Bauhaus-Universität Weimar
Faculty of Civil Engineering/Faculty of Media
Degree Programme Digital Engineering

Quantifying the Effects of Real-World Events on Wikipedia

Master's Thesis

Ali Moh'd Ali Saqallah
Born Dec. 12, 1995 in Amman, Jordan

Matriculation Number 123285

1. Referee: Prof. Dr. Benno Stein

Submission date: July 24, 2023

Declaration

Unless otherwise indicated in the text or references, this thesis is entirely the product of my own scholarly work.

Weimar, July 24, 2023

. . . .
Ali Moh'd Ali Saqallah

Abstract

This thesis presents a quantitative analysis of how real-world events influence editing behavior on the English Wikipedia, with the objective of comprehending their impact on overall editing activity and patterns. The study explores variations in editing behavior across different event categories and investigates Wikipedia's response to events through article protection. Our study reproduces previous work, establishes a novel analysis methodology, and quantifies the effects of 15 events across categories: Armed Conflicts and Wars, Elections, Natural Disasters, Sports and Entertainment Events, and Legal and Legislative Events.

Findings reveal a notable surge in editing activity following events, reflecting increased interest from editors. The Russian invasion of Ukraine event received the highest editing activity, while the Tigray War exhibited the highest effect size on total edits, despite less global attention. Reverted edits were most prevalent in articles related to the Tigray War, and vandalism-reverted edits were higher for controversial topics like same-sex marriage legislation in the United States. Editorial biases were evident, with events in the United States receiving more attention. Protective measures reduced vandalism during armed conflicts but had varied effectiveness in different categories. This research provides insights into the editing dynamics of Wikipedia during real-world events and highlights factors influencing editing behavior.

Contents

Acknowledgements

I would like to extend my sincere appreciation to Dr. Johannes Kiesel for his invaluable support and guidance during the completion of my thesis. His consistent availability for meetings, prompt responses to my queries, and unsolicited advice have played a crucial role in my academic journey. Furthermore, his expertise in presentation and public speaking has greatly contributed to the development of my confidence and communication abilities. I am profoundly grateful for Dr. Kiesel's unwavering assistance and mentorship. Additionally, I am thankful for the unwavering support of my parents throughout my entire academic endeavor.

Chapter 1

Introduction

Wikipedia, the biggest online encyclopedia globally, has owed its success to the active community of editors. These editors contribute, update, and maintain content, ensuring accuracy, quality, and comprehensiveness of the articles. Consequently, understanding the factors influencing Wikipedia editing behavior is of great interest. As highlighted by Chen and Iwaihara [2021], real-world events that attract substantial public attention often prompt the creation of Wikipedia articles. Furthermore, such events can significantly impact editing activity, especially when subsequent developments and findings emerge, leading to persistent revisions that can last for days or even months.

This thesis aims to quantitatively analyze the influence of real-world events on the editing behavior of the English Wikipedia. The research objectives encompass investigating the impact of real-world events on the overall editing activity of the English Wikipedia, exploring patterns and changes in editing behavior around these events, examining potential variations in editing behavior across different event categories, and studying Wikipedia's response to these events in terms of article protection.

This thesis presents the following contributions:

- Reproduction of [Kiesel et al., 2017]: We successfully reproduced the paper by Kiesel et al. [2017]. By replicating their findings with a newer Wikipedia history dump, we have verified the robustness and validity of their research, thereby establishing a strong foundation for our subsequent analyses.

- Development of Analysis Methodology: We have developed a novel analysis methodology to quantify the impact of real-world events on Wikipedia. This methodology provides a systematic and scalable approach that can be applied to different language versions of Wikipedia. By creating this methodology, we enable future researchers to analyze and compare the effects of events across various Wikipedia editions, thereby facilitating a deeper understanding of the platform's response to real-world events.

- Quantification of Event Impact: Our study quantifies the impact of real-world events on editing behavior on Wikipedia. By analyzing 15 different events across five distinct categories and examining patterns and changes in editing activity, we enhance our understanding of how events can shape the dynamics of editing on Wikipedia.

This thesis is structured as follows: Chapter 1 provides an introduction to the research topic. Chapter 2 reviews relevant literature to provide a theoretical background and context for the study. In Chapter 3, we replicate the findings of the paper by Kiesel et al. [2017] using a newer Wikipedia history dump, which serves to verify the functionality of the open-source software provided by the authors. This replication enables us to extract comprehensive editing data for our analysis, forming the foundational basis for our subsequent analyses. Chapter 4 outlines our research approach, detailing the data selection, analysis process, and defining key metrics. This chapter provides essential guidance for understanding the subsequent chapter. In Chapter 5, we present the results of our analysis, examining 15 real-world events categorized into armed conflicts and wars, elections, natural disasters, sports and entertainment, and legal and legislative events. Finally, Chapter 6 presents the conclusions drawn from our analysis, summarizing the findings of our study.

Chapter 2

Related Work

This chapter discusses relevant previous research related to the quantification of the effects of real-world events on the editing behavior of Wikipedia.

A study by Hu et al. [2021] titled "Predicting User Engagement on Twitter with Real-World Events" focuses on understanding factors influencing user engagement with real-world events on Twitter. The researchers analyze 2.7 billion English tweets, identifying 7,468 real-world event clusters and 22,957 Twitter users engaging with these events. Their study explores various predictors, including Twitter activities, tweet content, geolocation, and social network structure, translated into 17 numeric variables. The findings reveal that a user's prior Twitter activity and their social network significantly influence their engagement with events. Additionally, topical interests play a crucial role in engagement, particularly during political, business, sports, and sci-tech events. These insights can inform the understanding of how real-world events impact online behavior. While their research focuses on Twitter, this thesis specifically examines the editing behavior of Wikipedia in response to real-world events.

Liu et al. [2017] address the understanding and monitoring of real-time events, emphasizing the importance of tracking event evolution phases. The paper proposes a unified phase evolution mining model using k-means clustering, empirical rules, and burst detection algorithms to identify occurrence, development, climax, decline, and ending patterns of events based on post counts at specific time intervals. The TextRank algorithm is employed for topic analysis of each phase. The model's efficacy is demonstrated through experiments on real-world datasets from social media platforms. The paper's methodology and findings can provide insights into how real-world events impact online behavior and may reveal correlations with Wikipedia's editing behavior. While their study analyzes social media data, this thesis specifically focuses on the editing behavior of Wikipedia.

[García-Gavilanes et al., 2016] investigate the dynamics of attention to aircraft incidents and accidents on English and Spanish Wikipedia. The study analyzes Wikipedia's transactional data for 1606 English and 525 Spanish articles related to airline crashes. The research explores how factors such as the number of deaths, airline region, and event locale influence the level of attention given by Wikipedia editors and visitors. The study uses segmented regression analysis to model attention dynamics during the 50 days following an event, revealing patterns in communicative interaction, floating gap, and cultural memory phases. The findings shed light on attention biases and online information dissemination. While their research focuses on specific events, this thesis analyzes a broader range of real-world events.

[Fisichella and Ceroni, 2021] offers an effective approach for detecting dynamic relationships and events in Wikipedia edit records. The study proposes a novel approach for extracting complex event structures from Wikipedia using user edit records. The primary objective is to represent events involving multiple entities in a language-independent manner. The study introduces an evolution-aware entity-based enrichment algorithm to enhance entity accessibility and temporal retrieval on Wikipedia. By leveraging Wikipedia article links, the Explicit Relationship Identification method establishes connections between entities. Additionally, the Implicit Relationships Identification approach identifies entity connections using burst patterns with spikes based on real-world activities of the entities. These techniques, combined with fact detection, result in the detection of events. The paper also presents an event validation method that utilizes a supervised model to predict the presence of events in non-annotated corpora by incorporating the Web as an additional document source. Results indicate that the proposed approach achieves a high precision of 70% in event validation on a manually annotated corpus. Moreover, a comparison with Wikipedia's Current Event Portal demonstrates that the proposed method, named WikipEvent, along with the Co-References technique, can provide new and more data on events. While their study focuses on event detection using the editing behavior of Wikipedia, this thesis, on the other hand, specifically quantifies the effects of real-world events on the editing behavior of Wikipedia.

Chapter 3

Reproducing Spatio-Temporal Wikipedia Analysis

The objective of this chapter is to reproduce the results and conclusions presented in the original paper [Kiesel et al., 2017] and verify whether the software used in the analysis, which the authors made open source[1], is still functional and can be used in subsequent chapters, and if the reproduced results are consistent with the original ones.

The research conducted by Kiesel et al. [2017] addresses the issue of vandalism on Wikipedia, which poses a significant challenge to the community responsible for maintaining the integrity of articles. While Wikipedia is an open platform where anyone can edit articles, this freedom has also attracted vandals who damage articles instead of improving them. The authors highlight that the reviewing process of edits can only be handled manually up to a certain extent, which calls for the development of automated tools.

In their paper, the authors outline the methodologies used to address the problem of vandalism on Wikipedia and make three main contributions: ex post facto vandalism detection, historic editor geolocation, and spatio-temporal analysis. The authors conducted a systematic analysis of Wikipedia article revert graphs to identify vandalism and damaging edits, geolocated 77% of Wikipedia's anonymous editors since 2002, and conducted the first in-depth spatio-temporal analysis of Wikipedia's history, revealing a strong dependence of vandalism on various factors such as time of day, day of the week, country, culture, and Wikipedia language.

This chapter is structured the same way as the original paper, into three main sections: "Mining Vandalism," "Geolocating Editors," and "Spatio-Temporal Analysis." Each section comprises two subsections: "Results of Reproduction" and "Comparison with Original Results." In the first subsection, we present

[1] See `http://github.com/webis-de/ICWSM-17/`

the results of our reproduction, while in the second subsection, we compare our findings with the original paper, highlighting similarities and differences.

The reproduction was successful, yielding similar results with minor differences. These variations can be attributed to our utilization of a more recent and larger Wikipedia history dump.

3.1 Mining Vandalism

As we reproduce the results of the original paper, it is important to note that we use the same methodologies and approaches presented in Kiesel et al. [2017] to identify vandalism edits. Therefore, we adopt the approach outlined in the original paper to mine vandalism and define it as a ground truth for our analysis, relying on ex post facto evidence, specifically whether an edit had been manually or automatically reverted. We also follow the same self-reflection steps outlined by Howison et al. [2011] as in the original paper to ensure the validity of our approach. The identification of vandalism edits remains a crucial step in analyzing the history of Wikipedia articles and the adoption of the same methodology allows for consistency and comparability between our results and those presented in [Kiesel et al., 2017].

To identify vandalism, Kiesel et al. [2017] analyzed explicit comments left by community members, an approach that has also been utilized in previous studies [Kittur et al., 2007, Tran and Christen, 2013]. Another approach was to consider full-page reverts, which have been used in studies such as [Rzeszotarski and Kittur, 2012]. Moreover, they examined the revert graphs of Wikipedia article histories to filter out revert patterns that suggest good intentions on the part of an editor.

Kiesel et al. [2017] emphasized the importance of identifying past instances of vandalism as the ground truth in their research. In this context, ground truth refers to the definitive reference for determining whether an edit constitutes vandalism or not. In the case of Wikipedia, vandalism edits are typically reversed manually or automatically by restoring the most recent non-vandalized version of an article, resulting in what is known as a revert. A revert refers to the undoing of an edit, while the edits that are undone are referred to as reverted edits (see Figure 3.1 adapted from [Kiesel et al., 2017]).

To identify past instances of vandalism as the ground truth, we adopted the approach outlined in the original paper, which focused on full page reverts. A full page revert is a copy of the revision preceding the reverted edits, which is appended to the article's revision history (see Figure 3.1 adapted from [Kiesel et al., 2017]). We concur with their assertion that partial reverts are not reliable indicators of vandalism. As suggested by Kittur et al. [2007] and Flöck

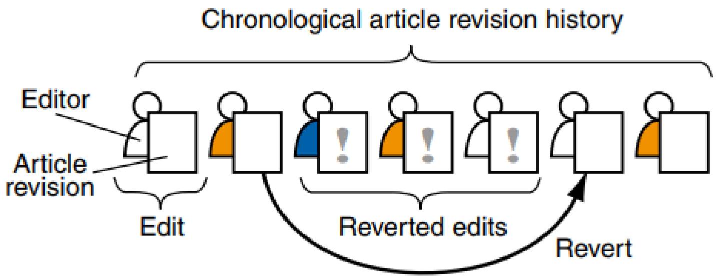


Figure 3.1: Taken from [Kiesel et al., 2017]. Illustration of a Wikipedia article revision history. Each revision is the result of an editor's changes to its preceding revision, yielding a chronological sequence of revisions by successive editors. Shade indicates different editors, arc arrows indicate reverts, where an old article revision is reinserted as new revision, undoing all intermediate revisions.

et al. [2012], partial reverts only account for a small percentage of vandalism cases.

3.1.1 Results of Reproduction

In this subsection, we present the results of our reproduction of the "Mining Vandalism" section from the original paper [Kiesel et al., 2017].

For our analysis, we use the full page reverts from all Wikipedia article histories comprised in the January 2023 Wikipedia history dumps. The English Wikipedia history dump, which is a 75 gigabyte compressed XML, contains a vast amount of edits and pages. We narrowed our focus to a specific subset of the data, which excludes user and discussion pages and only includes edits on articles. Additionally, it is important to note that due to article deletions, there are revision histories available for more articles than what is currently accessible on Wikipedia.

To ensure the reproducibility of the results presented in [Kiesel et al., 2017], we adopt the same approach and methodologies for detecting vandalism. We begin by identifying all full page reverts through the matching of SHA-1 hashes of article wikitexts ([Kittur et al., 2007]). If a SHA-1 value appears more than once in an article's revision history, any subsequent occurrences are considered reverts, and all the edits between the two instances are reverted. The first row of Table 3.1 displays the total number of detected reverts, which is 66.6 million, and the corresponding reverted edits, which amount to 175.4 million.

The patterns identified by [Kiesel et al., 2017] fall into four categories: pseudo-reverts, error-corrections, ambiguous reverts, and non-locatable editors. Pseudo-reverts (patterns (a) and (b)) involve unintended reverts, such as removing all content from an article or reverts that do not change the article. Error-corrections (patterns (c), (d), and (e)) capture reverts that are likely to be self-corrections or corrections of previous mistakes. Ambiguous reverts (patterns (f) and (g)) refer to reverts that are unclear or involve interleaved edits and edits by different editors. Non-locatable editors (pattern (h)) filters reverted edits made by registered editors who are responsible for less than 12% of reverted vandalism according to Kittur et al. [2007], bots, and editors whose location cannot be determined due to the lack of data in the geolocation databases.

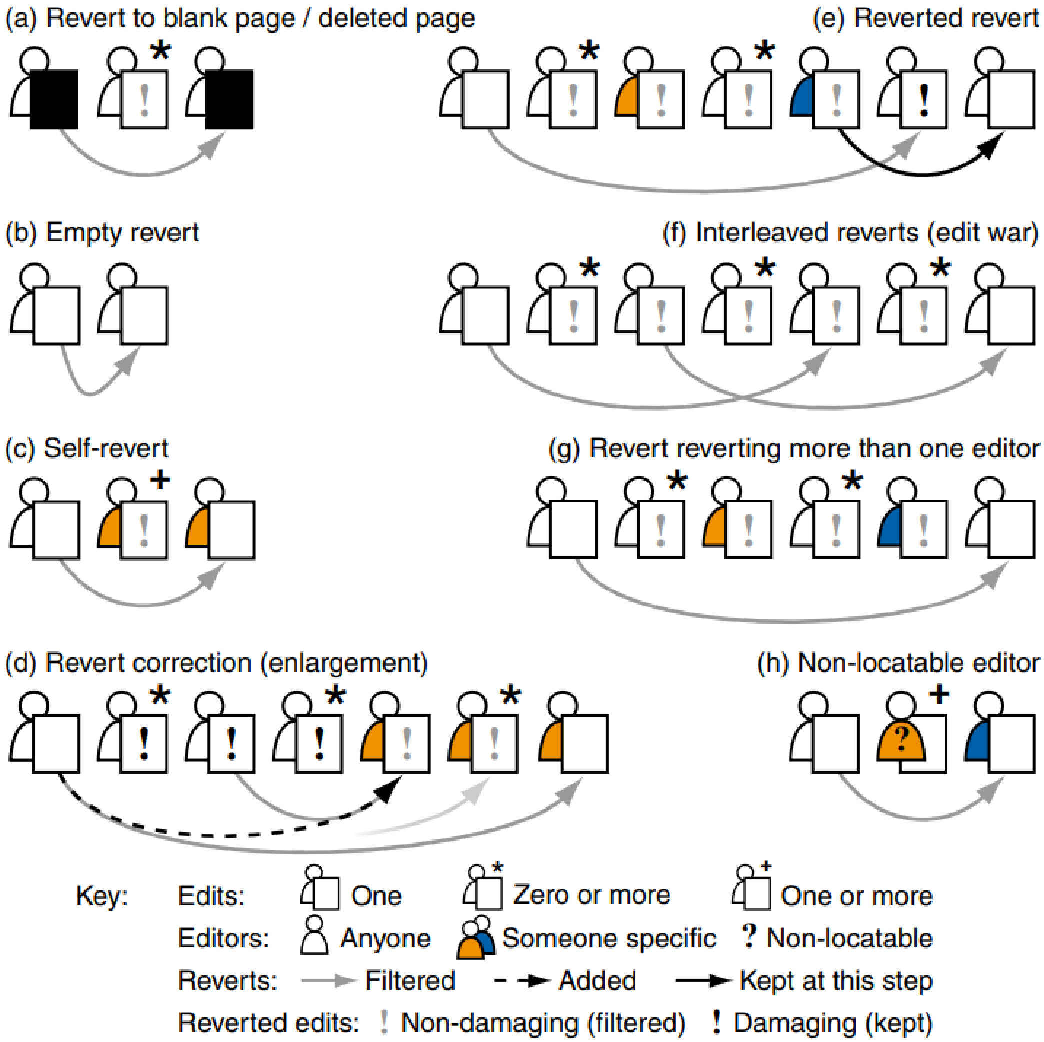


Figure 3.2: Taken from [Kiesel et al., 2017]. Revert patterns used for filtering full page reverts stepwise: first pseudo-reverts (a,b) are filtered, then error-corrections (c,d,e), ambiguous reverts (f,g), and finally reverts reverting edits of non-locatable editors (h). Each pattern depicts a regular expression that is matched against an article's revision history, filtering or reinterpreting reverts accordingly.

Table 3.1: The table corresponds to Table 1 in [Kiesel et al., 2017]. Step-by-step filtering of the English Wikipedia as per the revert patterns depicted in Figure 3.2 in the paper. Counts of full page reverts and counts of reverted edits affected by corresponding full page reverts are given. Full page reverts are analyzed for indications of vandalism in edit comments as per [Kittur et al., 2007], and reverted edits are divided into edits originating from editors who are anonymous, registered, or bots. Note that the approach by Kittur et al. [2007] uses specific words to classify comments and is thus likely less effective in finding vandalism for other languages than English.

Revert filtering step		**Full page reverts**				**Reverted edits**				
		Vandalism as per Kittur		Total		Editor			Total	
		No	Yes	Absolute	Relative	Anonymous	Registered	Bot	Absolute	Relative
Results of naive SHA-1 matching		58,038,755	8,607,778	66,646,533	100.0%	93,366,483	77,218,178	4,841,968	175,426,629	100.0%
(a) reverts to page blank		-1,424,729	-11,927	-1,436,656	-2.2%	-26,832,669	-37,437,581	-2,788,259	-67,058,509	-38.2%
(b) empty reverts due to renaming/removal/error		-3,228,493	-152,241	-3,380,734	-5.1%	0	0	0	0	0.0%
Results after filtering pseudo-reverts	Σ	53,385,533	8,443,610	61,829,143	92.8%	66,533,814	39,780,597	2,053,709	108,368,120	61.8%
(c) self reverts		-5,709,221	-11,453	-5,720,674	-8.6%	-3,573,793	-3,268,035	-309,247	-7,151,075	-4.1%
(d) revert corrections		-622,965	-84,673	-707,638	-1.1%	-1,326,594	-1,723,972	-42,081	-3,092,647	-1.8%
(e) reverted reverts		-477,367	-16,999	-494,366	-0.7%	-3,429,483	-5,264,954	-252,943	-8,947,380	-5.1%
Results after filtering error-corrections	Σ	46,575,980	8,330,485	54,906,465	82.4%	58,203,944	29,523,636	1,449,438	89,177,018	50.8%
(f) interleaved reverts		-7,053,336	-515,341	-7,568,677	-11.4%	-5,322,889	-7,714,637	-412,704	-13,450,230	-7.7%
(g) reverts reverting more than one editor		-2,604,164	-430,795	-3,034,959	-4.6%	-9,798,244	-7,246,434	-641,071	-17,685,749	-10.1%
Results after filtering ambiguous reverts	Σ	36,918,480	7,384,349	44,302,829	66.5%	43,082,811	14,562,565	395,663	58,041,039	33.1%
(h1) reverts reverting registered editors or bots		-9,845,910	-1,151,879	-10,997,789	-16.5%	0	-14,562,565	-395,663	-14,958,228	-8.5%
(h2) reverts reverting editors with IPv6 addresses		-2,523,980	-378,222	-2,902,202	-4.4%	-3,737,601	0	0	-3,737,601	-2.1%
Results after all filtering steps	Σ	24,548,590	5,854,248	30,402,838	45.6%	39,345,210	0	0	39,345,210	22.4%

3.1.2 Comparison with Original Results

In comparison to [Kiesel et al., 2017], our reproduction analysis utilizes the full page reverts from the January 2023 Wikipedia history dumps, which contain a larger amount of compressed file size, at 75 gigabytes. Similar to the original approach, we identify full page reverts by matching SHA-1 hashes of article wikitexts. Our analysis identified a total of 66.6 million reverts, which is higher than the 44.9 million identified by the original authors. Furthermore, we found a total of 175.4 million reverted edits resulting from the identified reverts, compared to 119.7 million reported by the original authors. These results are consistent with the expectation that more recent data would yield a higher number of reverts and reverted edits.

In the original paper, Kiesel et al. [2017] filtered out about 2.7 million reverts using patterns (a) and (b), while in the reproduction process, approximately 4.8 million reverts were filtered using the same patterns. This resulted in about 67 million unintentionally reverted edits being removed from the dataset, which is higher than the original paper's result of about 44.9 million reverted edits filtered.

Similarly, [Kiesel et al., 2017] filtered out about 4.7 million reverts using patterns (c), (d), and (e), while in the reproduction process, approximately 6.9 million reverts were filtered using the same patterns. This resulted in about 19.2 million reverted edits being filtered out, which is slightly higher than the original paper's result of about 13.7 million reverted edits.

Regarding the filtering of ambiguous reverts, Kiesel et al. [2017] filtered out about 7.1 million reverts using patterns (f) and (g), while in the reproduction process, approximately 10.6 million reverts were filtered using the same patterns. This resulted in about 31.1 million reverted edits being filtered out, which is higher than the original paper's result of about 21.5 million reverted edits.

Finally, Kiesel et al. [2017] filtered out about 7.2 million reverts using pattern (h), while in the reproduction process, approximately 13.9 million reverts were filtered using the same pattern. This resulted in about 18.7 million reverted edits being filtered out, which is slightly higher than the original paper's result of about 9.6 million reverted edits.

Overall, the reproduction process filtered out approximately 54.5% of all reverts and 77.6% of all reverted edits as harmless, ambiguous, or non-locatable, which is comparable to the results of [Kiesel et al., 2017] of 52% and about 75%, respectively. The reproduction process resulted in a ground truth of 39,345,210 edits that are considered to be vandalism originating from anonymous editors for subsequent analysis, which is similar to the original paper's result of 29,998,392 edits.

The original paper reported that out of all 44.9 million reverts analyzed, a total of 6,670,575 (14.9%) were classified as vandalism reverts. In comparison, our reproduction of the study identified 8,607,778 (12.9%) reverts as vandalism out of a total of 66.6 million analyzed. Our reproduction found a lower percentage of vandalism reverts overall.

Kiesel et al. [2017] filtered out 799,928 explicit vandalism reverts because they originated from registered users or bots, leaving a final recall rate of 73.3% of all explicit vandalism reverts. However, when disregarding those reverts from registered users and pseudo-reverts, the recall increases to 84.7% of the remainder. In contrast, our reproduction filtered out a greater number of explicit vandalism reverts from registered users or bots, with 1,151,879 reverts being excluded. This resulted in a lower recall rate of 68% of all explicit vandalism reverts. However, when excluding registered user and pseudo-reverts, it resulted in a slightly higher recall rate of 80.3% of the remainder.

In conclusion, our reproduction analysis of the study by [Kiesel et al., 2017] was able to replicate the findings of the original paper in terms of identifying vandalism reverts on Wikipedia, albeit with a lower recall rate. Our analysis, based on a newer and larger Wikipedia dump, identified a higher number of total reverts and reverted edits. These differences in recall rate and higher numbers of total reverts and reverted edits may be attributed to differences of the size of the data sets used. Importantly, the open source software provided by Kiesel et al. [2017] still works and can be used to mine vandalism with more recent data, as demonstrated by our reproduction analysis using the January 2023 Wikipedia history dumps.

3.2 Geolocating Editors

This section builds upon the same methodology presented in the original paper by Kiesel et al. [2017] for determining the geographic location of anonymous Wikipedia editors. The process involves utilizing geolocation databases (GeoDBs) to augment the server times and IP addresses provided in Wikipedia history dumps. However, since IP addresses may change location over time, special care must be taken to ensure reliable geolocation when dealing with historic IPs. Previous research on the accuracy of GeoDBs suggests that country information is reliable (with an accuracy above 95%) and that latitude/longitude coordinates typically have a tolerance of far below 1,000km [Poese et al., 2011].

In this section, we will show that the same approach as in the original paper can be used to reliably geolocate decade-old IP addresses in terms of country and time zone by combining GeoDBs with Regional Internet Registry (RIR) data. This will enable us to determine the geographic distribution of anonymous editors.

3.2.1 Results of Reproduction

To geolocate editors in our study, we followed the same methodology as described in the original paper [Kiesel et al., 2017]. To cross-check geolocations for consistency and remove inconsistent ones, we combined data from GeoDBs from IP2Location with RIR data.

The geolocation process for the editors' IP addresses was reproduced using the same methodology as outlined in [Kiesel et al., 2017]. The resulting flow diagram of decisions is depicted in Figure 3.3. The process involves the following steps:

1. Removal of the few IP addresses that have no corresponding RIR entry.

2. Check whether the IP addresses are contained in one or more GeoDBs that fall within a "RIR span" (characterized by the time span between the RIR entry directly before and the RIR entry directly after the Wikipedia time of an edit).

3. If yes, removal of IP addresses where RIR span and GeoDBs disagree on their country.

4. If they agree on the country, check whether they agree on their time zone as well.

5. In case of time zone disagreements within the GeoDBs, check whether the GeoDBs within the RIR span directly before and directly after an edit agree on time zones, and removal of all IP addresses where this is not the case. If yes, this corresponds to providers relocating an IP block within a multi-time-zone country, which is not recorded by RIRs.

6. When there is no GeoDB in the RIR span around an edit's time, check whether RIR geolocates to countries that have only one time zone, and removal of IP addresses where this is not the case.

By following the same set of rules and decisions as outlined in the original paper, we were able to reliably geolocate 85% of the 176,036,027 anonymous edits from the English Wikipedia, which corresponds to 149,742,112 edits.

Table 3.2 shows the numbers of edits removed/kept as a result of filtering IP addresses with unreliable geolocation, and the numbers of unique IPs whence they originated. The geolocated edits form an unbiased sample, with the ratio of reverted edits remaining identical at 22%. In total, our subsequent analysis is based on 33,487,624 reverted edits.

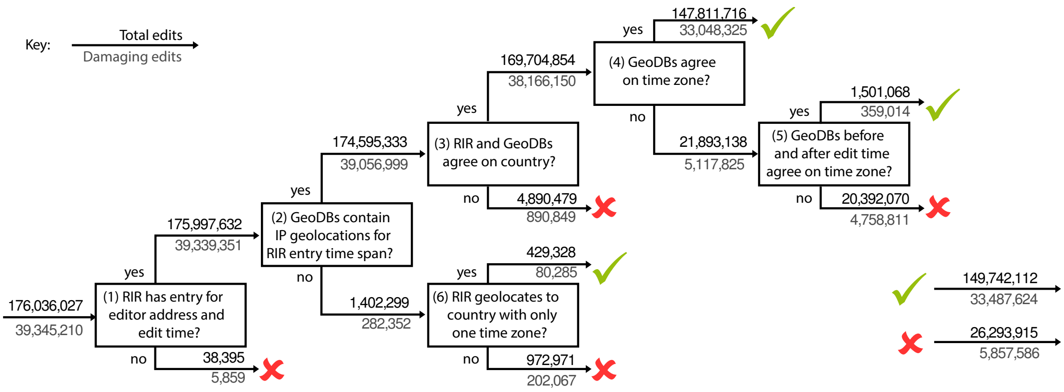


Figure 3.3: Decision tree to decide whether to trust the available geolocation information for an edit (✓), or not (✗). The numbers denote the total edits and reverted edits for the English Wikipedia that went through each branch.

Table 3.2: Historic geolocation success for all anonymous editors of the English Wikipedia in terms of edits and unique IP addresses whence they originated. Aside the totals, the subset of edits considered vandalism or damaging as per Section 3.1 are given, and their corresponding IP addresses. Numbers are given for each exit node of the decision tree in the Figure above, divided by whether or not the geolocation is trustworthy.

Decision Tree		**Edits**			**Unique IP addresses**	
Trusted	Exit Step	Vandalism as per Sec. 3		Total	Vandal IPs	Total
Entire Wikipedia		39,345,210	(22%)	176,036,027	15,517,050	44,376,630
No ✗	Step (1)	5,859	(15%)	38,395	2,579	8,017
	Step (3)	890,849	(18%)	4,890,479	338,340	1,150,972
	Step (5)	4,758,811	(23%)	20,392,070	1,838,212	5,051,004
	Step (6)	202,067	(20%)	972,971	80,345	229,516
	$\sum$	5,857,586	(22%)	26,293,915	2,259,116	6,437,426
Yes ✓	Step (4)	33,048,325	(22%)	147,811,716	13,101,420	37,527,106
	Step (5)	359,014	(23%)	1,501,068	145,038	375,367
	Step (6)	80,285	(18%)	429,328	39,538	138,973
	$\sum$	33,487,624	(22%)	149,742,112	13,285,756	38,040,070

3.2.2 Comparison with Original Results

In comparison to [Kiesel et al., 2017], our reproduced study achieved a higher percentage of reliably geolocated anonymous Wikipedia edits (85%) compared to the original study (77%), despite using a reduced number of GeoDBs. The difference in percentage can be attributed to the fact that our reproduced study only used GeoDBs from IP2Location, while the original study used 11 commercial GeoDBs from IPligence and IP2Location. This reduction in the number of GeoDBs may have affected the reliability of geolocations for some IP addresses, particularly in cases where the GeoDBs did not agree on the time zone. As a result, the reproduced study may have filtered out fewer IP addresses in step 4 of the filtering process, leading to a higher percentage of reliable geolocations. Despite this difference, both studies followed the same set of rules and decisions to filter out IP addresses with unreliable geolocation and obtained unbiased samples of geolocated edits. The ratio of reverted edits remained the same in both studies at 22%. In total, our reproduced study was able to subject 33,487,624 reverted edits to subsequent analysis, compared to 23,182,972 reverted edits in the original study.

3.3 Spatio-Temporal Analysis

This section builds upon the previous sections by investigating the patterns of vandalism edits in Wikipedia over time and space. Following the methodology of Kiesel et al. [2017], we calculated the ratio of vandalism edits among all Wikipedia edits per hour of the day and per location. Our analysis is limited to anonymous edits that can be reliably geolocated, which accounts for the majority of anonymous edits. As Kiesel et al. [2017] found no correlation between being geolocated and being vandalism, and we do not expect this restriction to affect our results significantly.

While our approach for detecting vandalism through the revert filter is designed to avoid mislabeling proper edits as vandalism, we acknowledge that some cases of vandalism may have been missed. Nonetheless, the vast majority of editors who indicate cleaning up vandalism do so by performing full page reverts (also found by [Kittur et al., 2007]), making it unlikely that our observations of the vandalism ratio and its spatio-temporal distribution are significantly different from what we observed.

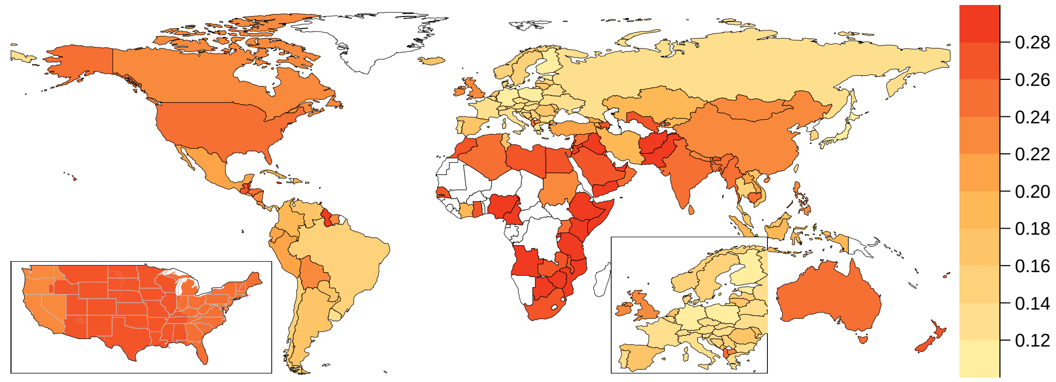
To estimate the vandalism ratio per hour of day, we averaged our results over all days since January 1, 2006. We excluded data from before this date, as the early stages of Wikipedia are known to have unstable vandalism ratios that are unreliable.

Based on the same methodologies used by the Kiesel et al. [2017], our reproduction findings for the spatio-temporal analysis of reverted Wikipedia edits are presented below. We used visual inspection, supported by thorough statistical analysis, to analyze the variances of the average vandalism ratios. Just like in the original paper, we employed Cohen's d to determine significant differences between visibly different graphs. We were able to confirm that large sample sizes (millions of edits) usually correspond to visibly different graphs and that high variances are indicative of vandalism ratios being influenced by other factors. Additionally, to account for vandalism ratio estimates based on few edits, we toned down these estimates in our figures. To establish the significance of all effects analyzed, as Kiesel et al. [2017], we utilized the Welch Two Sample t-test, and indicated p-values less than or equal to 0.05, 0.01, and 0.001 with one to three asterisks (*) respectively.

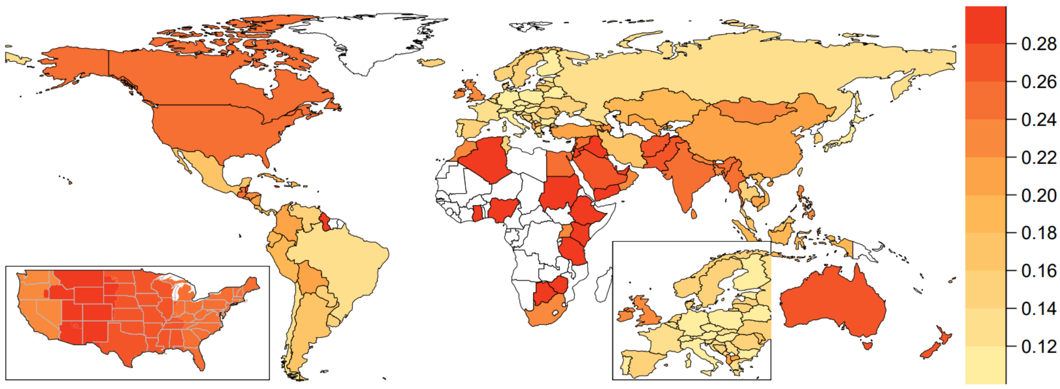
3.3.1 Results of Reproduction

According to our findings, we found that the vandalism ratio per country, as shown in Figure 3.4a[2], is highest in Africa. This finding is consistent with the original paper. Kiesel et al. [2017] hypothesized that this may be attributed to difficulties with the English language, leading native English editors to view edits from Africa as vandalism more frequently. However, only 1% of the geolocated edits to the English Wikipedia are from Africa, so we did not analyze the reasons behind this trend further. In Europe, the countries with the highest vandalism ratios were Albania, followed by Great Britain, Ireland, and North Macedonia, while Guyana had the highest vandalism ratios among South American countries.

[2]The map uses GADM 2.8 country/state data, http://www.gadm.org, and Efele 2016d timezone data, http://efele.net/maps/tz/.



(a) From our reproduction study.



(b) From the study conducted by Kiesel et al. [2017].

Figure 3.4: Ratio of vandalism to all edits in the English Wikipedia by country. Countries with less than 1,000 vandalism edits are not colored. The embedded small maps show (left) the vandalism ratio in the United States (without Alaska) by major time zone (from West to East: Pacific, Mountain, Central, and Eastern) with overlaid state borders and (right) Europe enlarged.

Vandalism Ratios in the United States

Figure 3.5a shows the variation in the vandalism ratio for edits to the English Wikipedia originating from the United States, along with the absolute number of all edits and vandalism edits. As a reference point, Figure 3.5a also presents a graph that only considers edits explicitly labeled as vandalism reverts with a corresponding editor comment, in accordance with Kittur et al. [2007]. The similarity between the two graphs provides further validation for using ex post facto vandalism detection method.

The majority of edits were made during the period between 14 and 17 hours. However, the ratio of vandalism to all edits showed distinctive peaks occurring much earlier at approximately 9 hours, with two additional peaks

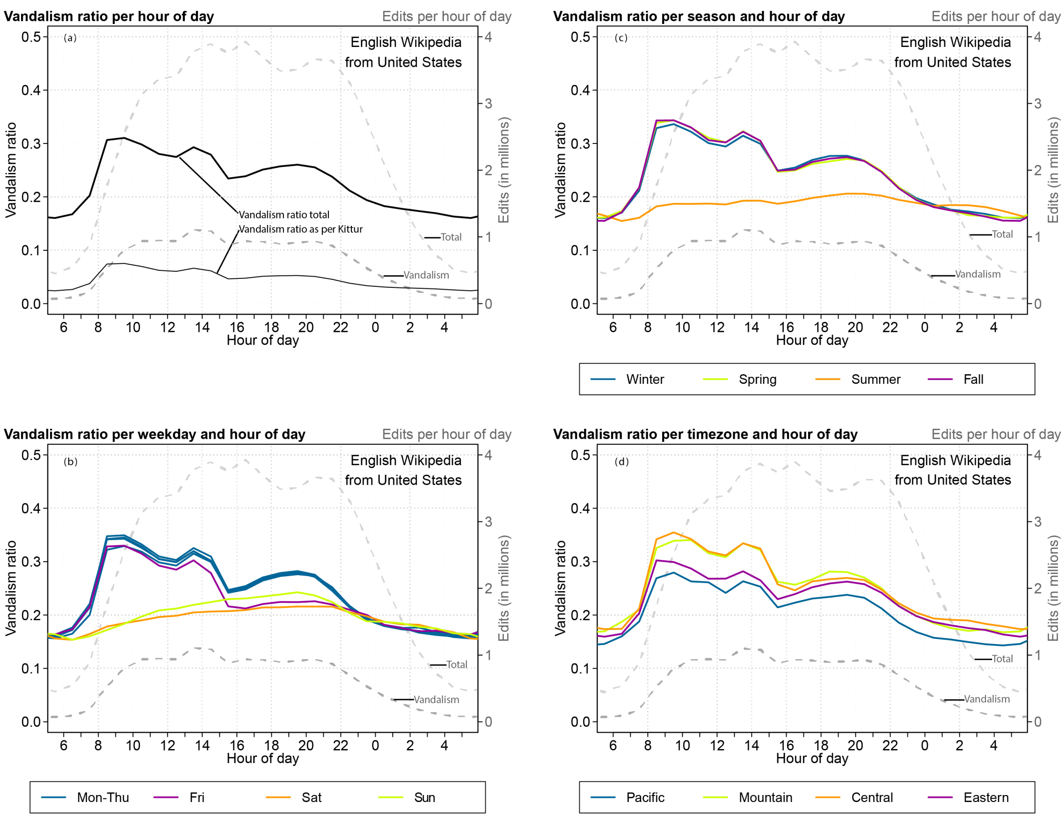


Figure 3.5: All plots show the ratio of vandalism to all edits per hour of day (left axis, solid lines), and for reference, the absolute number of edits and vandalism edits per hour of day (right axis, dashed lines), both averaged over Wikipedia's history. Plot (a) shows the overall ratio of vandalism edits on the English Wikipedia originating from the United States. Plots (b,c,d) divide the overall ratio by weekday, season, and US time zone. Ratios estimated from less than 1,000 vandalism edits are displayed with dotted lines.

identified at 13 and 19 hours. The hours between 23:00 and 08:00, considered as the "night time", had the lowest incidence of vandalism in both absolute and relative terms. During this time, approximately one in six edits was identified as vandalism, which sharply increased to approximately one in three edits during peak hours.

The difference in the vandalism ratio between night and day is visually apparent and is also supported by statistical analysis. The Cohen's d value between the average vandalism ratios for night and day demonstrates a strong statistical effect (d = 1.34***).

Our results from reproducing the study by Kiesel et al. [2017] confirm their findings that vandalism on Wikipedia is connected to working hours. Our plots also indicate that the highest peaks of vandalism occur in the morning (between 8 and 9 hours) and after lunch (between 13 and 14 hours). We support the hypothesis that vandalism is related to labor, as we observe an increase in vandalism ratio between 15 and 20 hours, which could be explained by people working long hours or relieving stress after work. This increase in vandalism ratio during workdays may also be due to an increase in negativity throughout the day, as found by Golder and Macy [2011] in their analysis of Twitter data.

Our statistical analysis shows a strong effect between Monday to Friday and Saturday plus Sunday for 8 to 15 hours (d = 1.40***), and a strong effect between Monday to Thursday and Friday to Sunday for 15 to 22 hours (d = 0.77***). The increase in the vandalism ratio on weekends has a medium effect, comparing the hour intervals (d = 0.46*** for Saturday and d = 0.61*** for Sunday). This increase could also be related to the increase in negativity found by Golder and Macy [2011].

Figure 3.5b provides further evidence for the labor-related vandalism hypothesis, as we observe a significant difference in vandalism ratios between workdays and weekends. The vandalism ratio is considerably higher on workdays than on Saturday and Sunday, which suggests that people are more likely to vandalize Wikipedia during working hours. On Fridays, the vandalism ratio graph is very similar to that of workdays up until around 16 hours, at which point it starts to resemble the graph of a weekend day.

Figure 3.5c shows that vandalism reduces during summer, possibly due to people going on vacation or being generally more relaxed. However, the effect size between summer and the other months for the time between 8 and 22 hours is only small (d = 0.33***), which is likely due to a large variance in the vandalism ratios from fall to spring. We agree with theKiesel et al. [2017]' statement that further investigation is necessary to establish a correlation between vandalism and other variables of interest. As they suggested, this is an area that could benefit from future research.

In line with Kiesel et al. [2017], we investigated the regional factors affecting vandalism ratios by dividing the United States into four distinct parts based on the commonly recognized time zones: Pacific, Mountain, Central, and Eastern. Our analysis produced results that aligned with the original study, as illustrated in Figure 3.4a (bottom left) and Figure 3.5d. Although the graphs exhibited similarities, we observed differences in vandalism ratios across the four time zones. However, the effect sizes were relatively small ($d < 0.24$). Our findings corroborate those of Kiesel et al. [2017], indicating that regional influences on vandalism ratios are minimal and that other factors may exert a more significant impact.

Vandalism Ratios across Countries

Our results, presented in Figures 3.6a-c, in line with the findings from the US, the analysis reveals a distinct variation in vandalism activity on Wikipedia during workdays versus weekends for the United Kingdom, Canada, and Australia. The corresponding effect sizes, denoted by d, are 0.94***, 1.04***, and 0.82*** for 8 to 15 hours, and 0.67***, 0.49***, and 0.65*** for 15 to 22 hours. Conversely, as shown in Figure 3.6d, the effect in India is much weaker, with effect sizes of 0.13** and 0.32*** for 8 to 15 hours and 15 to 22 hours, respectively. Kiesel et al. [2017] have hypothesized that this could be attributed to cultural differences in the value placed on work versus leisure time.

Interestingly, we observed that New Zealand had the highest vandalism ratios to the English Wikipedia among the countries we analyzed, as shown in Figure 3.6e.

Our reproduction shows that the vandalism ratio to the English Wikipedia is higher in countries with English as the official language. Our analysis of edits from Germany or France to the English Wikipedia and their respective "home" Wikipedias, presented in Figures 3.7(a-d), reveals that the vandalism ratios are indeed higher in the "home" Wikipedias. This finding is consistent with Kiesel et al. [2017] hypothesis that people tend to vandalize the Wikipedia variant of their mother tongue more frequently as it is an easier target and is usually ranked higher by search engines.

We note that there are differences in the magnitude of the vandalism ratios between the different language variants. For example, the English vandalism ratio for Germany is below 0.2 instead of reaching a striking 0.5 at 8 hours in the German Wikipedia, which is the highest ratio observed in our analysis. However, despite these differences in magnitude, the graphs in Figure 3.7(a-d) still exhibit similar peaks and valleys, indicating that people altogether follow a similar rhythm of life with vandalism ratios peaking when starting or continuing work or studies.

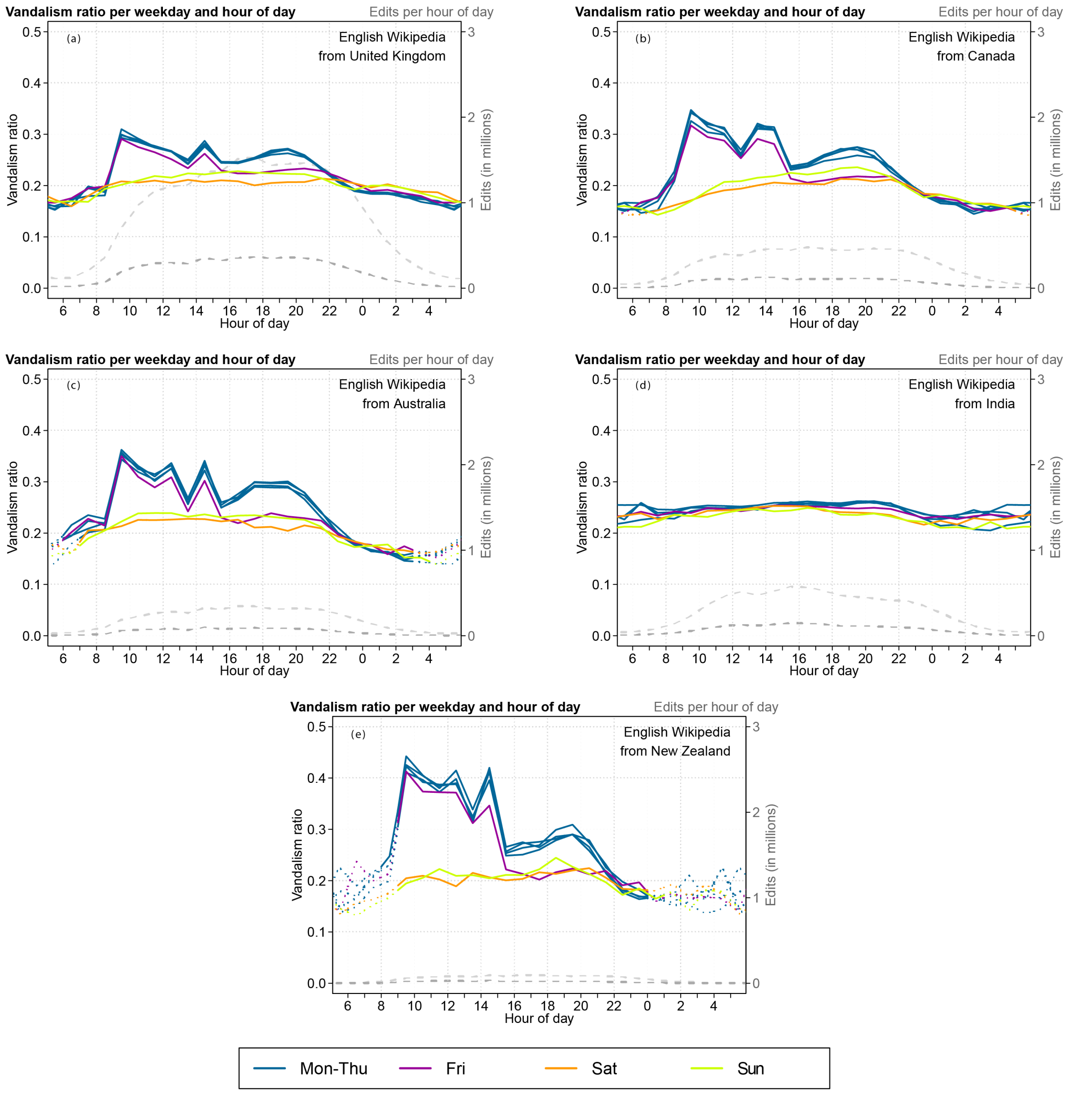


Figure 3.6: All plots show the ratio of vandalism to all edits per hour of day (left axis, solid lines), and for reference, the absolute number of edits and vandalism edits per hour of day (right axis, dashed lines), both averaged over Wikipedia's history. Plots (a-e) show vandalism ratios divided by weekday for the English Wikipedia edited from various countries. Ratios estimated from less than 1,000 vandalism edits are displayed with dotted lines.

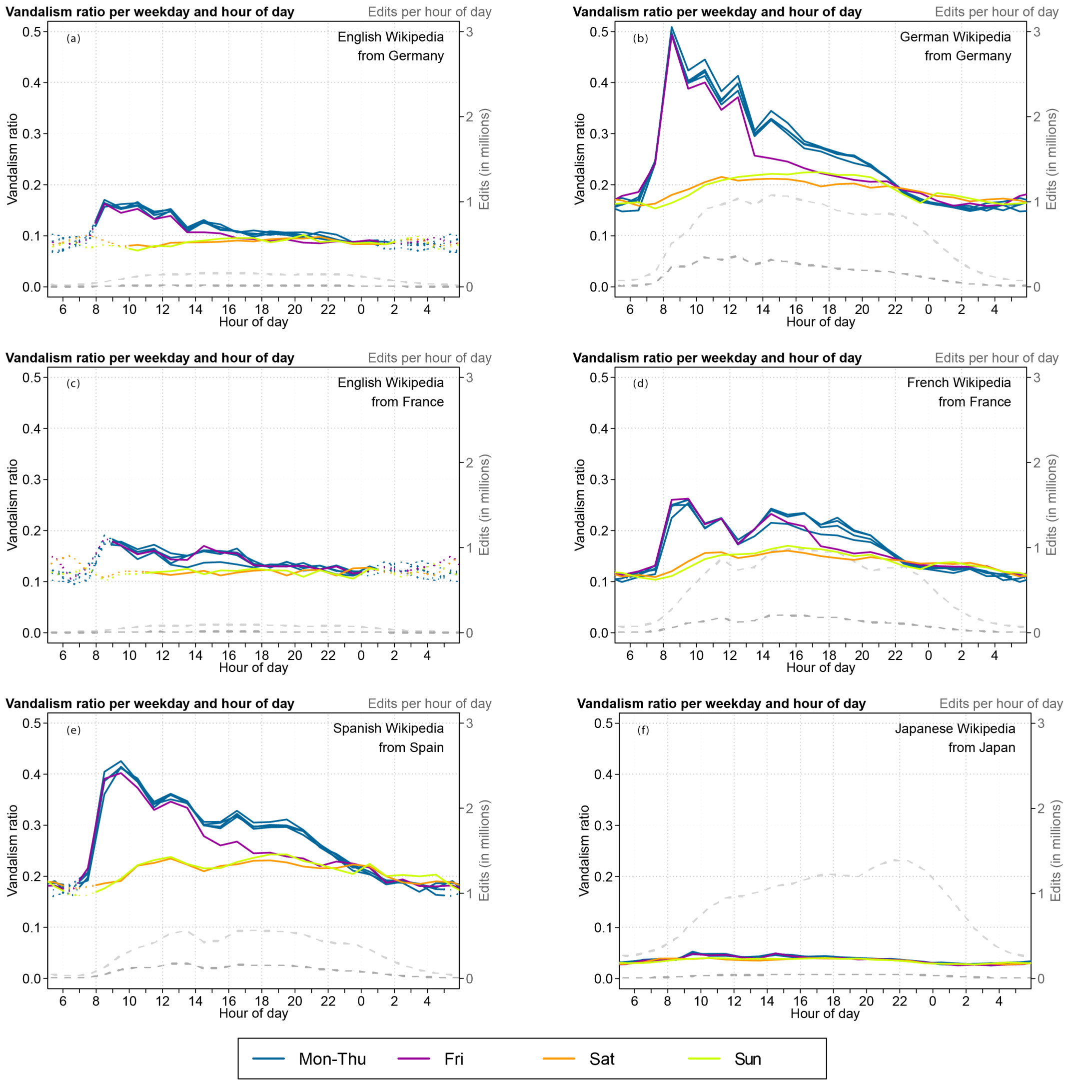


Figure 3.7: All plots show the ratio of vandalism to all edits per hour of day (left axis, solid lines), and for reference, the absolute number of edits and vandalism edits per hour of day (right axis, dashed lines), both averaged over Wikipedia's history. Plots (a,c) show vandalism ratios divided by weekday for the English Wikipedia when edited from Germany, and France. Plots (b,d) show vandalism ratios divided by weekday for the German, and French Wikipedias when edited from Germany, and France. Plots (e,f) show vandalism ratios divided by weekday for the Spanish, and Japanese Wikipedias when edited from Spain, and Japan. Ratios estimated from less than 1,000 vandalism edits are displayed with dotted lines.

We also observe the relatively low vandalism ratio for Wednesday afternoons for edits from France, is visible in the France-plots of both the French and the English Wikipedia, with a Cohen's d of 0.29*** and 0.12***, respectively.

Figure 3.7(e-f) depict the vandalism ratio for Spanish and Japanese Wikipedias, both of which are among the top 7 with the highest number of edits. The Spanish graph displays a pattern similar to that observed in the US, whereas the Japanese graph exhibits a remarkably low vandalism ratio with the only statistically significant variation being a higher rate of vandalism during the day as opposed to the night. This effect is not evident in the plot but is still of medium significance with a d value of 0.43*** owing to the limited variance. Therefore, in line with the original hypothesis, our findings demonstrate that time has a statistically substantial impact on the vandalism ratio. However, the results also reveal that cultural differences can exert an even greater influence, as evidenced by the low vandalism ratio observed in the Japanese Wikipedia despite variations in time.

3.3.2 Comparison with Original Results

Our findings, based on visual inspection using Figure 3.4, differ slightly from the original paper's observations on vandalism ratios in Europe. While Kiesel et al. [2017] found that countries with English as the official language, such as Great Britain and Ireland, had the highest vandalism ratios in Europe for the English Wikipedia, our results show that Albania, where Albanian is the official language, had the highest vandalism ratio. Great Britain and Ireland followed closely behind. However, North Macedonia had a higher vandalism ratio in our study, matching Great Britain and Ireland, whereas it had a lower ratio in the original study. For North America, our study found that Guyana had the highest vandalism ratio for the English Wikipedia, consistent with the original paper's findings. Despite these minor variations, our overall findings are consistent with the original paper's conclusions regarding the distribution of vandalism across different regions.

Vandalism Ratios in the United States

In comparison to the original findings, the reproduction study also identified distinctive peaks in the vandalism ratio occurring at around 9 hours, 13 hours, and 19 hours. The hours between 23:00 and 08:00 had the lowest incidence of vandalism in both absolute and relative terms. The visual difference in the vandalism ratio between night and day was also evident in the reproduction study, which found that approximately one in six edits was identified as vandalism during the night, increasing to approximately one in three edits during peak hours. However, the Cohen's d value between the average vandalism ratios for night and day was found to be lower in the reproduction study (d = 1.34***), indicating a slightly weaker statistical effect compared to the original study (d = 14.7***). The graph that only considers edits explicitly labeled as vandalism reverts with a corresponding editor comment also showed a high degree of similarity between the original and reproduction studies, providing further validation for using ex post facto vandalism detection method.

Our findings are consistent with the Kiesel et al. [2017]' conclusions that vandalism is connected to labor, with peaks of vandalism occurring when people start to work/study in the morning (8 to 9 hours) and after lunch (13 to 14 hours). Furthermore, our statistical analysis shows a similar pattern of a clear difference in vandalism ratios between workdays and weekends, with the vandalism ratio much higher on workdays than on Saturday and Sunday. However, our statistical analysis differs from [Kiesel et al., 2017] in terms of the effect size, which shows a slightly smaller effect between Monday to Friday and Saturday plus Sunday for 8 to 15 hours (d = 1.40***) and 15 to 22 hours (d = 0.77***), compared to the original paper's effect size of d = 1.49*** and d = 0.88***, respectively. Additionally, our analysis found a smaller increase in the vandalism ratio on weekends, with a medium effect size (d = 0.46*** for Saturday and d = 0.61*** for Sunday), compared to the original paper's medium to strong effect size (d = 0.53*** for Saturday and d = 0.68*** for Sunday).

The findings are very similar to those of Kiesel et al. [2017] in terms of the seasonality of vandalism. Like the original study, our analysis showed that vandalism reduces during the summer months, which could be attributed to people going on vacation or being generally more relaxed. The effect size between summer and other months for the time between 8 and 22 hours was also small (d = 0.33***), which is consistent with the original study's finding(d = 0.34***).

Our findings corroborate those of the original study by Kiesel et al. [2017] regarding regional influences, as we also observed minimal regional influences on vandalism ratios in the United States when dividing the country into four parts based on time zones. Specifically, we found that although there were slight differences in vandalism ratios across the four time zones, the effect sizes were relatively small ($d < 0.24$). These results align with the original study ($d < 0.30$), indicating that other factors may have a more significant impact on vandalism ratios.

Vandalism Ratios across Countries

We observe similar trends in vandalism activity across different countries and Wikipedia languages. Our results for the United Kingdom, Canada, and Australia are consistent with the findings reported by the original authors, with comparable effect sizes observed for the weekday versus weekend comparison. The effect sizes we obtained for these countries range between 0.82*** and 1.04*** for 8 to 15 hours, and between 0.49*** and 0.67*** for 15 to 22 hours, which are slightly smaller than the effect sizes reported byKiesel et al. [2017]. Conversely, we also find that the effect of the weekday versus weekend comparison is weaker for India, which is consistent with the original findings. The effect sizes we observed for India are similar to those reported by Kiesel et al. [2017], with effect sizes of 0.13** and 0.32*** for 8 to 15 hours and 15 to 22 hours, respectively. Interestingly, we also find that New Zealand had the highest vandalism ratios for the English Wikipedia among the countries we analyzed, which was not reported by Kiesel et al. [2017].

3.4 Summary

In this chapter, we have successfully replicated the results of the paper "Spatio-temporal Analysis of Reverted Wikipedia Edits" by Kiesel et al. [2017]. We have employed the same methodologies used in the mining vandalism, geolocating editors, and spatio-temporal analysis sections, utilizing the open-source software provided by the original authors. Our study has yielded similar results to the original paper, with minor variations resulting from our use of a larger and more recent Wikipedia history dump.

Importantly, our analysis indicates that the editing behavior on Wikipedia appears to have remained relatively consistent in recent years. This observation suggests a certain time-independence for our work, as the patterns and trends identified in the original study still hold true with the more recent dataset used in our reproduction analysis.

To summarize, this chapter has verified the validity and reproducibility of the original findings. The open-source software provided by the authors remains fully functional, and the methodologies utilized are effective for subsequent chapters of our thesis.

Chapter 4

Data Selection and Analysis Approach

This chapter serves as a guide to our research approach, data collection, and selection methods for studying the impact of real-world events on Wikipedia editing behavior. By reading this chapter, readers will gain an understanding of how we selected and analyzed relevant data, defined key metrics, determined analysis timeframes, and utilized statistical techniques to quantify the effects of real-world events on Wikipedia's editing activity. This understanding provides a solid foundation for comprehending the subsequent chapter 5 of our study, where we explore the relationship between real-world events and the editing dynamics within Wikipedia.

Using the editing data obtained from Section 3.2, we quantify the effects of real-world events on editing behavior within Wikipedia by following the approach depicted in Figure 4.1 and detailed in the following sections. Our analysis covers two timeframes: an 8-week period that examines the immediate effects before and after the event, and a 12-month period to capture any long-term effects. These timeframes include pre-event and post-event periods, as explained in Section 4.1. We measure key metrics such as total edits, reverted edits, reverted-vandalism edits, top articles contributing to the analysis of the events, and the number of protected articles, as outlined in Section 4.2. These metrics offer insights into the overall interest and engagement among the Wikipedia community, conflicts among editors, and any malicious activities during and after the events. We examine the top articles contributing to the analysis to understand the aspects of events that generate the most interest. By checking the anonymous editing protection status of articles, we can observe Wikipedia's response to real-world events. Additionally, we employ a title-based search approach, as described in Section 4.3, to identify and select articles directly associated with the events under investigation, ensuring the

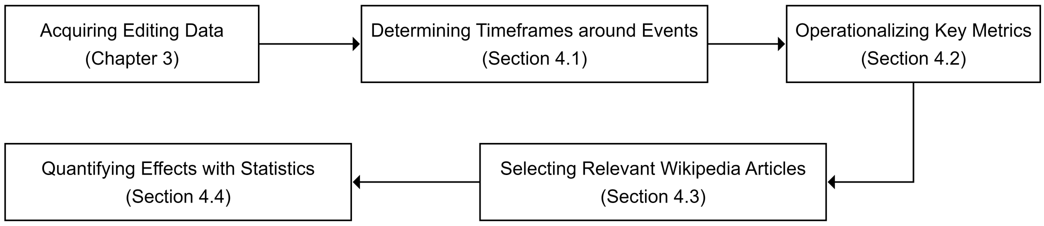


Figure 4.1: Flow chart illustrating the systematic approach to quantifying the effects of real-world events on Wikipedia editing behavior.

inclusion of relevant content while minimizing unrelated articles. Finally, we use statistical analysis techniques to analyze the effects of real-world events on the selected relevant articles, as explained in Section 4.4. By following this approach, we aim to identify significant patterns and trends related to the effects of real-world events on Wikipedia.

4.1 Determining Timeframes around Events

Determining an appropriate timeframe for analyzing the effects of real-world events on Wikipedia editing is crucial in order to capture relevant editing activity and filter out unrelated editing activity that occurs far from the event. This allows for meaningful comparisons between different events, identifying similarities or differences. After initially reviewing editing activity around the initial events, we determined an 8-week timeframe for event analysis. This entails analyzing 4 weeks before the event and 4 weeks after, enabling us to examine the editing behavior leading up to the event and the immediate effects of the event on editing behavior during the post-event period.

Additionally, apart from analyzing the effects of real-world events within an 8-week timeframe, we can broaden our perspective on editing dynamics and assess whether there are any distinct patterns or shifts beyond the immediate pre-event and post-event periods. This can be achieved by graphically interpreting edits over a 12-month period, including 6 months before and 6 months after the events.

4.2 Operationalizing Key Metrics

To examine the effects of real-world events on Wikipedia editing dynamics, this study focuses on five key metrics: "total edits," "reverted edits," "vandalism-reverted edits," "top articles and their contribution," and "anonymous editing protection" of articles. By operationalizing these key metrics, we aim to comprehensively analyze the relationship between real-world events and the editing behavior within the Wikipedia community, as well as Wikipedia's response to these events. Each metric is operationalized as follows:

Total Edits: Total edits are used as a measure of overall editing activity and engagement by editors. The number of edits made to Wikipedia articles related to selected events is recorded and compared in the four weeks preceding and following each event. By examining changes in the total number of edits, this metric aims to identify patterns that reflect the level of interest and engagement of editors in the specific event-related topics.

Reverted Edits: Reverted edits may indicate conflicts or disagreements among editors. The number of reverted edits is analyzed before and after real-world events to understand the extent of conflicts that arise during these periods. This metric helps determine whether the occurrence of events has an impact on the level of disagreement among editors and if certain topics provoke more conflicts than others.

Vandalism-Reverted Edits: Vandalism-reverted edits highlight malicious attempts to disrupt or manipulate article content. The number of vandalism-reverted edits is examined in relation to real-world events to find whether there is an increase in such behavior during these periods. This metric provides insights into how events may influence malicious activities on Wikipedia.

Top Articles and their Contributions: We identify the top articles related to the events of interest and examine their individual contributions in terms of the total number of edits. We aim to identify the specific topics that attract the most attention and editing activity during and after real-world events.

Number of Protected Articles: Wikipedia employs various types of article protections[1]. However, for the purpose of this analysis, we will focus on protections that prevent anonymous editing. We will count how many articles among the top 10 contributing articles by total edits were protected during the analyzed timeframe, to understand which articles editors perceive as more vulnerable to vandalism or controversial edits and protect them. This information helps us understand how Wikipedia editors react to real-world events.

[1] `https://en.wikipedia.org/wiki/Wikipedia:Protection_policy`

4.3 Selecting Relevant Wikipedia Articles

The selection of relevant Wikipedia articles for event analysis is crucial. A "relevant article" in the context of event analysis refers to an article that is directly related to the event being studied. It provides specific information, details, and perspectives about the event itself. To analyze an event, multiple relevant articles about that event are needed. In this study, four approaches were explored to identify articles related to real-world events. These approaches included solely focusing on the main article of the event, selecting articles based on the main category or related categories of the main article, and utilizing title search results. Among these approaches, the fourth approach was found to be the most effective and thus chosen for our analysis. The approaches can be summarized as follows:

Approach 1: This approach involved the selection of the main article manually corresponding to the event of interest for analysis. However, this approach presented a limitation in many cases, as our investigation of the effects of real-world events on Wikipedia required a comparison of editing activity before and after the event. The specified article, in most instances, was created at the time of the event, resulting in a lack of data regarding editing patterns leading up to the event. This limitation makes it difficult to conduct a comparison. Consequently, by solely analyzing the main article, we could primarily gain insights into the immediate aftermath of the event, rather than fully capturing the broader dynamics and long-term changes in editing activity. Therefore, this approach proved insufficient for capturing the impact of real-world events on Wikipedia's editing dynamics.

Approach 2: This approach involved selecting the main article corresponding to the event of interest and identifying its main category. The main category for the event was defined as the one that is titled the same as the main article. All the articles under this main category were then extracted and grouped for analysis. However, this method resulted in a significant number of unrelated articles within the category. For instance, when considering the first event analyzed in our study (Section 5.2.1), the specified main article was "Russian invasion of Ukraine," and all the articles from its main category were extracted. In total, there were 26 articles. Out of these 26 articles, only 5 of them had the words "Russian invasion of Ukraine" in their titles, indicating a direct relation to the event. However, it is important to note that 10 pages out of the 26 (38%) did not include any reference to Russia, Ukraine, or the invasion event in their titles. While these pages may still be related to Ukraine or Russia in some capacity, they did not appear directly relevant to the studied invasion event. Furthermore, out of these 10 articles not directly related to the event, 9 of them had zero revisions during the analysis period, suggesting a lack of activity or connection to the invasion event itself. Another obstacle encountered with this approach is the absence of a main category for certain events. In some cases, the main article might not have a corresponding category with the same title. This poses a difficulty in determining the main category and extracting relevant articles for analysis. Therefore, this approach proved insufficient for capturing the impact of real-world events on Wikipedia's editing dynamics.

Approach 3: This approach involved selecting the main article corresponding to the event of interest and identifying a directly related category. All the articles under this related category were then extracted and grouped for analysis. However, this approach introduced inconsistencies due to the presence of multiple related categories associated with an article, which made the selection of a specific related category subjective for articles that are associated with various related categories. Furthermore, for instance, when considering the first event analyzed in our study (Section 5.2.1), the specified main article was "Russian invasion of Ukraine," and the related category was identified as "Category:Russo-Ukrainian War." Extracting all articles within that category and grouping them for analysis resulted in 59 articles. Among these articles, only 6 of them included the words "Russian invasion of Ukraine" in their titles, indicating a direct association with the event of interest. Additionally, similar to the second approach, 25 articles out of the 59 articles (42%) within this category did not include any reference to Russia, Ukraine, or the invasion event in their titles, suggesting a lack of a direct connection to the invasion event. Therefore, this approach proved insufficient for capturing the impact of real-world events on Wikipedia's editing dynamics.

Approach 4 (Chosen Approach): The chosen approach involves searching on Wikipedia using the title of the main article related to the specific event we are investigating. We select a specific number of search result articles that include the keywords either in their titles or text content. However, we faced a limitation in determining the appropriate number of search result articles to select. To address this limitation, we examine the search results until we observe a point where the relevance begins to decrease or the articles become unrelated to the event. We capture the articles until that point, as they are the most relevant ones for the event or topic. The number of articles we capture depends on the search results obtained and the point at which the relevance begins to decline and ranges between 10 and 146. This iterative selection process ensures that the chosen articles are highly relevant to the event of interest. This approach has proven to be consistent in selecting relevant articles, ensuring a higher level of relevance to the event of interest. It captures a comprehensive range of articles directly related to the chosen real-world events, reducing the inclusion of unrelated content. Therefore, this approach was chosen.

4.4 Quantifying Effects with Statistics

We conducted a comparative analysis to compare the total number of edits, reverted edits, and vandalism-reverted edits between the periods before and after the event. This analysis aimed to determine the relative change, supported by Cohen's d, in order to quantify the effect of the event on engagement with these articles. This includes conflicts between editors, instances of vandalism, as well as the effect size and significance of these changes.

As part of the analysis, we reviewed the top 10 articles that contributed to the study, along with their protection status regarding anonymous editing. This information reveals which articles received the most attention from editors and which ones were protected as a preventive measure against vandalism during the event. This information helps us understand how Wikipedia editors react to real-world events.

Furthermore, we compared events within the same category and also compared categories with each other. This approach can help us to identify different or similar effects and trends associated with different events.

Chapter 5

Results of Event Impact on Editing Activity

This chapter investigates the impact of 15 real-world events on the editing behavior of the English Wikipedia, following the approach discussed in Chapter 4, and focusing on five categories. We will identify three events within each category and analyze the corresponding articles. Appendix A includes all the articles that were considered relevant for each event and were analyzed in this thesis. By comparing patterns and changes in editing behavior for these events within each category and across all categories, our goal is to understand how different events can impact the level of attention generated within the Wikipedia community, the level of disagreements among editors, and the potential impact on malicious attempts to disrupt or manipulate article content. Additionally, we aim to find which articles attract the most attention and editing activity during these events and explore strategies employed by editors in response to these events. The five event categories considered, along with the analyzed events, are shown in Table 5.1.

The selection of the five event categories analyzed in this chapter was based on their relevance and potential impact on the editing behavior of the English Wikipedia. These categories were chosen to represent a diverse range of real-world events that attract significant attention and have the potential to generate substantial editing activity on Wikipedia. Our objective is to examine how different types of events, each with distinct characteristics, can influence the editing behavior on Wikipedia. We also aim to explore the strategies employed by editors in response to these specific categories and events, as well as identify the articles that receive the most attention and editing activity during these periods. To illustrate further, armed conflicts and wars often evoke strong emotions and global interest, resulting in extensive updates and revisions of related Wikipedia articles. Elections, on the other hand, represent crucial po-

Table 5.1: Categories and events being analyzed.

Category	**Event**
Armed Conflicts and Wars	Russian invasion of Ukraine 2021 Israel–Palestine crisis Tigray War
Elections	2020 United States presidential election 2021 German federal election 2018 Bangladeshi general election
Natural Disaster	Hurricane Harvey 2018 Sulawesi earthquake and tsunami 2018 Kerala floods
Sports and Entertainment Events	2020 Summer Olympics Super Bowl LV 94th Academy Awards
Legal and Legislative Events	Same sex marriage legislation in the United States Legalization of cannabis in Canada General Data Protection Regulation

litical events that can trigger significant engagement from editors. Natural disasters have the potential to prompt a surge in editing activity as people seek to share information, provide aid resources, and document the impact of such events. Sports and entertainment events, with their wide viewership and public interest, can also drive increased editing and updating of related articles. Lastly, legal and legislative events, such as changes in laws or regulations, may lead to discussions and revisions of relevant Wikipedia articles to accurately reflect the new developments.

During our analysis of the events, we discovered that geolocated edits accounted for only a small percentage of the total edits made during the analyzed period. This can be attributed to the fact that the majority of edits were carried out by registered users who actively contributed to and updated information about the event developments. Furthermore, certain events had article protections in place, which restricted anonymous edits. As geolocated edits are limited to anonymous editors, the number of such edits was further reduced. Thus, geolocated edits are not the main focus of this thesis. However, we have provided the number of geolocated edits for the first event.

Our analysis involves comparing the count of editing activities before and after each event to determine their impact. However, the sample sizes vary for each event included in our analysis. To support our findings, we use Cohen's d to measure the effect size for edits, reverted edits, and vandalism-reverted edits before and after each event. Furthermore, we assess the significance of these effect sizes by conducting the Welch Two Sample t-test. The significance levels are denoted by asterisks (*), with one, two, or three asterisks indicating p-values less than or equal to 0.05, 0.01, and 0.001, respectively.

Table 5.2: Total Edits for the Top 10 Articles during Analyzed Events. The table covers 8 weeks, consisting of 4 weeks before and after the events. It displays the edit counts for each event before and after their occurrence, focusing specifically on the top 10 articles. Additionally, it provides counts of protected articles (🔒) and the instances of article protection implementation during the analysis period (@). Furthermore, the table includes the effect size (Cohen's d) analysis, which evaluates the impact of the events on the total, reverted, and vandalism-reverted edits across all articles, not limited to the top 10.

Category/Event	**Top 10 articles**					**Cohen's d**		
	$\sum$	$\leftarrow$	$\rightarrow$	🔒	@	Edits	Reverts	Vand.
Armed Conflicts and Wars								
Russian invasion of Ukraine	17,011	1,625	15,386	5	4	1.95***	0.91**	0.05
2021 Israel-Palestine crisis	3,023	101	2,922	9	5	1.52***	0.41	0.10
Tigray War	1,783	213	1,570	2	0	2.42***	1.45***	0.52
Elections								
2020 United States presidential election	5,290	1,885	3,405	1	0	0.89**	0.73**	0.15
2021 German federal election	1,321	293	1,028	0	0	0.59*	0.03	0.01
2018 Bangladeshi general election	350	133	217	0	0	0.20	0.00	
Natural Disasters								
Hurricane Harvey	5,200	744	4,456	2	1	1.83***	0.95***	0.42
2018 Sulawesi earthquake and tsunami	699	16	683	0	0	0.85**	0.55*	0.15
2018 Kerala floods	851	95	756	0	0	1.36***		
Sports and Entertainment Events								
2020 Summer Olympics	14,020	1,806	12,214	0	0	1.47***	1.06***	0.66*
Super Bowl LV	1,912	629	1,283	1	0	0.22	0.20	0.17
94th Academy Awards	4,410	1,429	2,981	1	0	0.33	0.29	0.18
Legal and Legislative Events								
Same-sex marriage legislation in the United States	1,791	452	1,339	1	0	0.68*	0.70*	0.70*
Legalization of cannabis in Canada	623	231	392	2	0	0.48	0.32	0.45
General Data Protection Regulation	441	189	252	0	0	0.17	0.19	0.27

5.1 Main Findings

To conduct a comparative analysis of editing behavior across events, we encountered variations in the number of relevant articles for each event. To address this challenge and ensure effective comparisons, we implemented a standardized approach. Table 5.2 shows the total number of edits for the top 10 articles contributing to each event, including edit counts for these articles four weeks before and after the analysis period. Additionally, the table provides information on how many articles were protected during the analysis period to safeguard them from vandalism. Furthermore, it includes the effect size (Cohen's d) analysis, evaluating the impact of the events on the total, reverted, and vandalism-reverted edits across all articles, not limited to the top 10.

In all 15 events analyzed, there was a notable surge in editing activity subsequent to the occurrences, indicating a heightened interest and involvement from editors in updating the information. However, in terms of the total number of edits from the top 10 articles contributing to each event, the Russian invasion of Ukraine event received the highest editing activity in total during the analysis period and following the event. This can be attributed to its significant global impact, extensive media coverage, and the high level of international concern it generated. The 2020 Summer Olympics ranked second, which can be attributed to the extended duration of the Olympic Games, spanning 17 days and generating ongoing engagement and updates for each day's competitive events. However, in terms of the effect size (Cohen's d), the Tigray War (which received less global attention) had the highest effect size on total edits, as it had ongoing developments, sustaining a high editing activity for the whole 4 weeks after the event began, with a $d = 2.42$***. Ranked second was the Russian invasion of Ukraine with $d = 1.95$***. From Table 5.2, we can see that events like Armed Conflicts and natural hazards have large effect sizes on edits, having the highest impact as expected since they usually have ongoing developments and editors keep updating information. On the other hand, events that take only one day had smaller effects, such as the Super Bowl LV and the 94th Academy Awards, have small to moderate effect sizes, $d = 0.22$ and $d = 0.33$, respectively.

The Tigray War exhibited not only the highest effect size on edits but also the highest effect size on reverted edits, with $d = 1.45$***. Approximately 4.6% of the total edits made following the event were reverted edits, indicating a greater level of disagreements and disputes among editors regarding this particular event. In comparison, The Russian invasion of Ukraine and the Israel-Palestine crisis experienced a reverting rate of 0.5% and 0.7% of total edits following the event, respectively.

For vandalism-reverted edits, the same-sex marriage legislation in the United States event received the highest effect size for vandalism-reverted edits with $d = 0.70$* across all categories. The controversial nature of the topic likely contributed to increased discussions and engagement from editors.

The extent of protection from anonymous editing varied across different event categories. Editors perceived armed conflict and war-related articles as more vulnerable to vandalism and therefore applied protective measures accordingly. Conversely, such measures were not as prevalent in other categories. Within the Armed Conflicts and Wars category, the analysis revealed that during the period under study, articles related to the Russian invasion of Ukraine and the 2021 Israel-Palestine crisis were protected from anonymous editing, whereas the Tigray War articles lacked similar protective measures. Consequently, vandalism-reverted edits decreased in the Russian invasion of Ukraine and the 2021 Israel-Palestine crisis articles, while the Tigray War articles experienced an increase in vandalism-related edits. This highlights the effectiveness of implementing measures such as article protection from anonymous editing in minimizing vandalism during armed conflicts.

However, it is important to note that even with article protection, certain topics, such as an article that is relevant to the same-sex marriage legislation in the United States event, titled "Same-sex marriage," remained prone to vandalism. Despite the restrictions on anonymous editing, the article still received 5 vandalism-reverted edits following the relevant event, suggesting that sometimes even with anonymous editing restrictions, vandalism couldn't be avoided when it comes to controversial subjects such as same-sex marriage.

Furthermore, it is worth noting that the "Will Smith–Chris Rock slapping incident" article received a notably higher number of reverted edits and vandalism compared to other articles related to the 94th Academy Awards event. Specifically, out of the 7 vandalism-reverted edits documented after the event, 5 were attributed to this particular article. This suggests that the article attracted a significant amount of vandalism, possibly due to its humorous or meme-like content. Surprisingly, the article was not protected from anonymous editing, which may have contributed to the increased frequency of such malicious edits.

Across three categories—Elections, Natural Disasters, and Legal and Legislative Events—, we included one event that occurred in the United States while the other two events took place in different countries. Remarkably, the event that occurred in the United States received significantly more attention and editing activity on Wikipedia compared to the other two events. For instance, in the Elections category, the top 10 articles related to the 2020 United States presidential election received four times more edits than the 2021 German federal election, and about 15 times more edits than the 2018 Bangladeshi general election. Similarly, in the Natural Disasters category, the top 10 articles related to Hurricane Harvey received nearly 7.5 times more edits than the 2018 Sulawesi earthquake and tsunami, and approximately 6 times more edits than the 2018 Kerala floods. In the Legal and Legislative Events category, the top 10 articles related to Same-sex marriage legislation in the US received three times more edits than the Legalization of cannabis in Canada, and four times more edits than the General Data Protection Regulation. The observed discrepancy in editing activity for events in the United States compared to similar events in other countries can be attributed to underlying editorial biases, influenced by the dominance of Western media in Wikipedia ([García-Gavilanes et al., 2016]). These biases favor events happening in North America and are consistent across different language Wikipedia editions, regardless of the origin of viewers or editors. Consequently, events occurring in the United States are more likely to receive detailed coverage and higher levels of engagement from editors and viewers on Wikipedia compared to events from other countries.

Another notable finding from our analysis is that a majority of the edits made during these events were contributed by registered users. The lowest percentage of registered edits, at 71%, was observed during the 94th Academy Awards event. In contrast, the 2021 Israel-Palestine crisis event exhibited the highest percentage of registered edits, accounting for 96% of the total edits. This pattern is consistent with expectations, as registered users typically show more dedication to updating information and contributing to the overall comprehensiveness of Wikipedia articles.

5.2 Armed Conflicts and Wars

In this section, we examine the impact of three Armed Conflicts and Wars events on English Wikipedia platform.

The first event is the Russian invasion of Ukraine[1] (Section 5.2.1), which began on February 24, 2022. Russia aimed to take over parts of Ukraine. This event was chosen due to its significant global impact and extensive media coverage. We can investigate how such a high-profile and widely covered armed conflict can affect editing behavior on Wikipedia.

The second event is the 2021 Israel-Palestine crisis[2] (Section 5.2.2). It began on May 10, 2021 and involved violent confrontations between Israel and Palestinian armed groups. This event was chosen to study an armed conflict in the Middle East that may not have received as much media coverage as the Russian invasion of Ukraine.

The third event is the Tigray War[3] (Section 5.2.3), an ongoing armed conflict that started on November 3, 2020, between the Ethiopian government and the Tigray People's Liberation Front in Ethiopia's Tigray region. This event was chosen to study an armed conflict that has received less media attention compared to the Russian invasion of Ukraine and took place in Africa.

In all three events, there was a notable surge in editing activity subsequent to the occurrences, indicating a heightened interest and involvement from editors in updating the information. However, it is worth noting that the Russian invasion of Ukraine received considerably more attention and engagement compared to the other two events. When comparing the number of edits received by the top 10 articles following each event, the Russian invasion of Ukraine accumulated a total of 15,386 edits. In contrast, the 2021 Israel-Palestine crisis received 2,922 edits, which accounts for approximately 19% of the edits seen during the Russian invasion, while the Tigray War received a mere 1,570 edits, which represents approximately 10% of the Russian invasion's editing activ-

[1] https://en.wikipedia.org/wiki/Russian_invasion_of_Ukraine
[2] https://en.wikipedia.org/wiki/2021_Israel%E2%80%93Palestine_crisis
[3] https://en.wikipedia.org/wiki/Tigray_War

ity. Editors were likely more drawn to the Russian invasion of Ukraine due to its significant global impact, extensive media coverage, and the high level of international concern it generated. Additionally, the geographical location of the conflict and the availability of information may have contributed to the higher number of edits observed.

Furthermore, the level of engagement in the six months following the events varied from one event to another. For instance, the Russian invasion of Ukraine and the Tigray War sustained a high level of engagement during the six months following their onset, indicating that these conflicts continued to generate ongoing developments that prompted editors to continuously update articles. However, in the case of the 2021 Israel-Palestine crisis, the level of engagement quickly returned to the daily average edit levels observed before the event took place.

Disagreements among editors were observed to increase, reflecting the sensitive and contentious nature of armed conflict topics. However, the percentage of reverted edits varied across different events. The Russian invasion of Ukraine and the Israel-Palestine crisis experienced a reverting rate of 0.5% and 0.7% of total edits, respectively. In contrast, the Tigray War had a significantly higher rate of 4.6% of reverted edits, indicating a greater level of disagreements and disputes among editors regarding this particular event.

Furthermore, it was noted that vandalism-related edits decreased in both the Russian invasion of Ukraine and the 2021 Israel-Palestine crisis, thanks to protective measures that were implemented, such as restrictions on anonymous editing. These measures effectively helped mitigate vandalism on the articles. Conversely, the Tigray War lacked similar protective measures, which resulted in an increase in vandalism-related edits. From this observation, we can conclude that implementing measures such as article protection from anonymous editing proves to be effective in minimizing vandalism during armed conflicts.

5.2.1 Russian invasion of Ukraine

To analyze the impact of the Russian invasion of Ukraine, which began on February 24, 2022, on the editing activity, we selected 44 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from January 27, 2022, to March 24, 2022). Table 5.3 presents the data obtained from these articles, which collectively received a total of 22,279 edits. These edits consisted of 20,386 registered edits (92% of the total) and 1,154 geolocated edits (5% of the total). Among these edits, 145 were reverted, and 19 instances of vandalism were identified.

A significant increase in editing activity is observed following the event. In the four weeks prior to the invasion, these articles received a total of 1,781

Table 5.3: Analysis of edits during the Russian invasion of Ukraine on 2022-02-24. Table shows total, registered, reverted, and vandalism-reverted edits for related articles, including geolocated versions. Covers 8 weeks (2022-01-27 - 2022-03-24), with 4 weeks before and after. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	22,279	1,781	20,498	18,717	1,051%	1.95***
Registered Edits	20,386	1,635	18,751	17,116	1,047%	1.85***
Geolocated Edits	1,154	105	1,049	944	899%	2.31***
Reverted Edits	145	34	111	77	226%	0.91**
Geolocated Reverted Edits	133	30	103	73	243%	0.92**
Vandalism Reverted Edits	19	10	9	1	-10%	0.05
Geolocated Vandalism Reverted Edits	17	8	9	1	12%	0.03

edits. However, in the four weeks after the invasion began, the number of edits increased to 20,498, accounting for 92% of all edits during the analysis period. This represents a substantial relative change of 1,051% and an absolute difference of 18,717 edits between the two periods. The effect size, as measured by $d = 1.95^{***}$, indicates a strong impact.

In terms of reverted edits, there were 34 instances before the event and 111 instances after, representing a 226% relative change. The effect size, as measured by $d = 0.91^{**}$, indicates a notable impact. This finding indicates an increment in conflicts or disagreements among editors, which is not surprising given the sensitive nature of articles related to armed conflicts. In such cases, opposing parties may repeatedly revert edits to promote their own viewpoints.

Notably, despite the expectation of an increase in instances of vandalism during such events, our analysis revealed that the number of reverted edits due to vandalism remained relatively stable, with a statistically insignificant effect size ($d = 0.05$). This stability can be attributed to the proactive measure of restricting articles from anonymous editing taken by Wikipedia editors, which likely deterred malicious activities. This is shown in 5.4, where we present the top 10 articles contributing to the total edits during the analysis. Out of these top 10 articles, five were protected, with four of them being protected during the analysis period. The main article titled 'Russian Invasion of Ukraine' garnered the majority of editing attention, accounting for 35% of these edits.

Regarding the temporal pattern of edits, Figure 5.1a shows the number of edits over time. The highest number of edits, totaling 1,991, occurred on February 25, 2022, the day following the event. Prior to the event, the highest number of edits was observed on February 22, 2022, with a total of 220 edits.

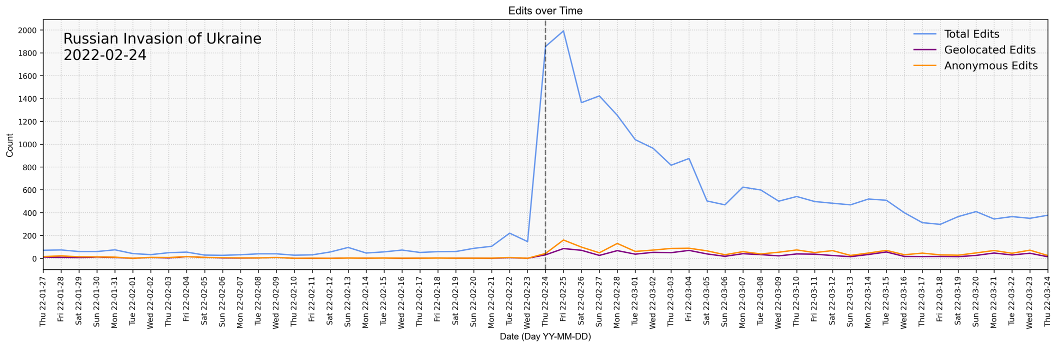
Figure 5.1b provides a broader perspective, covering a 12-month period

Table 5.4: Top 10 articles ranked by total edits during the Russian invasion of Ukraine on 2022-02-24. Covers 8 weeks (2022-01-27 - 2022-03-24), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

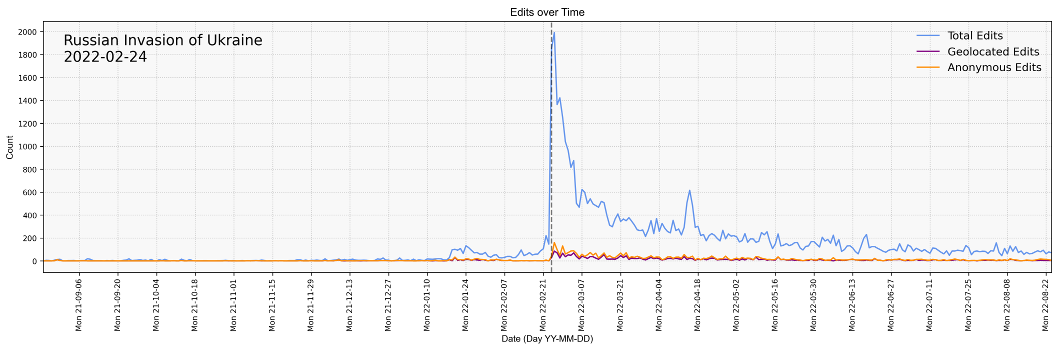
Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
Russian invasion of Ukraine	7,732	35%	6	7,726	0	2	0	0	✓	✓
Government and intergovernmental reactions to the...	2,039	9%	0	2,039	0	6	0	0	×	
Prelude to the Russian invasion of Ukraine	1,824	8%	1,384	440	18	0	3	0	×	
Ukrainian refugee crisis (2022–present)	1,094	5%	0	1,094	0	7	0	0	✓	✓
Russo-Ukrainian War	936	4%	235	701	8	0	5	0	✓	×
Anti-war protests in Russia (2022–present)	716	3%	0	716	0	2	0	0	×	
Timeline of the Russian invasion of Ukraine	705	3%	0	705	0	0	0	0	✓	✓
International Legion (Ukraine)	698	3%	0	698	0	5	0	1	×	
List of military aid to Ukraine during the Russo-Ukrainian War	635	3%	0	635	0	0	0	0	✓	✓
Order of battle for the Russian invasion of Ukraine	632	3%	0	632	0	13	0	0	×	
$\sum$	17,011	76%	1,625	15,386	26	35	8	1	5	4

from August 24, 2021, to August 24, 2022, with 6 months before and 6 months after the event. The plot reveals that the daily edit count did not return to pre-event levels for the next 6 months following the event, indicating sustained engagement. In contrast, in the 6 months leading up to the event, the daily edit count was close to zero until January 19, approximately a month prior to the event.

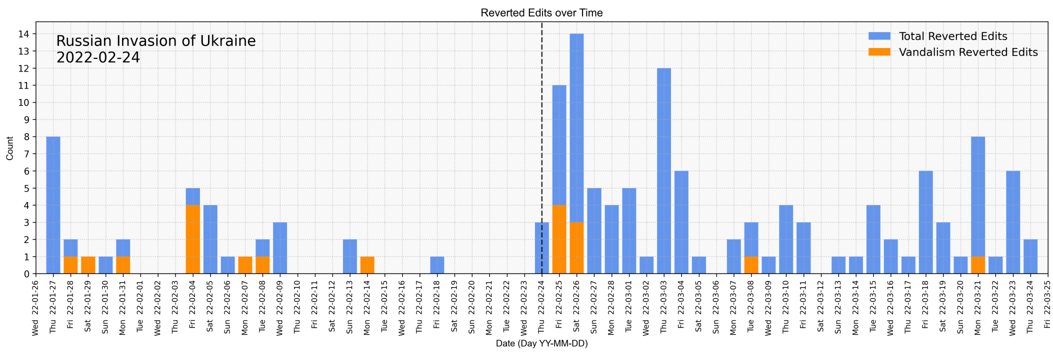
Figure 5.1c illustrates the number of reverted edits over time, showing a clear increase in daily reverted edits in the 4 weeks after the event compared to the 4 weeks before the event. The highest number of reverted edits occurred on February 26th, with 14 reverted edits.



(a) Total and anonymous edits over 8 weeks during the Russian invasion of Ukraine.



(b) Total and anonymous edits over 12 months during the Russian invasion of Ukraine.



(c) Reverted edits and reverted-vandalism edits over 8 weeks during the Russian invasion of Ukraine.

Figure 5.1: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the Russian invasion of Ukraine on 2022-02-24. Plots (a) and (c) cover an 8-week period (2022-01-27 - 2022-03-24), with 4 weeks before and after. Plot (b) covers a 12-month period (2021-08-24 - 2022-08-24), with 6 months before and after.

Table 5.5: Analysis of edits during the 2021 Israel-Palestine crisis on 2021-05-10. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2021-04-12 - 2021-06-07), with 4 weeks before and after. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	3,485	233	3,252	3,019	1,296%	1.52***
Registered Edits	3,335	205	3,130	2,925	1,427%	1.51***
Reverted Edits	34	11	23	12	109%	0.41
Vandalism Reverted Edits	5	3	2	1	-33%	0.10

5.2.2 2021 Israel–Palestine crisis

To analyze the impact of the 2021 Israel-Palestine crisis, which occurred on May 10, 2021, on the editing activity, we selected 44 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from April 12, 2021, to June 7, 2021). Table 5.5 presents the data obtained from these articles, which collectively received a total of 3,485 edits, with 3,335 of them being registered edits (96% of the total). Among these edits, 34 were reverted, and only 5 instances of vandalism were identified.

A significant increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 233 edits. However, in the four weeks after the event, the number of edits increased to 3,253 accounting for 93% of all edits during the analysis period. This represents a substantial relative change of 1,296% and an absolute difference of 3,019 edits between the two periods. The effect size, as measured by $d = 1.52^{***}$, indicates a strong impact.

In terms of reverted edits, there were 11 instances before the event and 23 instances after, representing a 109% relative change. The effect size, as measured by $d = 0.41$, indicates a small impact. However, the number of vandalism-reverted edits before and after the event remained low, showing a statistically insignificant effect size ($d = 0.1$). This outcome was expected since 9 out of the top 10 articles that contributed to the total edits during the analysis were protected from anonymous editing. It is worth mentioning that 5 out of these 9 articles were protected during the analysis period, presumably in response to Wikipedia editors' efforts to safeguard them from vandalism. These findings are presented in Table 5.6. The main article titled "2021 Israel-Palestine crisis" attracted the most editing attention, accounting for 66% of the total edits in this event analysis.

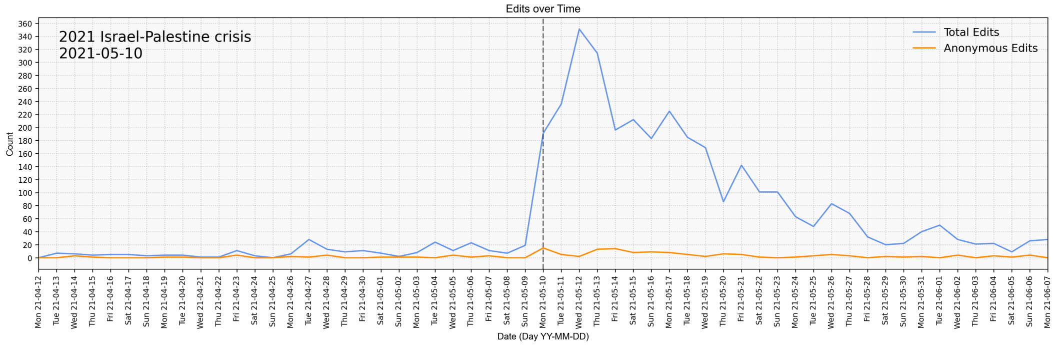
Table 5.6: Top 10 articles ranked by total edits during the 2021 Israel-Palestine crisis on 2021-05-10. Covers 8 weeks (2021-04-12 - 2021-06-07), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
2021 Israel–Palestine crisis	2,289	66%	16	2,273	0	3	0	2	✓	✓
Timeline of the Israeli–Palestinian conflict in 2021	205	6%	0	205	0	6	0	0	✓	✓
International reactions to the 2021 Israel–Palestine crisis	93	3%	0	93	0	0	0	0	✓	✓
International protests over the 2021 Israel–Palestine crisis	92	3%	0	92	0	0	0	0	✓	✓
International recognition of the State of Palestine	62	2%	21	41	3	5	0	0	✓	×
Wesley Fofana (footballer)	58	2%	25	33	4	2	2	0	×	
Mohammed el-Kurd	58	2%	0	58	0	0	0	0	✓	✓
Iron Dome	57	2%	0	57	0	0	0	0	✓	×
Israeli–Palestinian conflict	55	2%	3	52	0	0	0	0	✓	×
Israel and apartheid	54	2%	36	18	0	0	0	0	✓	×
$\sum$	3,023	87%	101	2,922	7	16	2	2	9	5

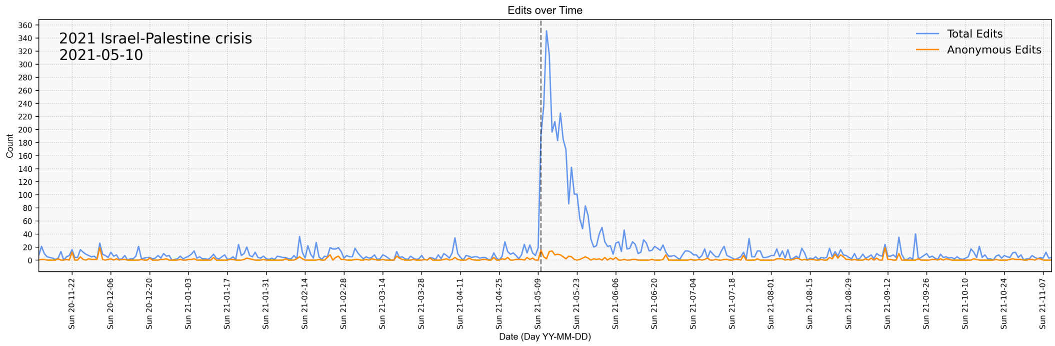
Regarding the temporal pattern of edits, Figure 5.2a shows that the highest number of edits, totaling 351, occurred on the second day following the event. Prior to the event, the average daily edits were low, with a maximum of 28 edits during the four weeks leading up to it. However, by the end of the sixth week after the event, the average daily edits were approaching their pre-event levels, with an average of fewer than 50 edits per day.

Additionally, Figure 5.2b shows that three weeks after the event, the average daily edits were very similar to the average during the six months prior to the event.

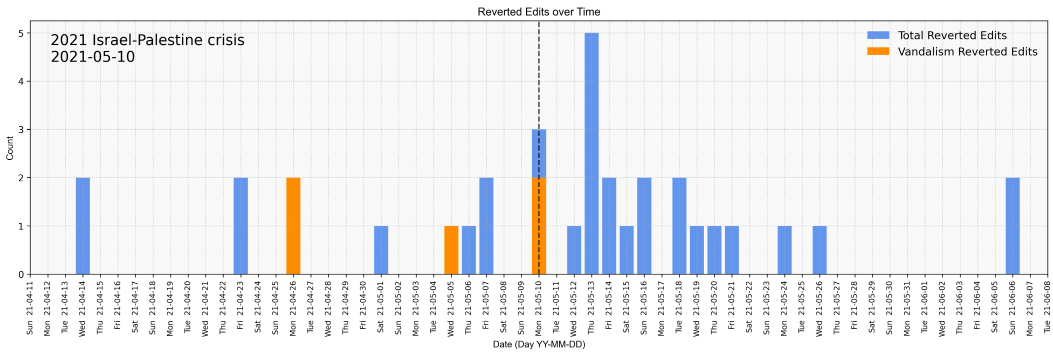
Figure 5.2c shows that the highest number of reverted edits occurred on the third day following the event, with 5 reverted edits.



(a) Total and anonymous edits over 8 weeks during the 2021 Israel-Palestine crisis.



(b) Total and anonymous edits over 12 months during the 2021 Israel-Palestine crisis.



(c) Reverted edits and reverted-vandalism edits over 8 weeks during the 2021 Israel-Palestine crisis.

Figure 5.2: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the 2021 Israel-Palestine crisis on 2021-05-10. Plots (a) and (c) cover an 8-week period (2021-04-12 - 2021-06-07), with 4 weeks before and after. Plot (b) covers a 12-month period (2020-11-10 - 2021-11-10), with 6 months before and after.

Table 5.7: Analysis of edits during the Tigray War on 2020-11-03. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2020-10-06 - 2020-12-01), with 4 weeks before and after. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	1,868	244	1,624	1,380	566%	2.42***
Registered Edits	1,416	170	1,246	1,076	633%	2.49***
Reverted Edits	87	13	74	61	469%	1.45***
Vandalism Reverted Edits	10	1	9	8	800%	0.52

5.2.3 Tigray War

To analyze the impact of the Tigray War, which began on November 3, 2020, on the editing activity, we selected 21 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from October 6, 2020, to December 1, 2020). Table 5.7 presents the data obtained from these articles, which collectively received a total of 1,868 edits, with 1,416 of them being registered edits (76% of the total). Among these edits, 87 (4.6% of the total) were reverted, and 10 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 244 edits. However, in the four weeks after the event, the number of edits increased to 1,624 accounting for 87% of all edits during the analysis period. This represents a relative change of 566% and an absolute difference of 1,380 edits between the two periods. The effect size, as measured by $d = 2.42$***, indicates a huge impact.

In terms of reverted edits, there were 13 instances before the Tigray War event and 74 instances after, representing a 469% relative change. The effect size, as measured by $d = 1.45$***, indicates a strong impact. This suggests that the Tigray War generated heated discussions, controversies, and conflicts among editors, leading to a higher rate of reverted edits. Furthermore, the number of vandalism-reverted edits also increased, with a moderate effect size ($d = 0.52$). This could be attributed to the lack of protection safeguarding from anonymous editing during the event, potentially contributing to the observed increase in vandalism-related edits. Table 5.8 shows that out of the top 10 articles contributing to this event analysis, only two were protected, and the protection was not implemented during the period of the event. Notably, the main article titled "Tigray War" attracted the most editing attention,

Table 5.8: Top 10 articles ranked by total edits during the Tigray War on 2020-11-03. Covers 8 weeks (2020-10-06 - 2020-12-01), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

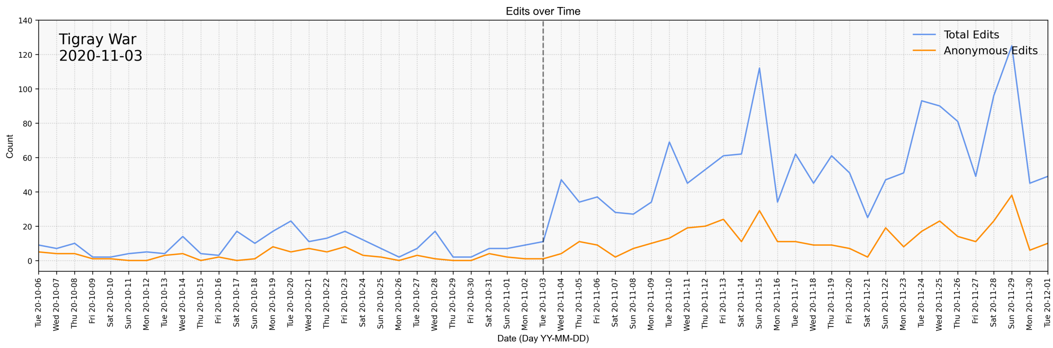
Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
Tigray War	774	41%	0	774	0	18	0	2	×	
Mai Kadra massacre	225	12%	0	225	0	3	0	0	×	
Baykar Bayraktar TB2	185	10%	129	56	6	0	0	0	✓	×
Tigray People's Liberation Front	178	10%	5	173	0	16	0	3	×	
Tigray Region	125	7%	11	114	0	6	0	0	×	
Abiy Ahmed	122	7%	33	89	2	15	1	1	✓	×
Timeline of the Tigray War	76	4%	0	76	0	0	0	0	×	
Eritrean–Ethiopian War	49	3%	11	38	0	6	0	1	×	
2022	25	1%	13	12	2	0	0	0	×	
List of war crimes	24	1%	11	13	1	4	0	2	×	
$\sum$	1,783	95%	213	1,570	11	68	1	9	2	0

accounting for 41% of the total edits in this event analysis.

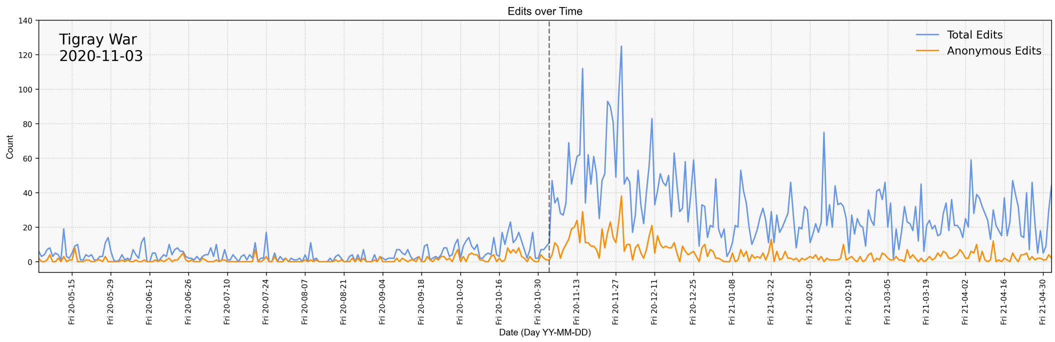
Regarding the temporal pattern of edits, Figure 5.3a depicts the highest number of edits occurring on November 29, 2020, which is 26 days after the initiation of the event. This notable surge in editing activity, amounting to 125 edits, can be attributed to a significant development during that time: the capture of Mekelle, the capital of the Tigray Region, by Federal allied forces on November 28, 2020. This particular event likely sparked increased interest and discussions, motivating editors to actively update and revise articles related to the Tigray War.

Moreover, Figure 5.3b displays multiple spikes in edits during the six months following the event. This observation indicates that the conflict continued to generate ongoing developments, news, and discussions, prompting editors to continuously update and revise articles concerning the Tigray War.

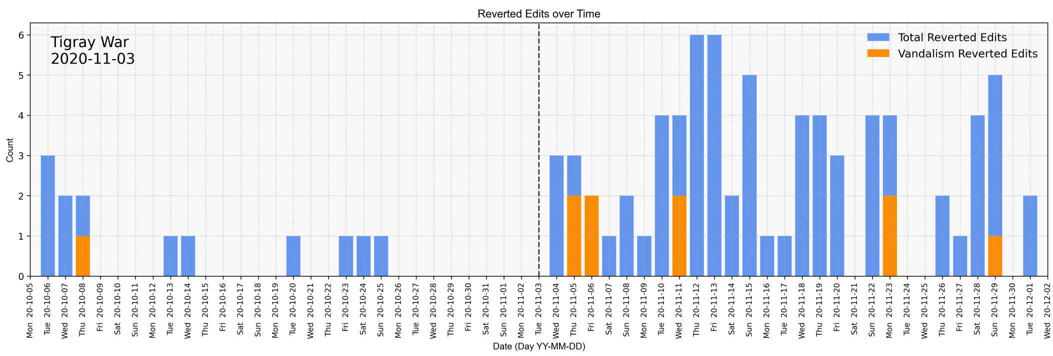
Figure 5.3c illustrates the increase in daily reverted edits following the event, peaking at six reverted edits.



(a) Number of total edits, and anonymous edits over an 8-week.



(b) Number of total edits, and anonymous edits over 12 months.



(c) Number of reverted edits, including reverted-vandalism edits over an 8-week.

Figure 5.3: Number of edits, reverted edits, including anonymous edits and vandalism-reverted edits during the Russian invasion of Ukraine on 2022-02-24. Plot (a) and (c) cover 8 weeks (2022-01-27 - 2022-03-24), with 4 weeks before and after. Plot (b) covers a 12-month period (2021-08-24 - 2022-08-24), with 6 months before and after.

5.3 Elections

In this section, we examine the impact of three Election events on English Wikipedia platform.

The first event is the 2020 United States presidential election[4] (Section 5.3.1), which began on November 3rd, 2020, and concluded on November 7th, 2020. Democrat Joe Biden defeated Republican Donald Trump to become the 46th President of the United States. This event was chosen because it is one of the most significant elections globally and received extensive media coverage. We can investigate how such a high-profile and widely covered election can affect the behavior and engagement of Wikipedia editors.

The second event is the 2021 German federal election, 2021.[5] (Section 5.3.2). which took place on September 26th, 2021, to elect the members of the 20th Bundestag. This event was chosen because Germany is a member of the European Union, which makes its federal elections relevant to all European countries. Additionally, since English is not the primary language in Germany, these elections may not have received as much media coverage as the 2020 United States presidential election in the English Wikipedia.

The third event is the 2018 Bangladeshi general election[6] (Section 5.3.3). It occurred on December 30, 2018, and led to a clear victory for the ruling Awami League and its allies, ensuring Sheikh Hasina's re-election as Prime Minister. This event was chosen because of its controversial nature, as concerns were raised about its fairness and transparency. However, it received less attention and engagement compared to the first two election events. Additionally, it was held in Asia where English is not the primary language.

In all three events, there was a significant surge in editing activity following the occurrences, indicating heightened interest and involvement from editors in updating the information. However, it is noteworthy that the 2020 United States presidential election received substantially more attention and engagement compared to the other two events. When comparing the number of edits received by the top 10 articles following each event, the 2020 United States presidential election accumulated a total of 1,885 edits. In contrast, the 2021 German federal election received 1,028 edits, accounting for approximately 54% of the edits of the 2020 United States presidential election. The 2018 Bangladeshi general election received only 217 edits, representing approximately 12% of the edits of the 2020 United States presidential election. The extensive international media coverage and the fact that English is the primary language in the United States likely contributed to the higher engagement and

[4] `https://en.wikipedia.org/wiki/2020_United_States_presidential_election`
[5] `https://en.wikipedia.org/wiki/2021_German_federal_election`
[6] `https://en.wikipedia.org/wiki/2018_Bangladeshi_general_election`

attention.

Furthermore, there was a noticeable buildup of daily edits for the 2020 United States presidential election, and the level of engagement remained high in the six months following the event. In contrast, the 2021 German federal election and the 2018 Bangladeshi general election had minimal buildup, and the level of engagement quickly returned to the average daily edit levels observed before the events took place. This indicates that the 2020 United States presidential election generated ongoing developments that prompted editors to continuously update the related articles, while the other two events did not.

Regarding reverted edits, there was an increase following the 2020 United States presidential election. However, for the 2021 German federal election and the 2018 Bangladeshi general election, including vandalism-reverted edits, the number remained similar, indicating a greater level of disagreements and disputes among editors specifically regarding the 2020 United States presidential election. This difference in reverted edits could be attributed to the higher level of international attention received by the 2020 United States presidential election. It is also worth noting that we are examining the English Wikipedia platform, where English is the primary language in the United States, while Germany and Bangladesh have different primary languages.

Furthermore, it was observed that there were no restrictions on anonymous editing for election-relevant articles. Only one article among the top 30 articles contributing to these three events was protected, and it was protected not during the event itself. This suggests that these articles do not attract a significant amount of vandalism, and editors do not perceive them as vulnerable to such vandalism.

5.3.1 2020 United States presidential election

To analyze the impact of the 2020 United States presidential election, which began on November 3, 2020, on the editing activity, we selected 55 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from October 6, 2020, to December 1, 2020). Table 5.9 presents the data obtained from these articles, which collectively received a total of 9,613 edits, with 8,400 of them being registered edits (87% of the total). Among these edits, 103 were reverted, and only 8 instances of vandalism were identified.

A significant increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 3,482 edits. However, in the four weeks after the event, the number of edits increased to 6,131. This represents a relative change of 76% and an absolute difference of 2,649 edits between the two periods. The effect size, as measured by $d =$

Table 5.9: Analysis of edits during the 2020 United States presidential election on 2020-11-03. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2020-10-06 - 2020-12-01), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's *d*, with one to three asterisks (*) indicating *p*-values less or equal to 0.05, 0.01, and 0.001.

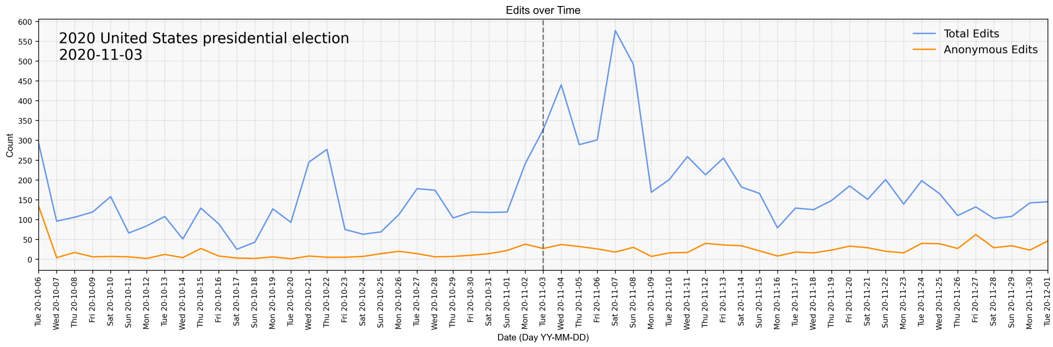
Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	9,613	3,482	6,131	2,649	76%	0.89**
Registered Edits	8,400	3,073	5,327	2,254	73%	0.78**
Reverted Edits	103	22	81	59	268%	0.73**
Vandalism Reverted Edits	8	3	5	2	67%	0.15

0.89**, indicates a large impact.

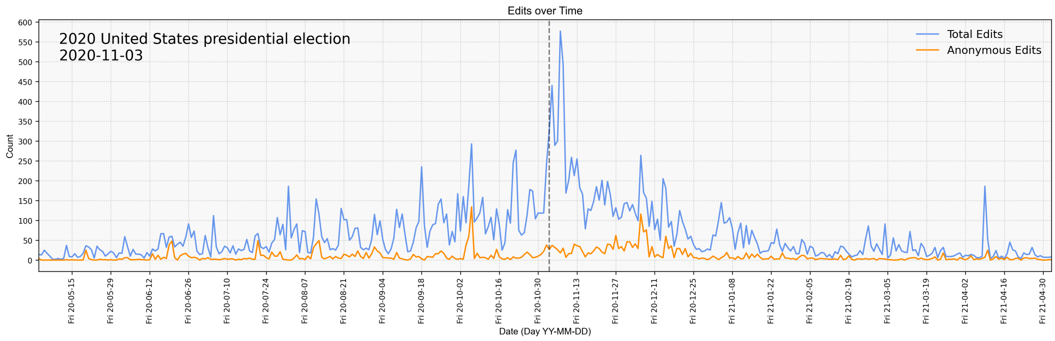
In terms of reverted edits, there were 22 instances before the event and 81 instances after, representing a 73% relative change. The effect size, as measured by $d = 0.73$**, indicates a moderate impact. The number of vandalism-reverted edits increased following the event; however, the absolute difference before and after is low, with a statistically insignificant effect size ($d = 0.15$). It was expected to have a higher impact on vandalism-reverted edits as 9 out of the top 10 articles that contributed to the total edits during the analysis were not protected from anonymous editing (see Table 5.10). The main article titled "2020 United States presidential election" attracted the most editing attention, accounting for 23% of the total edits in this event analysis, while the remaining 9 articles had relatively balanced contributions ranging from 2% to 5%.

Regarding the temporal pattern of edits, Figure 5.4a illustrates that the highest number of edits, totaling 577, occurred on November 7, 2020, which coincided with the declaration of the election winner. However, the plot demonstrates that the daily edits before and after the event were comparable. Furthermore, Figure 5.4b displays a buildup of spikes leading up to the event, indicating significant activity. After the event, it took approximately 10 weeks for daily edits to decline, indicating sustained engagement and continued attention to this event.

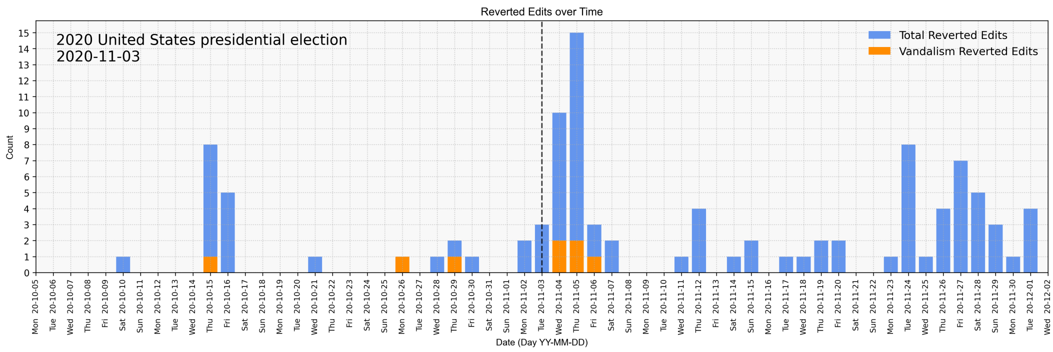
Figure 5.4c, the graph reveals an increase in daily reverted edits, with the highest number of 15 occurring on the third day following the event.



(a) Total and anonymous edits over 8 weeks during the 2020 United States presidential election.



(b) Total and anonymous edits over 12 months during the 2020 United States presidential election.



(c) Reverted edits and reverted-vandalism edits over 8 weeks during the 2020 United States presidential election.

Figure 5.4: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the 2020 United States presidential election on 2020-11-03. Plots (a) and (c) cover an 8-week period (2020-10-06 - 2020-12-01), with 4 weeks before and after. Plot (b) covers a 12-month period (2020-05-03 - 2021-05-03), with 6 months before and after.

Table 5.10: Top 10 articles ranked by total edits during the 2020 United States presidential election on 2020-11-03. Covers 8 weeks (2020-10-06 - 2020-12-01), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
2020 United States presidential election	2,195	23%	379	1,816	0	0	0	0	✓	×
2020 United States presidential election in Georgia	442	5%	157	285	1	4	0	0	×	
2020 United States presidential election in Pennsylvania	438	5%	223	215	0	7	0	0	×	
2020 United States presidential election in Arizona	365	4%	164	201	1	3	0	0	×	
2020 United States presidential election in Wisconsin	364	4%	243	121	0	2	0	0	×	
2020 United States presidential election in Florida	348	4%	186	162	0	3	0	1	×	
2020 United States presidential election in Michigan	344	4%	206	138	0	1	0	0	×	
2020 United States presidential election in North Carolina	280	3%	166	114	0	3	0	0	×	
2020 United States presidential election in Texas	277	3%	161	116	2	2	0	0	×	
Attempts to overturn the 2020 United States presidential election	237	2%	0	237	0	5	0	0	×	
$\sum$	5,290	55%	1,885	3,405	4	30	0	1	1	0

Table 5.11: Analysis of edits during the 2021 German federal election on 2021-09-26. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2021-08-29 - 2021-10-24), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	←	→	Abs.	Rel.	d
Total Edits	1,321	293	1,028	735	251%	0.59*
Registered Edits	1,191	200	991	791	396%	0.65*
Reverted Edits	19	9	10	1	11%	0.03
Vandalism Reverted Edits	2	1	1	0	0%	0.01

5.3.2 2021 German federal election

To analyze the impact of the 2021 German federal election, which began on September 26, 2021, on the editing activity, we selected 10 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from August 29, 2021, to October 24, 2021). Table 5.11 presents the data obtained from these articles, which collectively received a total of 1,321 edits, with 1,191 of them being registered edits (90% of the total). Among these edits, 19 were reverted, and only 2 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 293 edits. However,

Table 5.12: Top 10 articles ranked by total edits during the 2021 German federal election on 2021-09-26. Covers 8 weeks (2021-08-29 - 2021-10-24), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
2021 German federal election	725	55%	96	629	3	8	0	1	×	
Results of the 2021 German federal election	173	13%	2	171	0	0	0	0	×	
Opinion polling for the 2021 German federal election	162	12%	141	21	4	0	0	0	×	
Next German federal election	66	5%	0	66	0	0	0	0	×	
2017 German federal election	61	5%	15	46	1	0	0	0	×	
1994 German federal election	35	3%	11	24	0	0	0	0	×	
2009 German federal election	32	2%	8	24	0	0	0	0	×	
2005 German federal election	31	2%	11	20	0	0	0	0	×	
Candidates of the 2021 German federal election	25	2%	5	20	0	0	0	0	×	
Elections in Germany	11	1%	4	7	1	2	1	0	×	
$\sum$	1,321	100%	293	1,028	9	10	1	1	0	0

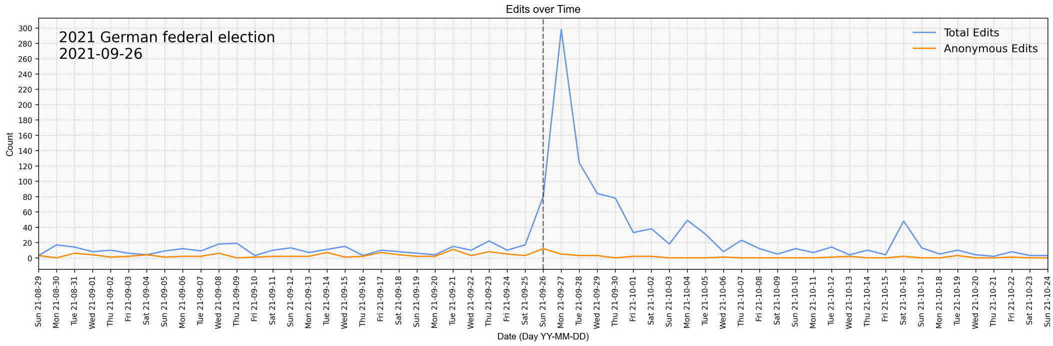
in the four weeks after the event, the number of edits increased to 1,028. This represents a relative change of 251% and an absolute difference of 735 edits between the two periods. The effect size, as measured by $d = 0.59$*, indicates a moderate impact.

In terms of reverted edits, there was minimal change before and after the event. Prior to the event, there were 9 instances, and after the event, there were 10 instances, representing a relative change of 11%. The effect size, as measured by $d = 0.03$, indicates an insignificant impact. This suggests that there were relatively limited discussions and controversies among editors following the event. Similarly, the number of vandalism-reverted edits remained the same, with only one instance before and one after the event, and a statistically insignificant effect size ($d = 0.01$). Table 5.12 shows that none of the top 10 articles contributing to this event analysis were protected from anonymous editing, suggesting that editors did not perceive them as vulnerable to vandalism. The main article titled "2021 German federal election" attracted the most editing attention, accounting for 55% of the total edits in this event analysis.

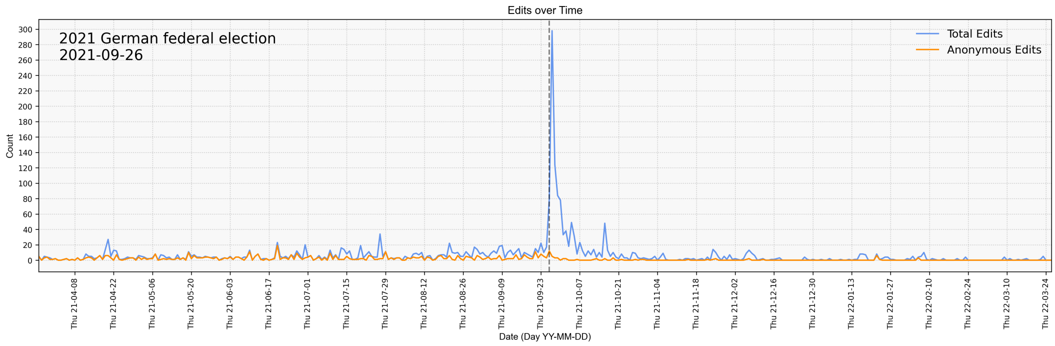
Regarding the temporal pattern of edits, Figure 5.5a demonstrates that the highest number of edits, totaling 298, occurred on the first day following the event. Subsequently, the daily edits rapidly returned to normal pre-event levels.

In Figure 5.5b, it can be observed that approximately three weeks after the event, the average number of edits returned to the pre-event levels, typically ranging between 0 and 30 edits per day.

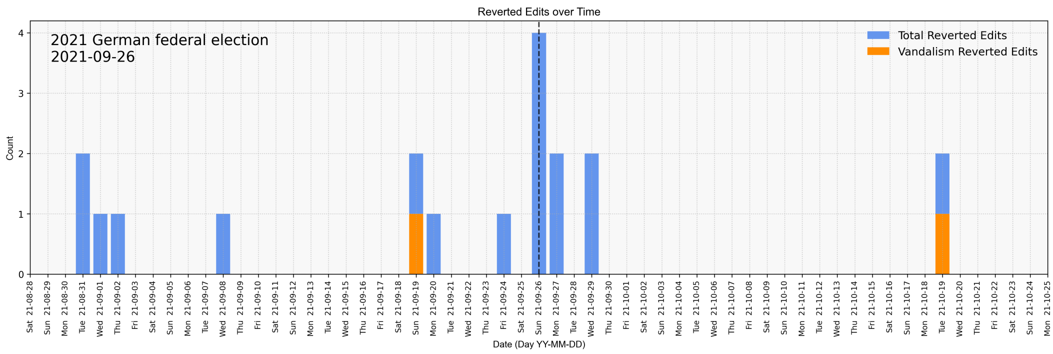
Figure 5.5c illustrates that out of the 28-day analysis period, only 11 days had reverted edits. The day of the event witnessed the highest number of reverted edits, with a total of 4 instances.



(a) Total and anonymous edits over 8 weeks during the 2021 German federal election.



(b) Total and anonymous edits over 12 months during the 2021 German federal election.



(c) Reverted edits and reverted-vandalism edits over 8 weeks during the 2021 German federal election.

Figure 5.5: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the 2021 German federal election on 2021-09-26. Plots (a) and (c) cover an 8-week period (2021-08-29 - 2021-10-24), with 4 weeks before and after. Plot (b) covers a 12-month period (2021-03-26 - 2022-03-26), with 6 months before and after.

Table 5.13: Analysis of edits during the 2018 Bangladeshi general election on 2018-12-30. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2018-12-02 - 2019-01-27), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's *d*, with one to three asterisks (*) indicating *p*-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	350	133	217	84	63%	0.20
Registered Edits	264	97	167	70	72%	0.24
Reverted Edits	2	1	1	0	0%	0
Vandalism Reverted Edits	0	0	0	0		

5.3.3 2018 Bangladeshi general election

To analyze the impact of the 2018 Bangladeshi general election, which occurred on December 30, 2018, on the editing activity, we selected 10 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from December 02, 2018, to January 27, 2019). Table 5.13 presents the data obtained from these articles, which collectively received a total of 350 edits, with 264 of them being registered edits (75% of the total). Among these edits, 2 were reverted, and 0 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 133 edits. However, in the four weeks after the event, the number of edits increased to 217. This represents a relative change of 63% and an absolute difference of 84 edits between the two periods. The effect size, as measured by $d = 0.2$, indicates a small impact.

In terms of reverted edits, there was no change before and after the event. Prior to the event, there was only 1 instance, and after the event, there was also 1 instance, indicating no significant impact. This suggests that there were minimal controversies among editors following the event. Additionally, there were no vandalism-reverted edits during the period of analysis. One would anticipate that an event as controversial as this, which raises concerns about fairness, would lead to disputes and disagreements among editors. However, the absence of reverted edits suggests that the event did not receive substantial global attention and was primarily of local interest. Given that Bangladesh and its neighboring countries do not have English as the primary language, even local interest might not have generated a significant number of edits on the English Wikipedia platform.

Table 5.14: Top 10 articles ranked by total edits during the 2018 Bangladeshi general election on 2018-12-30. Covers 8 weeks (2018-12-02 - 2019-01-27), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

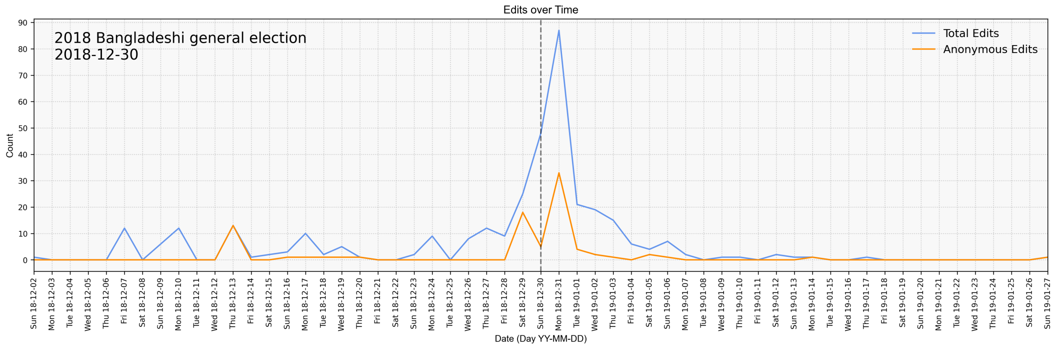
Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
2018 Bangladeshi general election	254	73%	66	188	1	1	0	0	×	
2001 Bangladeshi general election	20	6%	17	3	0	0	0	0	×	
2014 Bangladeshi general election	19	5%	16	3	0	0	0	0	×	
June 1996 Bangladeshi general election	14	4%	8	6	0	0	0	0	×	
1991 Bangladeshi general election	10	3%	4	6	0	0	0	0	×	
February 1996 Bangladeshi general election	9	3%	7	2	0	0	0	0	×	
2008 Bangladeshi general election	9	3%	4	5	0	0	0	0	×	
Elections in Bangladesh	8	2%	5	3	0	0	0	0	×	
1988 Bangladeshi general election	7	2%	6	1	0	0	0	0	×	
2024 Bangladeshi general election	0	0%	0	0	0	0	0	0	×	
$\sum$	350	100%	133	217	1	1	0	0	0	0

Table 5.14 reveals that none of the top 10 articles contributing to this event analysis were protected from anonymous editing, suggesting that editors did not perceive them as vulnerable to vandalism. The main article titled "2018 Bangladeshi general election" attracted the most editing attention, accounting for 73% of the total edits in this event analysis.

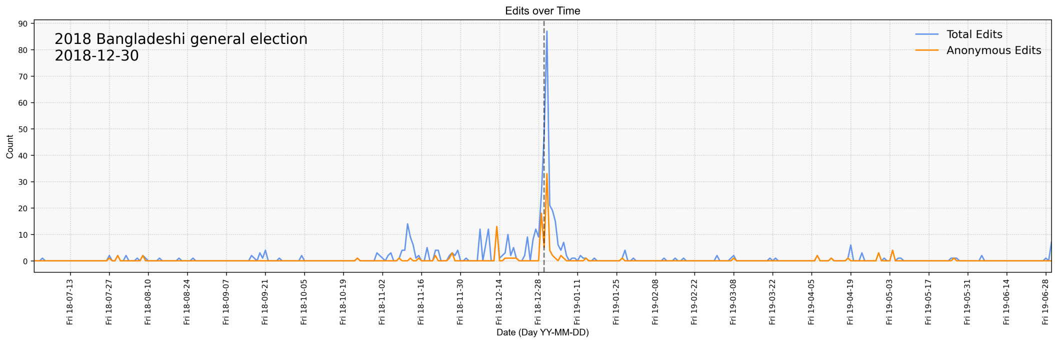
Regarding the temporal pattern of edits, Figure 5.6a demonstrates that the highest number of edits, totaling 87, occurred on the first day following the event. Subsequently, the daily edits rapidly returned to normal pre-event levels.

In Figure 5.6b, it can be observed that less than two weeks after the event, the average number of edits returned to the pre-event levels, typically ranging between 0 and 10 edits per day.

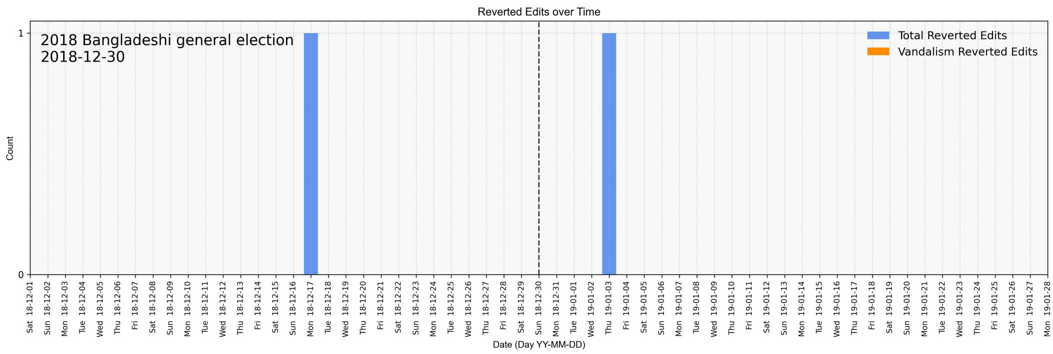
Figure 5.6c displays the occurrence of 2 instances of reverted edits, with one instance before the event and one instance after.



(a) Total and anonymous edits over 8 weeks during the 2018 Bangladeshi general election.



(b) Total and anonymous edits over 12 months during the 2018 Bangladeshi general election.



(c) Reverted edits and reverted-vandalism edits over 8 weeks during the 2018 Bangladeshi general election.

Figure 5.6: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the 2018 Bangladeshi general election on 2018-12-30. Plots (a) and (c) cover an 8-week period (2018-12-02 - 2019-01-27), with 4 weeks before and after. Plot (b) covers a 12-month period (2018-06-30 - 2019-06-30), with 6 months before and after.

5.4 Natural Disaster

In this section, we analyze the impact of three natural disaster events on the English Wikipedia platform.

The first event is Hurricane Harvey[7] (Section 5.4.1), which made landfall in Texas, United States, on August 25, 2017, causing extensive damage to infrastructure, displacing thousands of people, and resulting in loss of life. This event was chosen because it is considered one of the most destructive hurricanes in U.S. history and received a lot of media coverage.

The second event is the 2018 Sulawesi earthquake and tsunami[8] (Section 5.4.2). It struck the Indonesian island of Sulawesi on September 28, 2018, with a magnitude of 7.5. This earthquake caused a significant loss of life, displacement of communities, and widespread destruction of buildings and infrastructure. This event was chosen because it represents a different type of natural disaster that received extensive media coverage and occurred outside the U.S.

The third event is the 2018 Kerala floods[9] (Section 5.4.3), a catastrophic flooding event that occurred in the Indian state of Kerala on August 16, 2018. The floods claimed numerous lives, displaced millions of people, and caused extensive damage to infrastructure. This event was chosen because it represented a different type of natural disaster that occurred in India, where it might have received comparatively less media coverage than the first two events.

The analysis of these events reveals that all of them experienced a surge in editing activity following the events. However, the magnitude of this surge varied among the events. The top 10 articles related to Hurricane Harvey received nearly 7.5 times more edits compared to the 2018 Sulawesi earthquake and tsunami, and approximately 6 times more edits than the 2018 Kerala floods. This difference could be attributed to Hurricane Harvey's occurrence in the United States ([García-Gavilanes et al., 2016]), which garnered significant media attention. As we are examining the English Wikipedia and English being the native language of the US, there may be a correlation there. Reverted edits also increased noticeably, indicating a rise in disputes related to the articles. However, the effect size varied among the events, with Hurricane Harvey having the highest effect size. The effect size for the 2018 Kerala floods could not be calculated as there were no reverted edits in the pre-event analysis period. Instances of vandalism were low for all three events. Consequently, no measures of anonymous editing protection were employed by Wikipedia editors. Only one article out of the 30 top articles contributing to the analysis of these events

[7] `https://en.wikipedia.org/wiki/Hurricane_Harvey`
[8] `https://en.wikipedia.org/wiki/2018_Sulawesi_earthquake_and_tsunami`
[9] `https://en.wikipedia.org/wiki/2018_Kerala_floods`

Table 5.15: Analysis of edits during the Hurricane Harvey on 2017-08-25. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2017-07-28 - 2017-09-22), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	5,397	768	4,629	3,861	503%	1.83***
Registered Edits	4,613	619	3,994	3,375	545%	2.19***
Reverted Edits	189	28	161	133	475%	0.95***
Vandalism Reverted Edits	16	4	12	8	200%	0.42

was protected.

Regarding the temporal pattern of edits, similarities were observed across the events, with peaks occurring on the day following each event, reflecting the immediate impact of these events. However, the duration of heightened editing activity differed between the events. Hurricane Harvey sustained elevated daily edits for 15 weeks, indicating sustained interest and ongoing updates, while the 2018 Sulawesi earthquake and tsunami and the 2018 Kerala floods returned to pre-event edit levels within three weeks.

5.4.1 Hurricane Harvey

To analyze the impact of Hurricane Harvey, which occurred on August 25, 2017, on the editing activity, we selected 20 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from July 28, 2017, to September 22, 2017). Table 5.15 presents the data obtained from these articles, which collectively received a total of 5,397 edits, with 264 of them being registered edits (85% of the total). Among these edits, 189 were reverted, and 16 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 768 edits. However, in the four weeks after the event, the number of edits increased to 4,629. This represents a relative change of 503% and an absolute difference of 3,861 edits between the two periods. The effect size, as measured by $d = 1.83^{***}$, indicates a strong impact.

In terms of reverted edits, there were 28 instances before the event and 161 instances after, representing a 475% relative change. The effect size, as measured by $d = 0.95^{***}$, indicates a large impact. Furthermore, the number of vandalism-reverted edits also increased, with a moderate effect size ($d =$

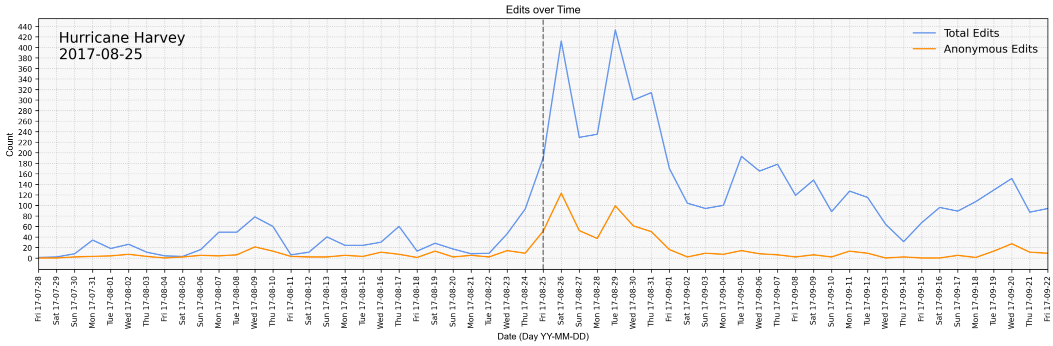
Table 5.16: Top 10 articles ranked by total edits during the Hurricane Harvey on 2017-08-25. Covers 8 weeks (2017-07-28 - 2017-09-22), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
2017 Atlantic hurricane season	2,885	53%	632	2,253	22	82	2	1	×	
Hurricane Harvey	1,432	27%	29	1,403	0	36	0	4	✓	✓
Joel Osteen	313	6%	10	303	1	23	1	6	×	
Houston	138	3%	20	118	3	7	0	0	×	
Timeline of the 2017 Atlantic hurricane season	102	2%	42	60	0	2	0	0	×	
Lakewood Church	83	2%	0	83	0	7	0	1	×	
Cajun Navy	71	1%	0	71	0	1	0	0	×	
Tropical cyclone	67	1%	8	59	0	0	0	0	✓	×
List of Texas hurricanes (1980–present)	56	1%	3	53	0	0	0	0	×	
Rockport, Texas	53	1%	0	53	0	0	0	0	×	
$\sum$	5,200	96%	744	4,456	26	158	3	12	2	1

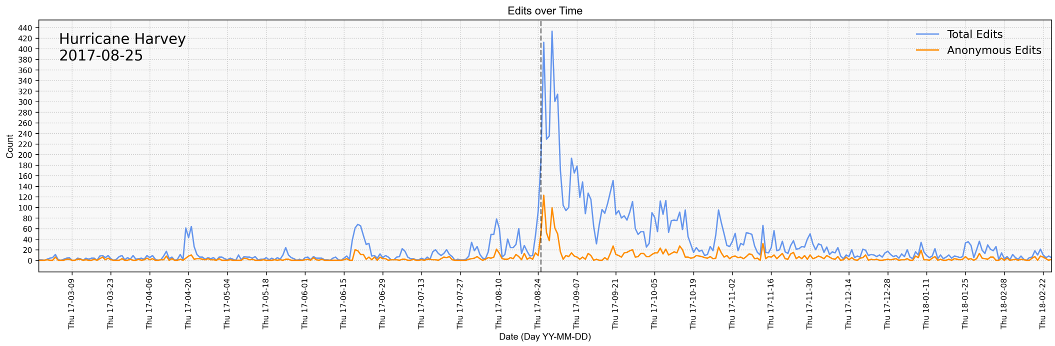
0.42).

Examining the top 10 articles ranked by total edits during Hurricane Harvey (Table 5.16), we found that the "2017 Atlantic hurricane season" article attracted the most editing attention, contributing 53% of the total edits for this event analysis. The "Hurricane Harvey" article itself received 27% of the total edits and was the only article that was protected from anonymous editing during the analysis period, likely to prevent vandalism.

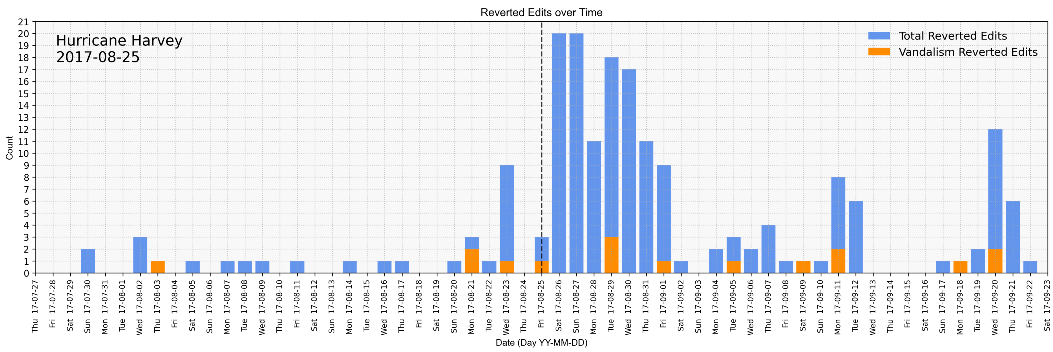
Regarding the temporal pattern of edits, Figure 5.7a demonstrates that there was a spike in daily edits on the day following the event, with 412 edits. However, a peak of 433 edits occurred four days after the start of the hurricane (August 29, 2017), coinciding with a significant development of the hurricane as it made its final landfall in Louisiana. Figure 5.7b indicates a decrease in average daily edits after the event; however, even after 15 weeks, the average daily edits remained higher than the pre-event levels, indicating sustained engagement for 15 weeks following the event. Additionally, Figure 5.7c highlights an increase in average daily reverted edits following the event, with a peak of 20 reverted edits on the day after the event and another peak of 20 reverted edits on the second day after the event. Notably, the week following the event contributed the most to the total number of reverted edits.



(a) Total and anonymous edits over an 8 weeks analysis period.



(b) Total and anonymous edits over a 12 months analysis period.



(c) Reverted edits and reverted-vandalism edits over an 8 weeks analysis period.

Figure 5.7: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the Hurricane Harvey on 2017-08-25. Plots (a) and (c) cover an 8-week period (2017-07-28 - 2017-09-22), with 4 weeks before and after. Plot (b) covers a 12-month period (2017-02-25 - 2018-02-25), with 6 months before and after.

Table 5.17: Analysis of edits during the 2018 Sulawesi earthquake and tsunami on 2018-09-28. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2018-08-31 - 2018-10-26), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	699	16	683	667	4,169%	0.85**
Registered Edits	592	13	579	566	4,354%	0.81**
Reverted Edits	15	1	14	13	1,300%	0.55*
Vandalism Reverted Edits	3	1	2	1	100%	0.15

5.4.2 2018 Sulawesi earthquake and tsunami

To analyze the impact of the 2018 Sulawesi earthquake and tsunami, which occurred on September 28, 2018, on the editing activity, we selected 10 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from August 31, 2018, to October 26, 2018). Table 5.17 presents the data obtained from these articles, which collectively received a total of 699 edits, with 592 of them being registered edits (85% of the total). Among these edits, 15 were reverted, and only 3 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 16 edits. However, in the four weeks after the event, the number of edits increased to 683. This represents a relative change of 4,169% and an absolute difference of 667 edits between the two periods. The effect size, as measured by $d = 0.85$**, indicates a large impact.

In terms of reverted edits, there was only 1 instance before the event and 14 instances after, representing a 1,300% relative change. The effect size, as measured by $d = 0.55$*, indicates a moderate impact. However, the number of vandalism-reverted edits remained relatively low, with a minimal effect size of $d = 0.15$.

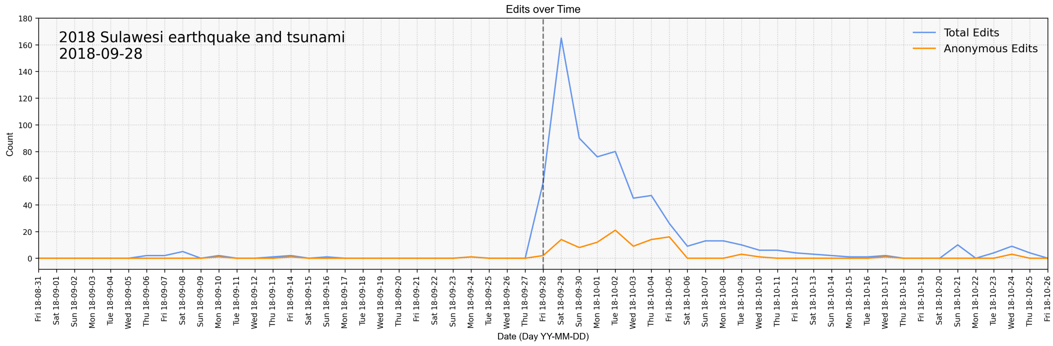
Examining the top 10 articles ranked by total edits during the 2018 Sulawesi earthquake and tsunami (Table 5.18), we found that the article specifically about the event itself attracted the most editing attention, contributing 82% of the total edits for this event analysis. It is noteworthy that no articles were protected from anonymous editing, indicating that editors did not perceive them as vulnerable to vandalism.

Regarding the temporal pattern of edits, Figure 5.8a demonstrates that

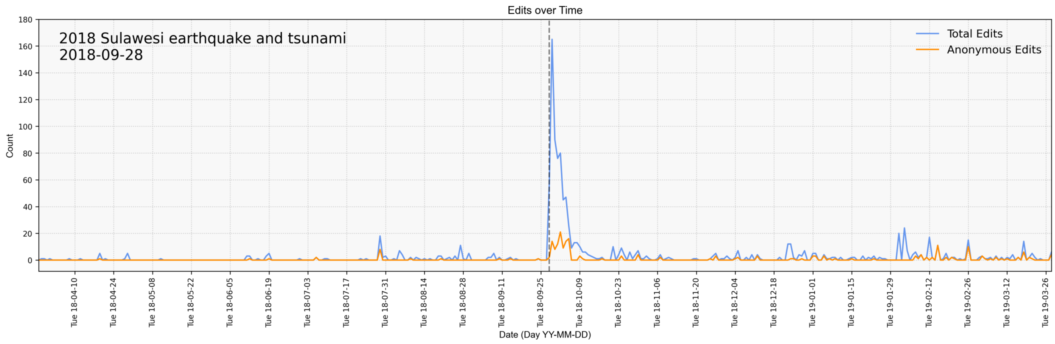
Table 5.18: Top 10 articles ranked by total edits during the 2018 Sulawesi earthquake and tsunami on 2018-09-28. Covers 8 weeks (2018-08-31 - 2018-10-26), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
2018 Sulawesi earthquake and tsunami	571	82%	0	571	0	8	0	1	×	
Lists of 21st-century earthquakes	49	7%	12	37	0	1	0	0	×	
List of tsunamis	27	4%	4	23	1	2	1	0	×	
2018 Indonesia earthquake	24	3%	0	24	0	1	0	0	×	
Operation Samudra Maitri	10	1%	0	10	0	0	0	0	×	
List of tsunamis affecting Indonesia	9	1%	0	9	0	2	0	1	×	
2018 Indonesian tsunami	4	1%	0	4	0	0	0	0	×	
2018 Indonesia tsunami	4	1%	0	4	0	0	0	0	×	
Catholic Relief Services	1	0%	0	1	0	0	0	0	×	
Floating Mosque of Palu	0	0%	0	0	0	0	0	0	×	
$\sum$	699	100%	16	683	1	14	1	2	0	0

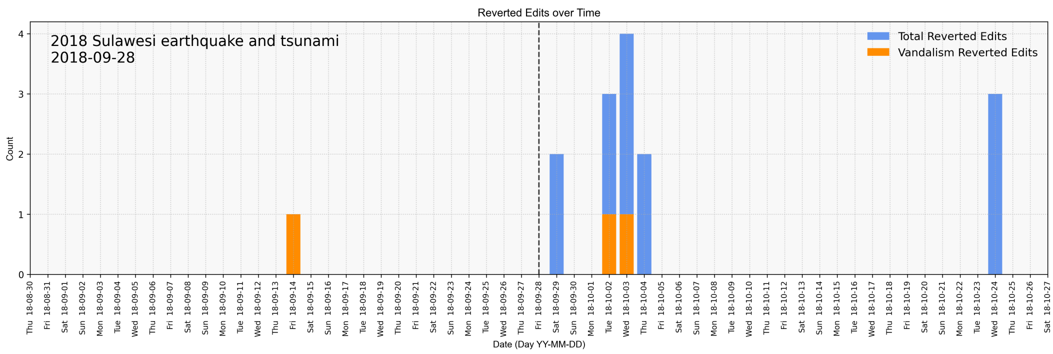
edits peaked on the day following the event, reaching a maximum of 165 edits. Figure 5.8b shows that within three weeks after the event, the average daily edits returned to their pre-event level. Figure 5.8c shows the daily reverted edits, which peaked with 4 reverted edits within a week of the event. Additionally, there were not many reverted edits or instances of vandalism after the event.



(a) Total and anonymous edits over an 8 weeks analysis period.



(b) Total and anonymous edits over a 12 months analysis period.



(c) Reverted edits and reverted-vandalism edits over an 8 weeks analysis period.

Figure 5.8: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the 2018 Sulawesi earthquake and tsunami on 2018-09-28. Plots (a) and (c) cover an 8-week period (2018-08-31 - 2018-10-26), with 4 weeks before and after. Plot (b) covers a 12-month period (2018-03-28 - 2019-03-28), with 6 months before and after.

Table 5.19: Analysis of edits during the 2018 Kerala floods on 2018-08-16. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2018-07-19 - 2018-09-13), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	851	95	756	661	696%	1.36***
Registered Edits	639	84	555	471	561%	1.16***
Reverted Edits	43	0	43	43		
Vandalism Reverted Edits	8	0	8	8		

5.4.3 2018 Kerala floods

To analyze the impact of the 2018 Kerala floods, which occurred on August 16, 2018, on the editing activity, we selected 10 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from July 19, 2018, to September 13, 2018). Table 5.19 presents the data obtained from these articles, which collectively received a total of 851 edits, with 639 of them being registered edits (75% of the total). Among these edits, 43 were reverted, and 8 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 95 edits. However, in the four weeks after the event, the number of edits increased to 756. This represents a relative change of 696% and an absolute difference of 661 edits between the two periods. The effect size, as measured by $d = 1.36^{***}$, indicates a large impact.

In terms of reverted edits, an increment can be observed as there were 43 instances recorded after the event, including 8 instances of vandalism-reverted edits. However, no reverted or vandalism-reverted edits were recorded before the event, thus making it impossible to compute the effect size for that period.

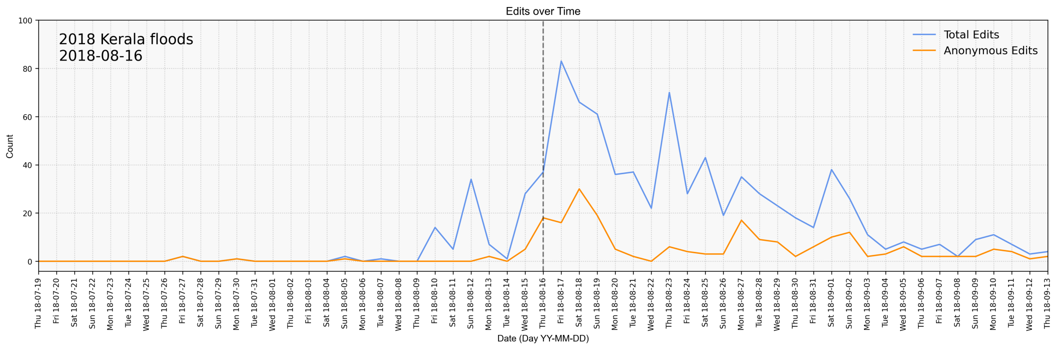
Examining the top 10 articles ranked by total edits during the 2018 Kerala floods (Table 5.20), we found that the article specifically about the event itself attracted the most editing attention, contributing 86% of the total edits for this event analysis. None of the articles were protected from anonymous editing, indicating that editors did not perceive them as vulnerable to vandalism.

Regarding the temporal pattern of edits, Figure 5.9a demonstrates that edits peaked on the day following the event, reaching a maximum of 87 edits. Another spike occurred after one week from the beginning of the event, possibly due to further developments related to the event. Figure 5.9b shows that after

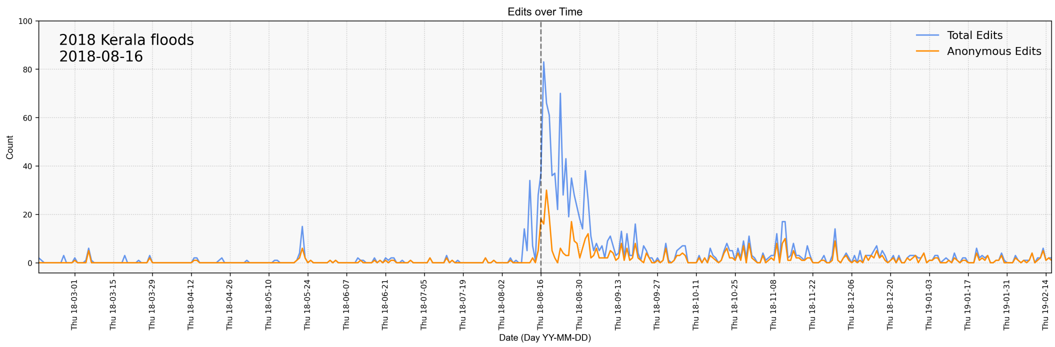
Table 5.20: Articles ranked by total edits during the 2018 Kerala floods on 2018-08-16. Covers 8 weeks (2018-07-19 - 2018-09-13), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\rightarrow$	$\leftarrow$	$\rightarrow$	🔒	@
2018 Kerala floods	733	86%	84	649	0	31	0	8	$\times$	
Great flood of 99	49	6%	5	44	0	2	0	0	$\times$	
2013 North India floods	37	4%	2	35	0	3	0	0	$\times$	
Floods in India	13	2%	1	12	0	3	0	0	$\times$	
Seva Bharati	10	1%	0	10	0	4	0	0	$\times$	
Operation Madad (Indian Navy)	5	1%	0	5	0	0	0	0	$\times$	
Kalaiyarasan	2	0%	2	0	0	0	0	0	$\times$	
Chalakudy River	1	0%	0	1	0	0	0	0	$\times$	
Gauthami Nair	1	0%	1	0	0	0	0	0	$\times$	
2019 Kerala floods	0	0%	0	0	0	0	0	0	$\times$	
$\sum$	851	100%	95	756	0	43	0	8	0	0

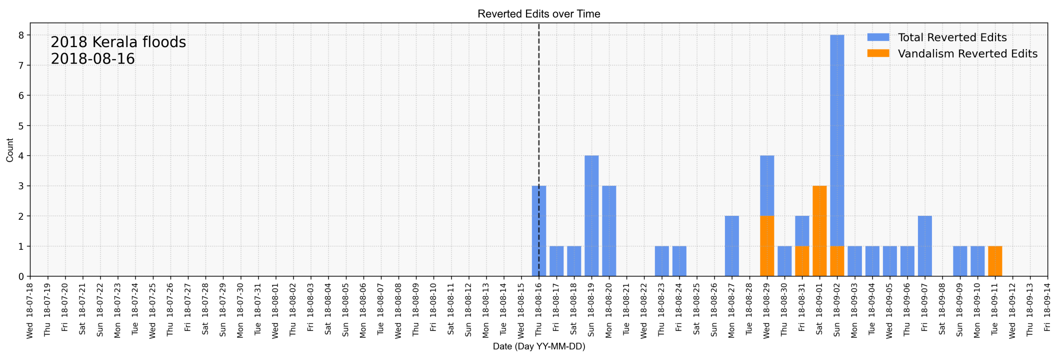
almost two and a half weeks, daily edits dropped to less than 20 edits per day, approaching the pre-event daily edit levels. Figure 5.9c shows the daily reverted edits, which peaked with 8 reverted edits after 17 days from the event. We researched why there were 8 reverts on that day, but it seems to be a normal occurrence of minor reverted edits and 1 vandalism-reverted edit. There wasn't a significant new development in the event to explain the peak on that particular day.



(a) Total and anonymous edits over an 8 weeks analysis period.



(b) Total and anonymous edits over a 12 months analysis period.



(c) Reverted edits and reverted-vandalism edits over an 8 weeks analysis period.

Figure 5.9: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the 2018 Kerala floods on 2018-08-16. Plots (a) and (c) cover an 8-week period (2018-07-19 - 2018-09-13), with 4 weeks before and after. Plot (b) covers a 12-month period (2018-02-16 - 2019-02-16), with 6 months before and after.

5.5 Sports and Entertainment Events

In this section, we analyze the impact of three Sports and Entertainment events on the English Wikipedia platform.

The first event is the 2020 Olympic Games[10] (Section 5.5.1), which took place in Tokyo, Japan, from July 23 to August 8, 2021. The Olympic Games is one of the biggest international sporting events, attracting athletes from around the world to compete in a wide range of sports. This event was chosen as it received intensive media attention due to its delayed schedule and the challenges posed by the COVID-19 pandemic. With an extended period of 17 days duration, the impact of this event on Wikipedia can be seen in the big amount of articles, including detailed articles on each sport, athlete profiles, and medal standings for each country. its interesting to see the impact of an long duration event,

The second event is the Super Bowl[11] (Section 5.5.2), an annual championship game of the National Football League (NFL) in the United States. Super Bowl LV, held on February 7, 2021, in Florida. This event was chosen because it received immense media attention as it is one of the most-watched television events in the country. In addition to the game itself, the Super Bowl halftime show featuring The Weeknd, a globally renowned Canadian singer, further amplified the entertainment value of the event.

The third event is the 94th Academy Awards[12] (Section 5.5.3). Held on February 27, 2022, in Los Angeles, California, the Academy Awards, also known as the Oscars, is one of the most prestigious events in the entertainment industry. This event was chosen because it gained significant media attention, not only for celebrating outstanding achievements in filmmaking but also due to a controversial incident involving famous actor Will Smith and comedian Chris Rock. Will Smith's unexpected action of slapping Chris Rock during Rock's presentation for Best Documentary Feature generated a lot of attention and discussion in the media.

The analysis of these events reveals that all of them experienced a surge in editing activity on relevant Wikipedia articles, indicating a heightened interest in updating and maintaining the information related to these events. However, the magnitude of the impact varied across the events. The top 10 articles related to the 2020 Summer Olympics received approximately 7 times more edits compared to Super Bowl LV and about 3 times more edits than the 94th Academy Awards. This difference can be attributed to the extended duration of the Olympic Games, which spanned 17 days and generated ongoing

[10] `https://en.wikipedia.org/wiki/2020_Summer_Olympics`
[11] `https://en.wikipedia.org/wiki/Super_Bowl_LV`
[12] `https://en.wikipedia.org/wiki/94th_Academy_Awards`

engagement and updates for each day's competitive events. The 2020 Summer Olympics exhibited a significant effect size of $d = 1.47$*** for the total number of edits, indicating a substantial impact on editing activity. In contrast, both Super Bowl LV and the 94th Academy Awards had small and insignificant effect sizes, suggesting a minimal impact on editing activity for these events.

In terms of reverted edits, there was an increase in disputes and disagreements between editors across all three events; however, the effect size varied. The 2020 Summer Olympics had a moderate effect size of 1.06***, indicating a notable impact on reverted edits and highlighting a higher level of controversies. In contrast, both Super Bowl LV and the 94th Academy Awards had small and insignificant effect sizes, suggesting a lower level of disputes among editors for these events.

Moreover, the 2020 Summer Olympics had a moderate impact on vandalism, with an effect size of 0.66*. On the other hand, both Super Bowl LV and the 94th Academy Awards had small and insignificant effect sizes, suggesting a relatively lower occurrence of vandalism for these events.

It is worth noting that there were no measures of anonymous editing protection employed by Wikipedia editors during the events for any of the top 30 articles contributing to the analysis of these events. This suggests that editors did not perceive these articles as vulnerable to vandalism and trusted the collaborative editing process to maintain the integrity of the information.

Regarding the temporal pattern of edits, all events experienced peaks immediately following each event, reflecting the immediate impact and interest in updating the articles. However, the daily edit levels returned to pre-event levels relatively quickly after the events. This indicates that while editors were highly engaged in updating articles during and immediately after the events, the sustained interest gradually declined as the initial excitement waned and new information became less available.

5.5.1 2020 Summer Olympics

To analyze the impact of the 2020 Summer Olympics, which took place on August 25, 2017, on the editing activity, we selected 100 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from June 25, 2021, to August 20, 2021). Table 5.21 presents the data obtained from these articles, which collectively received a total of 33,111 edits, with 26,829 of them being registered edits (81% of the total). Among these edits, 423 were reverted, and only 42 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 4,946 edits. However,

Table 5.21: Analysis of edits during the 2020 Summer Olympics on 2021-07-23. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2021-06-25 - 2021-08-20), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	33,111	4,946	28,165	23,219	469%	1.47***
Registered Edits	26,829	3,818	23,011	19,193	503%	1.46***
Reverted Edits	423	128	295	167	130%	1.06***
Vandalism Reverted Edits	42	8	34	26	325%	0.66*

in the four weeks after the event, the number of edits increased to 28,165. This represents a relative change of 469% and an absolute difference of 23,219 edits between the two periods. The effect size, as measured by $d = 1.47$***, indicates a substantial impact on edits.

In terms of reverted edits, there were 128 instances recorded before the event and 295 instances recorded after, representing a relative change of 130%. The effect size, as measured by $d = 1.06$***, indicates a large impact and a significant rise in disputes between editors. Similarly, the number of vandalism-reverted edits also experienced a significant relative change of 325%; however, the effect size for vandalism-reverted edits was moderate at 0.66*, suggesting a smaller impact compared to all reverted edits and total edits.

Interestingly, when examining the top 10 articles ranked by total edits during the 2020 Summer Olympics (Table 5.22), it was found that the main article specifically about the event, titled "2020 Summer Olympics," did not appear in the top 10 articles. However, the article titled "United States at the 2020 Summer Olympics" received the highest editing attention, contributing 6% of the total edits during the analysis period. It is worth noting that each article in the table contributed between 3% and 6%, indicating a relatively balanced distribution of editing attention among the top articles. Notably, none of the articles were protected from anonymous editing, suggesting that editors did not consider them vulnerable to vandalism and trusted the collaborative editing process.

Regarding the temporal pattern of edits, Figure 5.10a illustrates that the highest number of edits, totaling 2,106, occurred on the second day following the event. The daily editing activity remained consistently high throughout the duration of the 2020 Summer Olympics, reflecting the ongoing updates and information related to various sports and races taking place during the event. This sustained engagement can be attributed to the continuous cover-

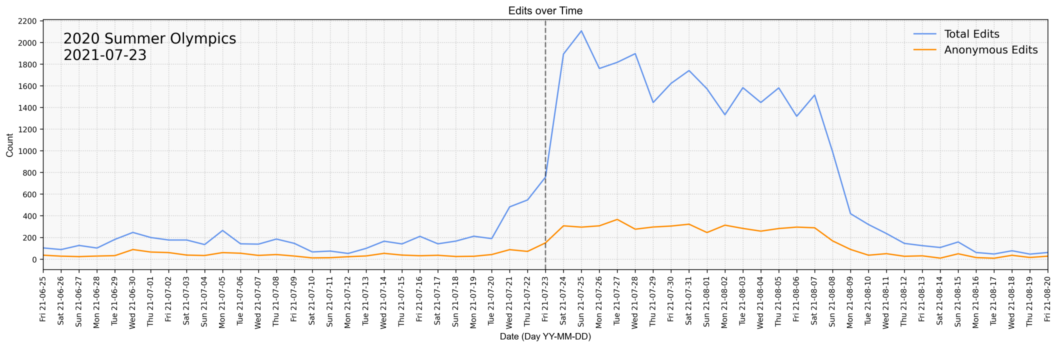
Table 5.22: Top 10 articles ranked by total edits during the 2020 Summer Olympics on 2021-07-23. Covers 8 weeks (2021-06-25 - 2021-08-20), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
United States at the 2020 Summer Olympics	2,040	6%	191	1,849	0	4	0	0	×	
India at the 2020 Summer Olympics	1,769	5%	363	1,406	12	7	0	1	×	
Great Britain at the 2020 Summer Olympics	1,755	5%	319	1,436	0	4	0	0	×	
China at the 2020 Summer Olympics	1,465	4%	79	1,386	2	10	1	0	×	
Russian Olympic Committee athletes at...	1,379	4%	147	1,232	0	2	0	1	×	
Australia at the 2020 Summer Olympics	1,207	4%	219	988	1	4	0	0	×	
2020 Summer Olympics medal table	1,144	3%	7	1,137	0	29	0	1	×	
Germany at the 2020 Summer Olympics	1,115	3%	115	1,000	0	0	0	0	×	
Canada at the 2020 Summer Olympics	1,110	3%	232	878	12	0	2	0	×	
Netherlands at the 2020 Summer Olympics	1,036	3%	134	902	0	3	0	0	×	
$\sum$	14,020	42%	1,806	12,214	27	63	3	3	0	0

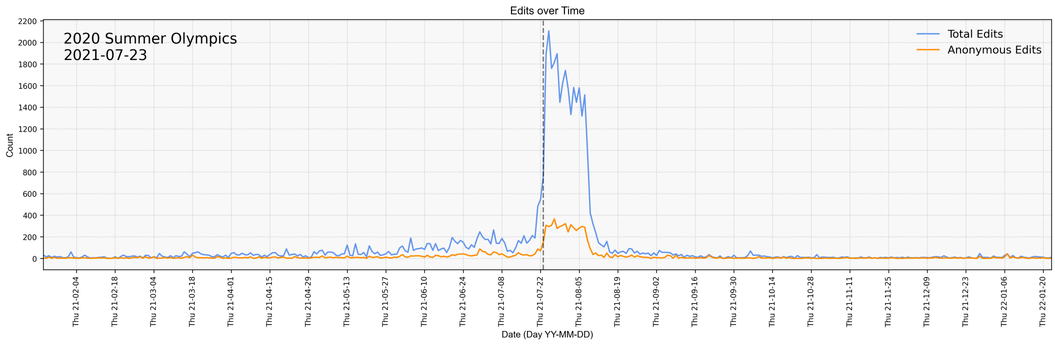
age and excitement surrounding the Olympics, with editors actively updating information until the closing ceremony on August 8, 2021.

Figure 5.10b demonstrates that the average daily edits declined rapidly after the closing ceremony, as the excitement subsided and the availability of new information decreased. The average daily edits eventually returned to pre-event levels, reaching nearly zero daily edits. This decline in editing activity indicates a decrease in interest and ongoing updates related to the 2020 Summer Olympics following its conclusion.

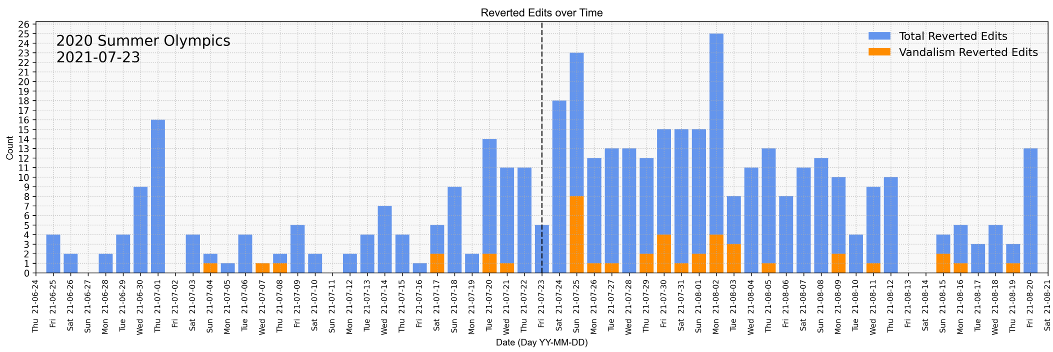
Figure 5.10c presents the increase in average daily reverted edits following the event, with a peak of 25 reverted edits. This suggests a rise in disputes or conflicts among editors regarding the content of the articles.



(a) Total and anonymous edits over an 8 weeks analysis period.



(b) Total and anonymous edits over a 12 months analysis period.



(c) Reverted edits and reverted-vandalism edits over an 8 weeks analysis period.

Figure 5.10: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the 2020 Summer Olympics on 2021-07-23. Plots (a) and (c) cover an 8-week period (2021-06-25 - 2021-08-20), with 4 weeks before and after. Plot (b) covers a 12-month period (2021-01-23 - 2022-01-23), with 6 months before and after.

Table 5.23: Analysis of edits during the Super Bowl LV on 2021-02-07. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2021-01-10 - 2021-03-07), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	2,069	703	1,366	663	94%	0.22
Registered Edits	1,632	563	1,069	506	90%	0.20
Reverted Edits	76	30	46	16	53%	0.20
Vandalism Reverted Edits	21	8	13	5	62%	0.17

5.5.2 Super Bowl LV

To analyze the impact of the Super Bowl LV, which took place on February 7, 2021, on the editing activity, we selected 18 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from January 10, 2021, to March 7, 2021). Table 5.23 presents the data obtained from these articles, which collectively received a total of 2,069 edits, with 1,632 of them being registered edits (79% of the total). Among these edits, 76 were reverted, and 21 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 703 edits. However, in the four weeks after the event, the number of edits increased to 1,366. This represents a relative change of 94% and an absolute difference of 663 edits between the two periods. The effect size, as measured by $d = 0.22$, indicates a moderate impact.

In terms of reverted edits, there were 30 instances recorded before the event and 46 instances recorded after, representing a relative change of 53%. The effect size, as measured by $d = 0.2$, indicates a small impact. The number of vandalism-reverted edits increased by 62%, with a small effect size of 0.17.

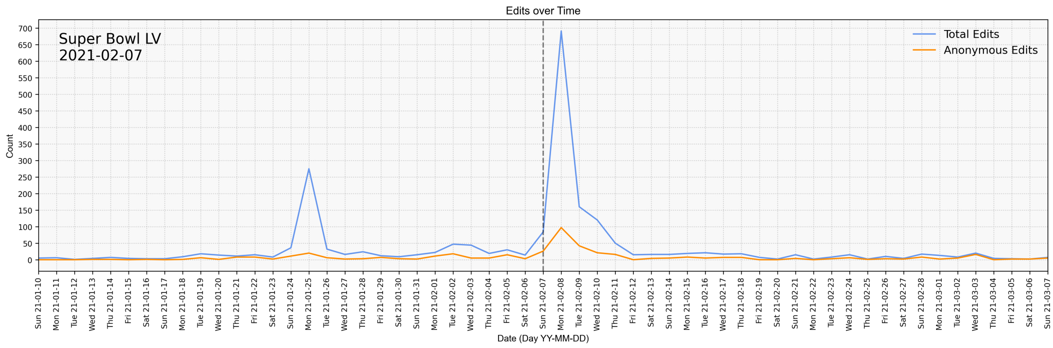
When examining the top 10 articles ranked by total edits during Super Bowl LV (Table 5.24), it was found that the main article specifically about the event, titled "Super Bowl LV," received the highest attention, contributing 43% of the total edits during the analysis period. The "Super Bowl LV halftime show" article received the second-highest editing attention, contributing 13% of the total edits. Notably, only a single article was protected from anonymous editing, and it was not implemented during the event, suggesting that editors did not consider the articles vulnerable to vandalism.

Regarding the temporal pattern of edits, Figure 5.11a demonstrates that daily edits following the event peaked only once, on the day after the event,

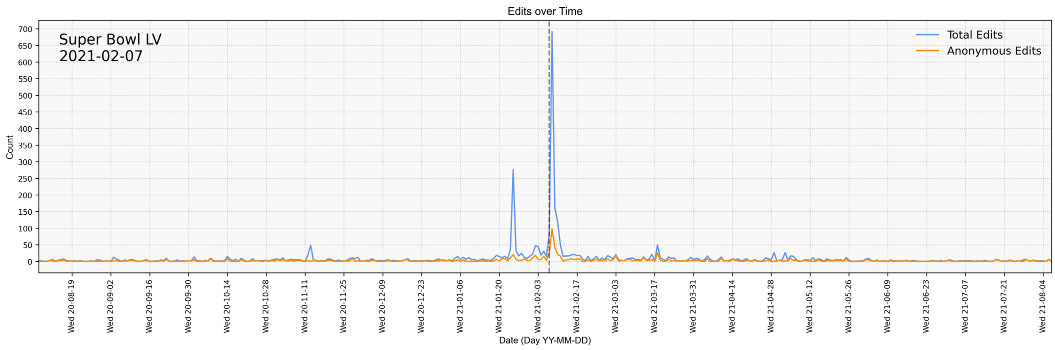
Table 5.24: Top 10 articles ranked by total edits during the Super Bowl LV on 2021-02-07. Covers 8 weeks (2021-01-10 - 2021-03-07), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
Super Bowl LV	889	43%	333	556	0	5	0	2	×	
Super Bowl LV halftime show	268	13%	20	248	1	8	0	2	×	
Super Bowl	200	10%	110	90	8	11	0	8	×	
Super Bowl Most Valuable Player Award	125	6%	18	107	8	7	5	0	×	
List of Super Bowl records	117	6%	15	102	2	2	0	1	×	
Super Bowl LVI	91	4%	25	66	3	6	3	0	×	
Super Bowl curse	65	3%	50	15	0	1	0	0	×	
Super Bowl LIV	55	3%	24	31	4	0	0	0	×	
Super Bowl LI	52	3%	23	29	1	0	0	0	×	
List of Super Bowl champions	50	2%	11	39	0	0	0	0	✓	×
$\sum$	1,912	92%	629	1,283	27	40	8	13	1	0

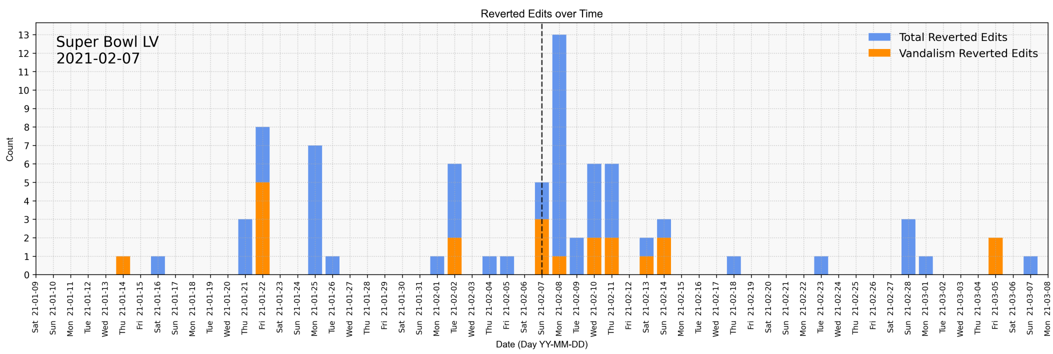
with 691 edits. However, there was a small peak in the pre-event period on January 25, 2021, with 275 edits, likely due to the introduction of new information about the upcoming event. Figure 5.11b shows that the average daily edits dropped rapidly after four days of Super Bowl LV and returned to pre-event levels, ranging between zero and 25 daily edits. Figure 5.11c displays the increase in average daily reverted edits following the event, peaking at 13 reverted edits on the day after the event.



(a) Total and anonymous edits over an 8 weeks analysis period.



(b) Total and anonymous edits over a 12 months analysis period.



(c) Reverted edits and reverted-vandalism edits over an 8 weeks analysis period.

Figure 5.11: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the Super Bowl LV on 2021-02-07. Plots (a) and (c) cover an 8-week period (2021-01-10 - 2021-03-07), with 4 weeks before and after. Plot (b) covers a 12-month period (2020-08-07 - 2021-08-07), with 6 months before and after.

Table 5.25: Analysis of edits during the 94th Academy Awards on 2022-03-27. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2022-02-27 - 2022-04-24), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	8,116	3,180	4,936	1,756	55%	0.33
Registered Edits	5,871	2,188	3,683	1,495	68%	0.37
Reverted Edits	336	147	189	42	29%	0.29
Vandalism Reverted Edits	32	18	14	4	-22%	0.18

5.5.3 94th Academy Awards

To analyze the impact of the 94th Academy Awards, which took place on March 27, 2022, on the editing activity, we selected 146 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from February 27, 2022, to April 24, 2022). Table 5.25 presents the data obtained from these articles, which collectively received a total of 8,116 edits, with 5,871 of them being registered edits (71% of the total). Among these edits, 336 were reverted, and 32 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 3,180 edits. However, in the four weeks after the event, the number of edits increased to 4,936. This represents a relative change of 55% and an absolute difference of 1,756 edits between the two periods. The effect size, as measured by $d = 0.33$, indicates a small impact.

In terms of reverted edits, there were 147 instances recorded before the event and 189 instances recorded after, representing a relative change of 29%. The effect size, as measured by $d = 0.29$, indicates a moderate impact. However, the number of vandalism-reverted edits decreased, with a minimal effect size of $d = 0.18$.

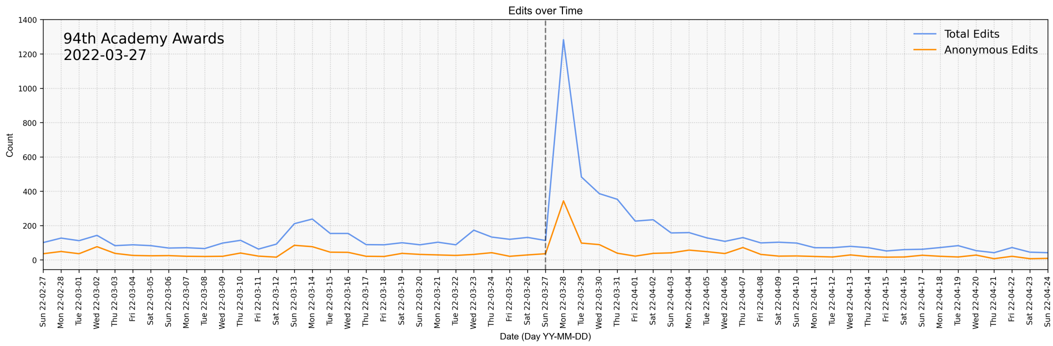
When examining the top 10 articles ranked by total edits during the 94th Academy Awards (Table 5.26), both the main article of the event titled "94th Academy Awards" and the "Will Smith–Chris Rock slapping incident" article were the highest contributors, each accounting for 16% of the total edits during the analysis period. The remaining articles in the top 10 contributed between 2% and 4% of the total edits, indicating a relatively balanced distribution of editing attention. Notably, none of the articles were protected from anonymous editing, suggesting that editors did not consider them vulnerable

Table 5.26: Top 10 articles ranked by total edits during the 94th Academy Awards on 2022-03-27. Covers 8 weeks (2022-02-27 - 2022-04-24), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

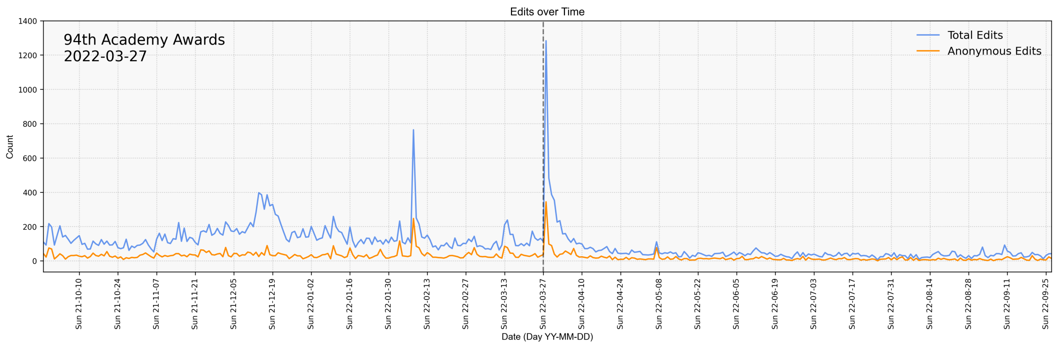
Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
94th Academy Awards	1,308	16%	268	1,040	19	12	6	1	×	
Will Smith–Chris Rock slapping incident	1,265	16%	0	1,265	0	36	0	5	×	
Spider-Man: No Way Home	365	4%	239	126	24	0	2	0	✓	×
West Side Story (2021 film)	314	4%	240	74	13	5	0	0	×	
Encanto	264	3%	173	91	8	0	1	0	×	
CODA (2021 film)	264	3%	80	184	1	5	0	1	×	
75th British Academy Film Awards	181	2%	168	13	5	1	1	0	×	
The Power of the Dog (film)	165	2%	112	53	10	2	4	0	×	
Rachel Zegler	151	2%	90	61	3	3	0	0	×	
Ariana DeBose	133	2%	59	74	0	9	0	0	×	
$\sum$	4,410	54%	1,429	2,981	83	73	14	7	1	0

to vandalism. Interestingly, the "Will Smith–Chris Rock slapping incident" article had the highest number of reverted edits among the top 10 articles, with 36 instances. Additionally, out of the 7 vandalism-reverted edits recorded following the event, 5 of them were from this article, suggesting that it attracted a significant amount of vandalism, possibly due to its humorous or meme-like nature. However, the article was not protected from anonymous editing.

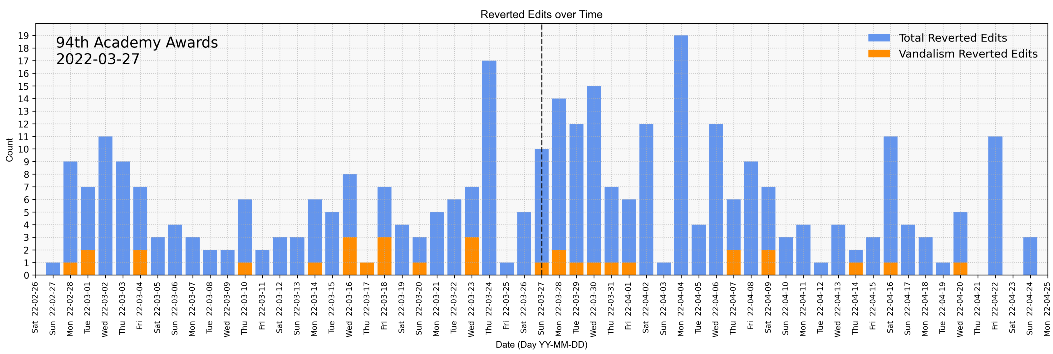
Regarding the temporal pattern of edits, Figure 5.12a demonstrates that daily edits peaked only once, on the day following the event, with 1,283 edits. Figure 5.12b shows that the average daily edits dropped rapidly after the 94th Academy Awards, within a week, and returned to pre-event levels. However, there was a spike in activity in the six-month period prior to the event on February 8, 2022, likely due to the inclusion of the movie "Spider-Man: No Way Home" article and its upcoming release. Figure 5.12c displays the average daily reverted edits, showing that the pre-event and post-event periods have a similar pattern of reverted edits.



(a) Total and anonymous edits over an 8 weeks analysis period.



(b) Total and anonymous edits over a 12 months analysis period.



(c) Reverted edits and reverted-vandalism edits over an 8 weeks analysis period.

Figure 5.12: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the 94th Academy Awards on 2022-03-27. Plots (a) and (c) cover an 8-week period (2022-02-27 - 2022-04-24), with 4 weeks before and after. Plot (b) covers a 12-month period (2021-09-27 - 2022-09-27), with 6 months before and after.

5.6 Legal and Legislative Events

In this section, we analyze the impact of three Legal and Legislative events on the English Wikipedia platform.

The first event is the Same-sex marriage legislation in the United States[13] (Section 5.6.1). which occurred On June 26, 2015, where the Supreme Court of the United States ruled in that a fundamental right to marry is guaranteed to same-sex couples. This event was chosen as This decision sparked intense debate and received extensive media attention due to its controversial nature.

The second event is the legalization of cannabis in Canada[14] (Section 5.6.2). The Cannabis Act is a law which legalized recreational cannabis use in Canada which was effective on October 17, 2018. This event was chosen as it brought cannabis into the spotlight, generating significant media attention and debates due to its controversial nature.

The third event is the General Data Protection Regulation[15] (Section 5.6.3) which this regulation was implement in 25 May 2018, in the European Union. This event was chosen as it was not as controversial as the previous two events, and may have not recieved as much as media attention.

The examination of these events revealed significant impacts on Wikipedia editing activity. All three events experienced a surge in editing, indicating a heightened interest in updating and maintaining relevant articles. However, there were differences in the magnitude of the impact among the events. The top 10 articles related to Same-sex marriage legislation received almost three times more edits compared to the Legalization of cannabis in Canada and four times more edits than the General Data Protection Regulation (GDPR). This difference could be attributed to the controversial nature of same-sex marriage, which likely generated more discussions and engagement from editors. This hypothesis is supported by the increase in reverted edits observed for the same-sex marriage event, whereas both the legalization of cannabis and the GDPR experienced a decrease in the number of reverted edits.

In terms of the top articles by total edits, the Same-sex marriage legislation and the Legalization of cannabis had a notable impact on directly related articles, while the GDPR had a concentrated effect on the main article dedicated to the regulation itself. Interestingly, the Same-sex marriage article was the only one protected from anonymous editing, indicating the recognition of the potential for vandalism and controversies surrounding the topic. However,

[13] https://en.wikipedia.org/wiki/Same-sex_marriage_legislation_in_the_United_States

[14] https://en.wikipedia.org/wiki/Same-sex_marriage_legislation_in_the_United_States

[15] https://en.wikipedia.org/wiki/General_Data_Protection_Regulation

Table 5.27: Analysis of edits during the Same sex marriage legislation in the United States on 2015-06-26. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2015-05-29 - 2015-07-24), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	2,178	572	1,606	1,034	181%	0.68*
Registered Edits	1,677	442	1,235	793	179%	0.71*
Reverted Edits	97	15	82	67	447%	0.70*
Vandalism Reverted Edits	12	2	10	8	400%	0.70*

even with protection, instances of vandalism were still present, highlighting the challenges of safeguarding controversial subjects.

The temporal patterns of edits exhibited similarities across the events, with peaks occurring shortly after the respective events took place. This reflects the immediate impact and the initial surge of editing activity. However, the duration of heightened editing activity varied among the events. The Same-sex marriage legislation sustained an increase in daily edits, potentially due to ongoing discussions and updates related to the topic. On the other hand, the Legalization of cannabis and the GDPR saw a quicker return to pre-event edit levels, indicating a more short-lived impact.

5.6.1 Same sex marriage legislation in the United States

To analyze the impact of the Same sex marriage legislation in the United States, which took place on June 26, 2015, on the editing activity, we selected 69 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from May 29, 2015, to July 24, 2015). Table 5.27 presents the data obtained from these articles, which collectively received a total of 2,178 edits, with 1,677 of them being registered edits (77% of the total). Among these edits, 97 were reverted, and only 12 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 572 edits. However, in the four weeks after the event, the number of edits increased to 1,606. This represents a relative change of 181% and an absolute difference of 1,034 edits between the two periods. The effect size, as measured by $d = 0.68^*$, indicates a moderate impact.

Reverted edits also saw a significant increase of 447%, with a similar ef-

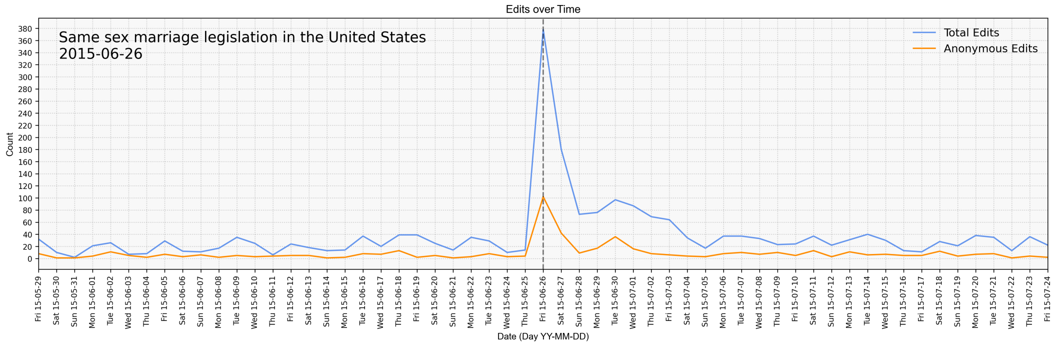
Table 5.28: Top 10 articles ranked by total edits during the Same sex marriage legislation in the United States on 2015-06-26. Covers 8 weeks (2015-05-29 - 2015-07-24), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
Same-sex marriage	459	21%	141	318	5	48	2	5	✓	×
Same-sex marriage in the United States	415	19%	60	355	6	12	0	2	×	
Obergefell v. Hodges	317	15%	2	315	0	3	0	0	×	
Recognition of same-sex unions in Europe	157	7%	79	78	2	2	0	0	×	
Same-sex union legislation	135	6%	77	58	0	1	0	0	×	
Same-sex marriage in Mexico	98	4%	45	53	0	0	0	0	×	
Same-sex marriage law in the United States by state	67	3%	1	66	0	5	0	1	×	
Legal status of same-sex marriage	58	3%	12	46	0	2	0	0	×	
Same-sex marriage in Australia	46	2%	26	20	0	1	0	0	×	
Timeline of same-sex marriage	39	2%	9	30	0	1	0	0	×	
$\sum$	1,791	82%	452	1,339	13	75	2	8	1	0

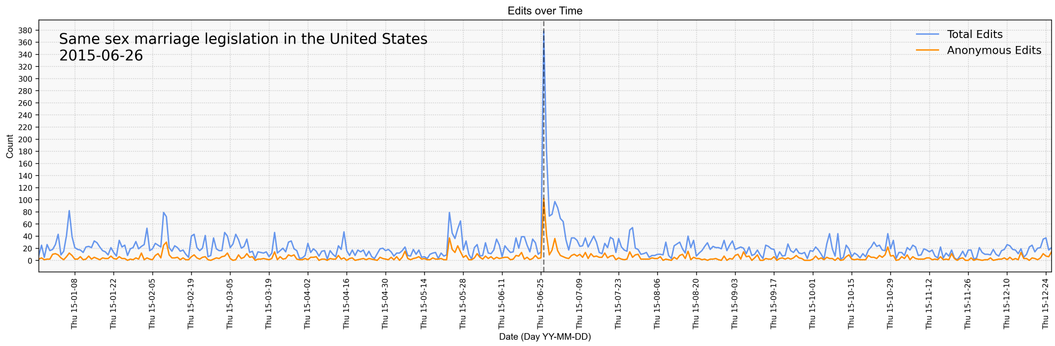
fect size to total edits of $d = 0.7$*. The number of vandalism-reverted edits increased by 8 additional edits, representing a relative change of 400%. The effect size for vandalism-reverted edits was moderate at $d = 0.7$*.

When examining the top 10 articles ranked by total edits during the Same-sex marriage legislation in the United States (Table 5.28), the article "Same-sex marriage" received the highest attention, contributing 21% of the edits to the analysis. The article "Same-sex marriage in the United States" came in second place, contributing 19% of the edits. Given that the event occurred in the United States, it is expected that these articles would receive significant editing attention. Notably, only one article was protected from anonymous editing, which was the "Same-sex marriage" article. However, the protection was implemented before the event and not specifically for this event, indicating that the restriction was not a measure specifically implemented to safeguard the article during the event. Interestingly, even with protection, the "Same-sex marriage" article experienced the highest number of instances of vandalism, with 5 instances out of a total of 10 vandalism incidents following the event. This suggests that even with restrictions on anonymous editing, vandalism may still occur, particularly for controversial subjects like same-sex marriage.

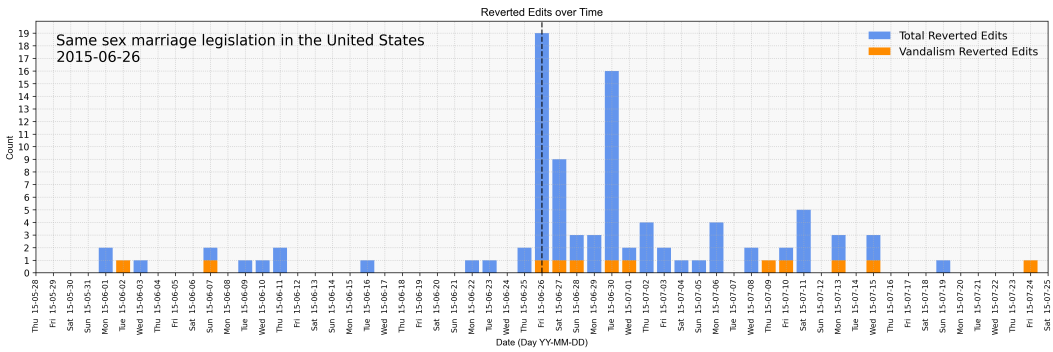
Regarding the temporal pattern of edits, Figure 5.13a illustrates that daily edits reached their peak on the day of the event, with a total of 378 edits. Figure 5.13b demonstrates that the average daily edits dropped rapidly after the event, returning to pre-event levels. Figure 5.13c showcases the increase in average daily reverted edits following the event, peaking at 19 reverted edits on the day of the event.



(a) Total and anonymous edits over an 8 weeks analysis period.



(b) Total and anonymous edits over a 12 months analysis period.



(c) Reverted edits and reverted-vandalism edits over an 8 weeks analysis period.

Figure 5.13: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the Same sex marriage legislation in the United States on 2015-06-26. Plots (a) and (c) cover an 8-week period (2015-05-29 - 2015-07-24), with 4 weeks before and after. Plot (b) covers a 12-month period (2014-12-26 - 2015-12-26), with 6 months before and after.

Table 5.29: Analysis of edits during the Legalization of cannabis in Canada on 2018-10-17. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2018-09-19 - 2018-11-14), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	805	295	510	215	73%	0.48
Registered Edits	589	192	397	205	107%	0.51
Reverted Edits	34	21	13	8	-38%	0.32
Vandalism Reverted Edits	4	4	0	4	-100%	0.45

5.6.2 Legalization of cannabis in Canada

To analyze the impact of the Legalization of cannabis in Canada, which took place on October 17, 2018, on the editing activity, we selected 37 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from September 19, 2018, to November 14, 2018). Table 5.29 presents the data obtained from these articles, which collectively received a total of 805 edits, with 589 of them being registered edits (73% of the total). Among these edits, 34 were reverted, and only 4 instances of vandalism were identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 295 edits. However, in the four weeks after the event, the number of edits increased to 510. This represents a relative change of 73% and an absolute difference of 215 edits between the two periods. The effect size, as measured by $d = 0.48$, indicates a small impact.

In terms of the number of reverted edits, it would be expected to see a high number of disputes between editors due to the controversial nature of the topic. However, the number of reverted edits, including vandalism, actually decreased. Specifically, the reverted edits decreased by 38%, with a smaller effect size of $d = 0.32$ in relation to total edits. Furthermore, there were no vandalism-reverted edits following the event, with an effect size of 0.45. It is possible that this event received less attention because it took place in Canada rather than the United States. However, further research is needed to confirm this hypothesis.

When analyzing the top 10 articles ranked by total edits during the legalization of cannabis in Canada (refer to Table 5.30), the article "same-sex marriage" garnered the most attention, contributing to 21% of the edits in the

Table 5.30: Top 10 articles ranked by total edits during the Legalization of cannabis in Canada on 2018-10-17. Covers 8 weeks (2018-09-19 - 2018-11-14), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
Legality of cannabis	174	22%	109	65	9	6	2	0	×	
Cannabis in Canada	136	17%	51	85	1	1	1	0	×	
Cannabis laws of Canada by province or territory	112	14%	28	84	0	0	0	0	×	
Legality of cannabis by U.S. jurisdiction	40	5%	10	30	0	0	0	0	×	
Cannabis (drug)	39	5%	4	35	0	0	0	0	✓	×
Legal history of cannabis in Canada	39	5%	8	31	3	0	0	0	×	
Cannabis edible	25	3%	2	23	0	0	0	0	×	
Legalization of non-medical cannabis in the United States	23	3%	3	20	0	0	0	0	×	
Cannabis	19	2%	10	9	0	0	0	0	✓	×
Cannabis in the United States	16	2%	6	10	1	0	1	0	×	
$\sum$	623	77%	231	392	14	7	4	0	2	0

analysis. The article "Same-sex marriage in the United States" ranked second, accounting for 19% of the edits. This was expected since the event occurred in the US. It is worth noting that only one article was protected from anonymous editing, which was the article that received the most attention, "Same-sex marriage." However, the protection was implemented before the event and was not specifically aimed at safeguarding it during this particular event. Interestingly, even with protection, this article experienced the most instances of vandalism, with a total of 5 instances out of 10 following the event. This suggests that, at times, vandalism cannot be completely avoided, even with restrictions on anonymous editing, particularly when it comes to controversial subjects such as same-sex marriage.

Regarding the temporal pattern of edits, Figure 5.14a illustrates a spike in daily edits on the day of the event, with 52 edits. However, the peak of daily edits occurred on October 28, 2018, 11 days after the event, with 95 edits. We investigated the reasons behind this spike and found that it can be attributed to two articles: "Cannabis laws of Canada by province or territory" and "Cannabis in Canada." On that day, these articles experienced an unusually high number of edits, specifically 41 and 39 edits respectively, accounting for 84% of the total edits made that day. Both articles were extensively revised and overhauled by the same two individuals, leading to the spike in edits. Another spike in edits, totaling 47, occurred on November 7, 2018. Upon further investigation, we discovered that this spike was due to developments in the legality of cannabis in various U.S. jurisdictions.

Table 5.31: Analysis of edits during the General Data Protection Regulation on 2018-05-25. Table shows total, registered, reverted, and vandalism-reverted edits for related articles. Covers 8 weeks (2018-04-27 - 2018-06-22), 4 before and 4 after the event. Counts provided for both periods, with absolute difference, relative change, and Cohen's d, with one to three asterisks (*) indicating p-values less or equal to 0.05, 0.01, and 0.001.

Edits Analysis	**Counts**			**Change**		
	$\sum$	$\leftarrow$	$\rightarrow$	Abs.	Rel.	d
Total Edits	455	198	257	59	30%	0.17
Registered Edits	349	140	209	69	49%	0.26
Reverted Edits	25	15	10	5	-33%	0.19
Vandalism Reverted Edits	1	1	0	1	-100%	0.27

Figure 5.14b demonstrates that both the pre-event and post-event periods experienced small spikes in edits throughout the analyzed period. This suggests that developments related to this subject occurred consistently over time, leading to these minor spikes.

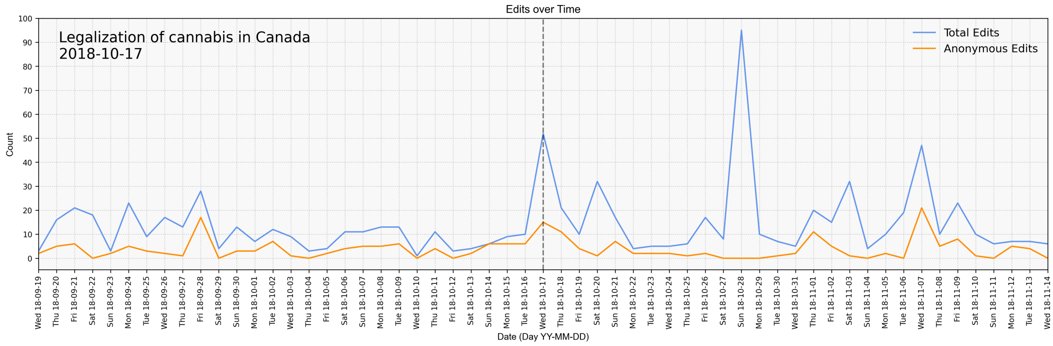
Finally, Figure 5.14c displays the decrease in average daily reverted edits following the event, as well as the absence of vandalism-reverted edits after the event.

5.6.3 General Data Protection Regulation

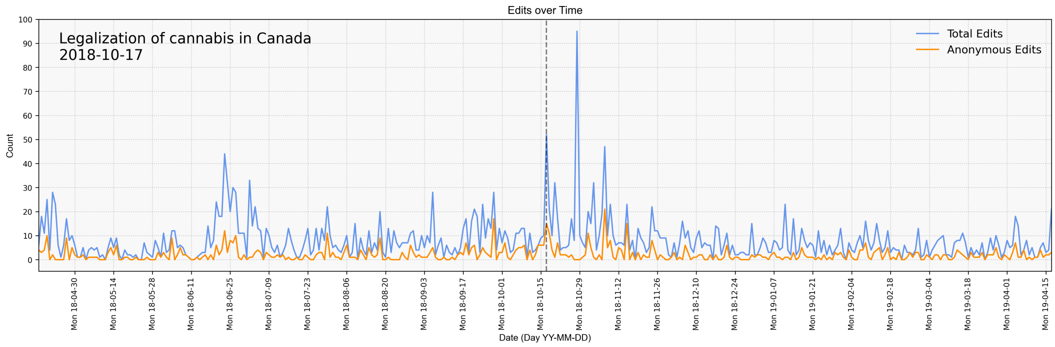
To analyze the impact of the 2021 German federal election, which began on September 26, 2021, on the editing activity, we selected 15 relevant articles to analyze over a period of 8 weeks, encompassing four weeks before and four weeks after the event (from April 27, 2018, to June 22, 2018). Table 5.31 presents the data obtained from these articles, which collectively received a total of 455 edits, with 349 of them being registered edits (77% of the total). Among these edits, 25 were reverted, and only 1 instance of vandalism was identified.

An increase in editing activity is observed following the event. In the four weeks prior to the event, these articles received a total of 198 edits. However, in the four weeks after the event, the number of edits increased to 257. This represents a relative change of 30% and an absolute difference of 59 edits between the two periods. The effect size, as measured by $d = 0.17$, indicates a small impact.

The number of reverted edits actually decreased following the event, with a reduction of 33% and a small effect size of $d = 0.19$. Moreover, there was only one instance of reverted edits before the event, and no vandalism-reverted edits were observed after the event. This suggests that the event was not highly



(a) Total and anonymous edits over an 8 weeks analysis period.



(b) Total and anonymous edits over a 12 months analysis period.

(c) Reverted edits and reverted-vandalism edits over an 8 weeks analysis period.

Figure 5.14: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the Legalization of cannabis in Canada on 2018-10-17. Plots (a) and (c) cover an 8-week period (2018-09-19 - 2018-11-14), with 4 weeks before and after. Plot (b) covers a 12-month period (2018-04-17 - 2019-04-17), with 6 months before and after.

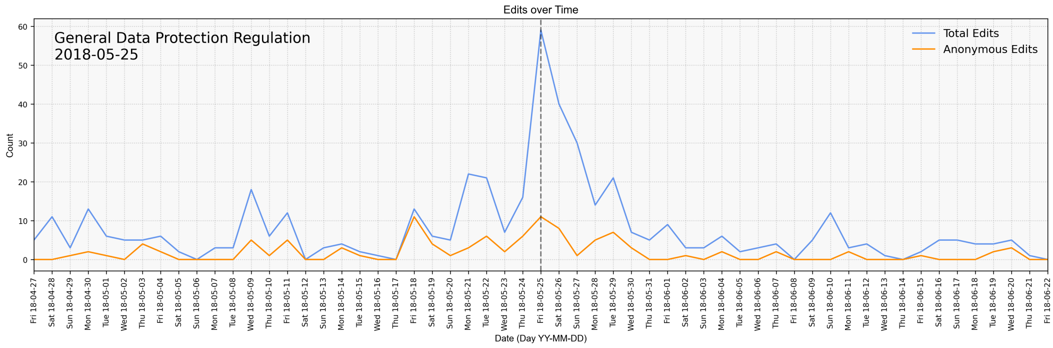
Table 5.32: Top 10 articles ranked by total edits during the General Data Protection Regulation on 2018-05-25. Covers 8 weeks (2018-04-27 - 2018-06-22), with 4 weeks before and after. Table shows counts of total, reverted, and vandalism-reverted edits for each article before and after the event. Includes protection status (🔒) and if the article was protected during the analysis period (@). Summation row provides overall statistics for all 10 articles. Percentage contribution indicates edits proportion compared to total analysis edits.

Top 10 Articles by Total Edits	**Edits**				**Reverts**		**Vand.**		**Protec.**	
	$\sum$	%	←	→	←	→	←	→	🔒	@
General Data Protection Regulation	330	73%	112	218	8	9	0	0	×	
Data Protection Act 1998	23	5%	13	10	6	0	1	0	×	
Information privacy	14	3%	13	1	0	0	0	0	×	
Pseudonymization	14	3%	14	0	0	0	0	0	×	
Data security	13	3%	11	2	0	0	0	0	×	
Data Protection Directive	13	3%	8	5	1	1	0	0	×	
Data portability	9	2%	3	6	0	0	0	0	×	
Personal data	9	2%	7	2	0	0	0	0	×	
Data Protection Act 2018	9	2%	2	7	0	0	0	0	×	
Data protection officer	7	2%	6	1	0	0	0	0	×	
$\sum$	441	97%	189	252	15	10	1	0	0	0

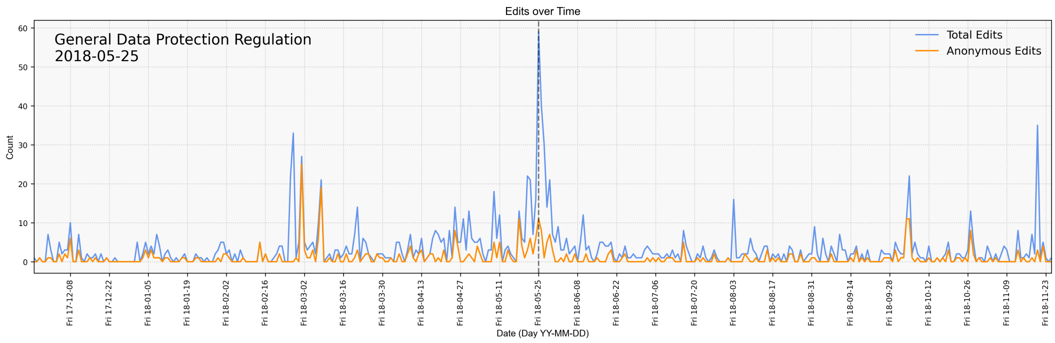
controversial, resulting in fewer disputes between editors.

When analyzing the top 10 articles ranked by total edits during the General Data Protection Regulation (refer to Table 5.32), the primary article related to the event, "General Data Protection Regulation," received the highest attention, contributing to 73% of the edits in the analysis. The second-highest contributing article accounted for 5% of the edits, indicating that the impact of the event was primarily concentrated on this specific article. Interestingly, none of the top 10 articles by total edits were protected from anonymous editing, suggesting that editors did not perceive them as vulnerable to vandalism.

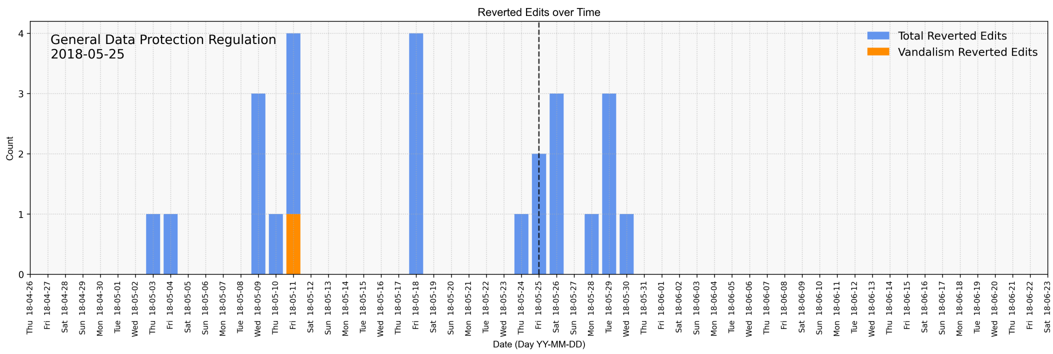
Examining the temporal pattern of edits, Figure 5.15a demonstrates that daily edits reached their peak on the day of the event, with 59 edits. Subsequently, the number of daily edits rapidly declined, returning to pre-event levels. Figure 5.15b shows that the average daily edits experienced a sharp decline after the event, eventually stabilizing at pre-event levels. However, there were fluctuations in daily edits throughout the analyzed period. Figure 5.15c displays the low number of daily reverted edits following the event, with a peak of four reverted edits occurring two days before the event.



(a) Total and anonymous edits over an 8 weeks analysis period.



(b) Total and anonymous edits over a 12 months analysis period.



(c) Reverted edits and reverted-vandalism edits over an 8 weeks analysis period.

Figure 5.15: Edits, reverted edits, including anonymous edits and vandalism-reverted edits during the General Data Protection Regulation on 2018-05-25. Plots (a) and (c) cover an 8-week period (2018-04-27 - 2018-06-22), with 4 weeks before and after. Plot (b) covers a 12-month period (2017-11-25 - 2018-11-25), with 6 months before and after.

Chapter 6

Conclusion

This chapter presents the conclusion of our study on quantifying the effects of real-world events on the editing behavior of the English Wikipedia. The research questions posed in the introduction guided our investigation, and we analyzed five categories of events: Armed Conflicts and Wars, Elections, Natural Disasters, Sports and Entertainment Events, and Legal and Legislative Events. Through examining the findings within each category, our aim was to understand the impact of these events on the level of attention, disagreements among editors, and potential disruptions to article content.

This thesis has made contributions to the understanding of real-world event impacts on Wikipedia and the dynamics of editing behavior. Firstly, we successfully reproduced [Kiesel et al., 2017], validating its findings with a newer Wikipedia history dump and laying a foundation for subsequent analyses. Secondly, we introduced a novel analysis methodology, enabling the quantification of the impact of events on Wikipedia across different language editions, facilitating comparisons and insights into the platform's response to events. Lastly, our study quantified the effect of real-world events on editing behavior, providing insights into how events shape the editing dynamics on Wikipedia.

Our analysis indicates a notable surge in editing activity after the occurrences of the events, reflecting increased interest and engagement from editors in updating information. The Russian invasion of Ukraine garnered the highest total editing activity, attributed to its significant global impact and extensive media coverage. Surprisingly, the Tigray War, with less global attention, had the highest effect size on edits, suggesting sustained involvement due to ongoing developments. Vandalism-reverted edits were most prevalent in the same-sex marriage legislation in the United States article, highlighting the challenges of protecting controversial topics. Implementing protective measures against anonymous editing proved effective in reducing vandalism during armed conflicts but not always for other subjects. Editorial biases were evident, as events in the United States received significantly more attention than similar events in other countries, as suggested by García-Gavilanes et al. [2016], and our study backed it up. Moreover, registered users contributed the majority of edits, with higher dedication and engagement compared to anonymous users. Overall, these findings reveal insights into the dynamics of Wikipedia editing activity during various events and offer insights into the factors influencing editing behavior.

It is important to acknowledge the limitations of our study. We focused on the English Wikipedia platform, which may not fully represent editing behavior in other language editions. Additionally, the categorization of events: although the categories we use in this study are quite general and capture a large portion of events, arguments can certainly be made in support of finer-grained categories that will support more nuanced analysis. However, these findings lay the groundwork for future research to explore new dimensions and expand the scope of investigation. As we have introduced a novel analysis methodology, future studies can quantify the impact of real-world events on different language versions of Wikipedia, focusing on finer-grained categories. By building upon this work, researchers can further enrich our understanding of the relationship between real-world events and the editing behavior on Wikipedia.

Appendix A

List of Analyzed Articles

The following is a list of all the articles analyzed in this thesis, categorized by event.

Russian invasion of Ukraine

Russian invasion of Ukraine
Government and intergovernmental reactions to the Russian invasion of Ukraine
Prelude to the Russian invasion of Ukraine
2022–present Ukrainian refugee crisis
Russo-Ukrainian War
Anti-war protests in Russia (2022–present)
Timeline of the Russian invasion of Ukraine
International Legion (Ukraine)
List of military aid to Ukraine during the Russo-Ukrainian War
Order of battle for the Russian invasion of Ukraine
War crimes in the Russian invasion of Ukraine
Protests against the Russian invasion of Ukraine
Casualties of the Russo-Ukrainian War
2022 Snake Island campaign
List of military engagements during the Russian invasion of Ukraine
List of aircraft losses during the Russo-Ukrainian War
Z (military symbol)
List of equipment of the Armed Forces of Ukraine
Economic impact of the Russian invasion of Ukraine
Corporate responses to the Russian invasion of Ukraine
Capture of Chernobyl
Disinformation in the Russian invasion of Ukraine
Belarusian involvement in the Russian invasion of Ukraine
Russia–Ukraine relations

List of Russian generals killed during the Russian invasion of Ukraine
List of invasions and occupations of Ukraine
Russian people's militias in Ukraine
Peace negotiations in the Russian invasion of Ukraine
Russian-occupied territories of Ukraine
Legality of the Russian invasion of Ukraine
List of equipment of the Russian Ground Forces
Russian cruiser Moskva
Georgian Legion (Ukraine)
Ukrainian–Soviet War
Wikipedia and the Russian invasion of Ukraine
International sanctions during the Russian invasion of Ukraine
United States and the Russian invasion of Ukraine
Chechen involvement in the Russian invasion of Ukraine
Russian invasion
Impact of the Russian invasion of Ukraine on nuclear power plants
Women in the Russian invasion of Ukraine
China and the Russian invasion of Ukraine
Open-source intelligence in the Russian invasion of Ukraine
Marinka, Ukraine

2021 Israel–Palestine crisis

2021 Israel–Palestine crisis
Timeline of the Israeli–Palestinian conflict in 2021
International reactions to the 2021 Israel–Palestine crisis
International protests over the 2021 Israel–Palestine crisis
International recognition of the State of Palestine
Wesley Fofana (footballer)
Mohammed el-Kurd
Iron Dome
Israeli–Palestinian conflict
Israel and apartheid
Islam in Israel
Shahed 149 Gaza
Israel
Destruction of al-Jalaa Building
China Global Television Network
Gaza electricity crisis
List of towns and villages depopulated during the 1947–1949 Palestine war
Geraldo Rivera
List of killings and massacres in Mandatory Palestine

History of Israel
History of the Israeli–Palestinian conflict
Palestine Action
Israel–Turkey relations
Arab–Israeli conflict
Military operations of the Israeli–Palestinian conflict
List of wars involving Israel
Democratic Front for the Liberation of Palestine
Siege of Beirut
9M133 Kornet
Arab citizens of Israel
Itch.io
1947–1948 civil war in Mandatory Palestine
Israeli demolition of Palestinian property
Armenians in Israel and Palestine
Popular Front for the Liberation of Palestine
Anti-war movement
Ricardo Menéndez March
Gaza War
Israel–New Zealand relations
Mariam Barghouti
Druze in Israel
Battle of Gaza
Arab–Israeli War
Mick Whitley

Tigray War
Tigray War
Mai Kadra massacre
Baykar Bayraktar TB2
Tigray People's Liberation Front
Tigray Region
Abiy Ahmed
Timeline of the Tigray War
Eritrean–Ethiopian War
2022
List of war crimes
Bellingcat
Wartime sexual violence
Ethiopian Air Force
Getachew Reda

2021 Ethiopian general election
Eritrean Defence Forces
Lalibela
Tigray
Abala, Ethiopia
Shire (Tigray)
Church of Our Lady Mary of Zion

2020 United States presidential election
2020 United States presidential election
2020 United States presidential election in Georgia
2020 United States presidential election in Pennsylvania
2020 United States presidential election in Arizona
2020 United States presidential election in Wisconsin
2020 United States presidential election in Florida
2020 United States presidential election in Michigan
2020 United States presidential election in North Carolina
2020 United States presidential election in Texas
Attempts to overturn the 2020 United States presidential election
2020 United States presidential election in Ohio
2020 United States presidential election in New Jersey
2020 United States presidential election in California
2020 United States presidential election in Maine
2020 United States presidential election in Iowa
2020 United States presidential election in Nevada
2020 United States presidential election in Alaska
2020 United States presidential election in Colorado
2020 United States presidential election in New York
2020 United States presidential election in Montana
2020 United States presidential election in Virginia
2020 United States presidential election in Illinois
2020 United States presidential election in Indiana
2020 United States presidential election in Missouri
2020 United States presidential election in Utah
2020 United States presidential election in Alabama
2020 United States presidential election in Kansas
2020 United States presidential election in New Mexico
2020 United States presidential election in Oklahoma
2020 United States presidential election in Massachusetts
2020 United States presidential election in Washington (state)
2020 United States presidential election in Nebraska

2020 United States presidential election in South Carolina
2020 United States presidential election in Mississippi
2020 United States presidential election in Maryland
2020 United States presidential election in Connecticut
2020 United States presidential election in Kentucky
2020 United States presidential election in Arkansas
2020 United States presidential election in Delaware
2020 United States presidential election in Idaho
2020 United States presidential election in West Virginia
2020 United States presidential election in Vermont
2020 United States presidential election in Hawaii
2020 United States presidential election in Louisiana
2020 United States presidential election in Oregon
2020 United States presidential election in South Dakota
2020 United States presidential election in Wyoming
2020 United States presidential election in the District of Columbia
2020 United States presidential election in North Dakota
2020 United States presidential election in Tennessee

2021 German federal election
2021 German federal election
Results of the 2021 German federal election
Opinion polling for the 2021 German federal election
Next German federal election
2017 German federal election
1994 German federal election
2009 German federal election
2005 German federal election
Candidates of the 2021 German federal election
Elections in Germany

2018 Bangladeshi general election
2018 Bangladeshi general election
2001 Bangladeshi general election
2014 Bangladeshi general election
June 1996 Bangladeshi general election
1991 Bangladeshi general election
February 1996 Bangladeshi general election
2008 Bangladeshi general election
Elections in Bangladesh
1988 Bangladeshi general election

2024 Bangladeshi general election

Hurricane Harvey
2017 Atlantic hurricane season
Hurricane Harvey
Joel Osteen
Houston
Timeline of the 2017 Atlantic hurricane season
Lakewood Church
Cajun Navy
Tropical cyclone
List of Texas hurricanes (1980–present)
Rockport, Texas
Sylvester Turner
Hurricane Katrina
Harvey
Federal Emergency Management Agency
1981 Atlantic hurricane season
Lockheed C-130 Hercules
Minute Maid Park
Refugio, Texas
Effects of Hurricane Harvey in Texas
1993 Atlantic hurricane season

2018 Sulawesi earthquake and tsunami
2018 Sulawesi earthquake and tsunami
Lists of 21st-century earthquakes
List of tsunamis
2018 Indonesia earthquake
Operation Samudra Maitri
List of tsunamis affecting Indonesia
2018 Indonesian tsunami
2018 Indonesia tsunami
Catholic Relief Services
Floating Mosque of Palu

2018 Kerala floods
2018 Kerala floods
Great flood of 99
2013 North India floods
Floods in India

Seva Bharati
Operation Madad (Indian Navy)
Kalaiyarasan
Chalakudy River
Gauthami Nair
2019 Kerala floods

2020 Summer Olympics
United States at the 2020 Summer Olympics
India at the 2020 Summer Olympics
Great Britain at the 2020 Summer Olympics
China at the 2020 Summer Olympics
Russian Olympic Committee athletes at the 2020 Summer Olympics
Australia at the 2020 Summer Olympics
2020 Summer Olympics medal table
Germany at the 2020 Summer Olympics
Canada at the 2020 Summer Olympics
Netherlands at the 2020 Summer Olympics
2020 Summer Olympics
Japan at the 2020 Summer Olympics
Brazil at the 2020 Summer Olympics
List of 2020 Summer Olympics medal winners
France at the 2020 Summer Olympics
2020 Summer Olympics opening ceremony
Hungary at the 2020 Summer Olympics
Indonesia at the 2020 Summer Olympics
Greece at the 2020 Summer Olympics
Turkey at the 2020 Summer Olympics
Philippines at the 2020 Summer Olympics
Serbia at the 2020 Summer Olympics
Volleyball at the 2020 Summer Olympics – Men's tournament
Tennis at the 2020 Summer Olympics – Men's singles
Baseball at the 2020 Summer Olympics
All-time Olympic Games medal table
Sport climbing at the 2020 Summer Olympics – Men's combined
Badminton at the 2020 Summer Olympics – Men's singles
Badminton at the 2020 Summer Olympics – Women's singles
Volleyball at the 2020 Summer Olympics – Women's tournament
Athletics at the 2020 Summer Olympics
2032 Summer Olympics
Athletics at the 2020 Summer Olympics – Men's 100 metres

Tennis at the 2020 Summer Olympics – Women's singles
Basketball at the 2020 Summer Olympics – Men's tournament
Sweden at the 2020 Summer Olympics
India at the Olympics
Swimming at the 2020 Summer Olympics
Sport climbing at the 2020 Summer Olympics – Women's combined
Tennis at the 2020 Summer Olympics
Football at the 2020 Summer Olympics – Women's tournament
Wrestling at the 2020 Summer Olympics
Softball at the 2020 Summer Olympics
Summer Olympic Games
2024 Summer Olympics
Norway at the 2020 Summer Olympics
Football at the 2020 Summer Olympics – Men's tournament
Fencing at the 2020 Summer Olympics
Boxing at the 2020 Summer Olympics
Judo at the 2020 Summer Olympics
Tennis at the 2020 Summer Olympics – Men's doubles
Basketball at the Summer Olympics
Basketball at the 2020 Summer Olympics – Men's team rosters
2020 Summer Paralympics
Football at the 2020 Summer Olympics
List of Olympic Games host cities
2028 Summer Olympics
IOC Refugee Olympic Team at the 2020 Summer Olympics
Rowing at the 2020 Summer Olympics
Cycling at the 2020 Summer Olympics
Gymnastics at the 2020 Summer Olympics
Japan at the Olympics
Cycling at the 2020 Summer Olympics – Men's individual road race
Weightlifting at the 2020 Summer Olympics
Shooting at the 2020 Summer Olympics
Tennis at the 2020 Summer Olympics – Mixed doubles
Badminton at the 2020 Summer Olympics
Taekwondo at the 2020 Summer Olympics
Canoeing at the 2020 Summer Olympics
Rugby sevens at the 2020 Summer Olympics
Golf at the 2020 Summer Olympics – Men's individual
Bids for the 2036 Summer Olympics
Basketball at the 2020 Summer Olympics
Volleyball at the 2020 Summer Olympics

Karate at the 2020 Summer Olympics
Skateboarding at the 2020 Summer Olympics
Table tennis at the 2020 Summer Olympics
Water polo at the 2020 Summer Olympics
Diving at the 2020 Summer Olympics
2016 Summer Olympics
Sailing at the 2020 Summer Olympics
Archery at the 2020 Summer Olympics
Field hockey at the 2020 Summer Olympics
Equestrian at the 2020 Summer Olympics
Sport climbing at the 2020 Summer Olympics
Golf at the 2020 Summer Olympics
Triathlon at the 2020 Summer Olympics
Surfing at the 2020 Summer Olympics
Olympic sports
Handball at the 2020 Summer Olympics
Cycling at the Summer Olympics
1964 Summer Olympics
Venues of the 2020 Summer Olympics and Paralympics
Baseball at the Summer Olympics
Artistic swimming at the 2020 Summer Olympics
Wrestling at the Summer Olympics
1940 Summer Olympics
2026 Summer Youth Olympics
Istanbul bid for the 2020 Summer Olympics
Tokyo Olympics

Super Bowl LV

Super Bowl LV
Super Bowl LV halftime show
Super Bowl
Super Bowl Most Valuable Player Award
List of Super Bowl records
Super Bowl LVI
Super Bowl curse
Super Bowl LIV
Super Bowl LI
List of Super Bowl champions
Super Bowl XLV
Super Bowl LIII
List of Super Bowl broadcasters

Super Bowl commercials
Super Bowl LVII
Super Bowl XXXVIII
Super Bowl VI
Super Bowl Sunday

94th Academy Awards
94th Academy Awards
Will Smith–Chris Rock slapping incident
Spider-Man: No Way Home
West Side Story (2021 film)
Encanto
CODA (2021 film)
75th British Academy Film Awards
The Power of the Dog (film)
Rachel Zegler
Ariana DeBose
The Eyes of Tammy Faye (2021 film)
Luca (2021 film)
List of Disney live-action adaptations and remakes of Disney animated films
Free Guy
Belfast (film)
Licorice Pizza
Snow White (2024 film)
Nightmare Alley (2021 film)
Cyrano (film)
Raya and the Last Dragon
Academy Awards
Academy Award for Best Picture
No Time to Die
Academy Award for Best Actress
Shang-Chi and the Legend of the Ten Rings
Drive My Car (film)
Academy of Motion Picture Arts and Sciences
Cruella (film)
List of accolades received by Dune (2021 film)
Jane Campion
The Mitchells vs. the Machines
House of Gucci
Denis Villeneuve
Academy Award for Best Animated Feature

Summer of Soul
List of Walt Disney Animation Studios films
Regina Hall
Timothée Chalamet
Riz Ahmed
List of accolades received by The Power of the Dog (film)
The Worst Person in the World (film)
Being the Ricardos
The Lost Daughter (film)
List of accolades received by Belfast (film)
Academy Award for Best Original Song
Academy Award for Best Actor
List of Academy Award records
List of James Bond films
Parallel Mothers
List of awards and nominations received by Jane Campion
Army of the Dead
List of accolades received by Licorice Pizza
List of Nordic Academy Award winners and nominees
List of Jamie Dornan performances
Westbrook (company)
Caitríona Balfe
A Hero
Academy Award for Best Supporting Actress
List of accolades received by CODA (2021 film)
Jonas Poher Rasmussen
Wanda Sykes
Titane
The Dress (2020 film)
The Hand of God (film)
List of Academy Award-winning families
Memoria (2021 film)
Reinaldo Marcus Green
Academy Award for Best Supporting Actor
List of awards and nominations received by Steven Spielberg
Carlos López Estrada
Emancipation (2022 film)
Three Songs for Benazir
Escape from Mogadishu
Academy Award for Best Costume Design
Dos Oruguitas

Paul Lambert (special effects artist)
Ala Kachuu – Take and Run
Rooney Mara
Coming 2 America
The Windshield Wiper
Bestia (2021 film)
List of French Academy Award winners and nominees
Joachim Trier
Academy Award for Best Production Design
Minamata (film)
Academy Award for Best Adapted Screenplay
Lamb (2021 film)
Great Freedom
Amy Schumer
Bad Luck Banging or Loony Porn
Compartment No. 6
Brighton 4th
Lunana: A Yak in the Classroom
Amira (film)
Audible (film)
To Leslie
List of Polish Academy Award winners and nominees
List of awards and nominations received by Judi Dench
The Long Goodbye (Riz Ahmed album)
Nayattu (2021 film)
List of submissions to the 94th Academy Awards for Best International Feature Film
Kevin Messick
List of Big Five Academy Award winners and nominees
Nicholas Britell
40th Academy Awards
Affairs of the Art
List of superlative Academy Award winners and nominees
Vanessa Hudgens
List of Indian winners and nominees of the Academy Awards
Jonathan Fawkner
Dan Oliver
Pawo Choyning Dorji
Hive (film)
I'm Your Man (2021 film)
Olga (2021 film)

The Gravedigger's Wife
Don Phillips (casting director)
Lead Me Home
73rd Academy Awards
Clara Sola
When We Were Bullies
Jon Spaihts
78th Academy Awards
Do Not Hesitate
David Korins
Lingui, The Sacred Bonds
Private Desert
49th Academy Awards
48th Academy Awards
46th Academy Awards
35th Academy Awards
7th Academy Awards
List of Academy Award nominees presented under false names
Should the Wind Drop
Foscadh
Pebbles (film)
Mandela (2021 film)
2021 Academy Awards
Rehana Maryam Noor
Nothing but the Sun
Costa Brava, Lebanon
Yuni (film)
The Falls (2021 film)
Oasis (2020 film)
Plaza Catedral
Jay Rosenblatt (filmmaker)

Same-sex marriage legislation in the United State
Same-sex marriage
Same-sex marriage in the United States
Obergefell v. Hodges
Recognition of same-sex unions in Europe
Same-sex union legislation
Same-sex marriage in Mexico
Same-sex marriage law in the United States by state
Legal status of same-sex marriage

Same-sex marriage in Australia
Timeline of same-sex marriage
Timeline of same-sex marriage in the United States
Same-sex marriage in the United Kingdom
Defense of Marriage Act
Same-sex marriage in Texas
Same-sex marriage in Guam
Same-sex marriage in the United States Virgin Islands
Same-sex marriage in Ohio
History of same-sex marriage in the United States
Same-sex marriage in tribal nations in the United States
Same-sex marriage legislation in the United States
Same-sex marriage in Slovenia
Same-sex unions in the United States
Religious views on same-sex marriage
Same-sex adoption in the United States
Same-sex marriage in Taiwan
Same-sex marriage in Switzerland
Same-sex marriage in Spain
Same-sex marriage in Michigan
Same-sex marriage in Canada
Timeline of civil marriage in the United States
Same-sex marriage in Chile
Same-sex marriage in Jersey
Same-sex marriage in Norway
Same-sex marriage in Indiana
Rights and responsibilities of marriages in the United States
Civil partnership in the United Kingdom
Same-sex marriage in Denmark
Same-sex marriage in Finland
Divorce of same-sex couples
Same-sex marriage in Arizona
Same-sex marriage in Belgium
Same-sex marriage in Arkansas
Same-sex marriage in France
Same-sex marriage in New York
Same-sex marriage in Argentina
Same-sex marriage in Minnesota
Same-sex marriage in Pennsylvania
Recognition of same-sex unions in India
Respect for Marriage Act

Marriage age in the United States
Same-sex marriage in Austria
Same-sex marriage in Maryland
Same-sex marriage in Yukon
Same-sex marriage in Colombia
Same-sex marriage in Ontario
Same-sex marriage in Manitoba
Same-sex marriage in Nunavut
Same-sex marriage in Alberta
Same-sex marriage in Saskatchewan
Same-sex marriage in California
Same-sex marriage in Cuba
Recognition of same-sex unions in Nepal
Same-sex marriage in Portugal
Marriage (Same Sex Couples) Act 2013
Hawaii Marriage Equality Act
Same-sex marriage in Washington (state)
Same-sex marriage in Uruguay
Same-sex marriage in Germany
Same-sex marriage in Florida

Legalization of cannabis in Canada

Legality of cannabis
Cannabis in Canada
Cannabis laws of Canada by province or territory
Legality of cannabis by U.S. jurisdiction
Cannabis (drug)
Legal history of cannabis in Canada
Cannabis edible
Legalization of non-medical cannabis in the United States
Cannabis
Cannabis in the United States
Cannabis Act
420 (cannabis culture)
Medical cannabis
Cannabis in India
Cannabis in Michigan
Cannabis in New Jersey
Cannabis in Illinois
Cannabis in California
Minors and the legality of cannabis

Cannabis in Mexico
Cannabis in Hawaii
Cannabis in Japan
Legal history of cannabis in the United States
Effects of legalized cannabis
Cannabis in Washington (state)
History of cannabis
Cannabis and the Canadian military
Cannabis in Colorado
Grassroots—Legalize Cannabis Party
Cannabis in Maryland
Cannabis in Maine
Cannabis in Norway
Cannabis in Portugal
Cannabis in North Dakota
Cannabis in Minnesota
Cannabis in New York
Dispensary

General Data Protection Regulation

General Data Protection Regulation
Data Protection Act 1998
Information privacy
Pseudonymization
Data security
Data Protection Directive
Data portability
Personal data
Data Protection Act 2018
Data protection officer
Right to be forgotten
Children's Online Privacy Protection Act
EPrivacy Regulation
Information Commissioner's Office
Do Not Track

Bibliography

Zihang Chen and Mizuho Iwaihara. Detection of editing bursts and extraction of significant keyphrases from wikipedia edit history. In Wookey Lee, Carson K. Leung, and Aziz Nasridinov, editors, *Big Data Analyses, Services, and Smart Data*, pages 45–65, Singapore, 2021. Springer Singapore. ISBN 978-981-15-8731-3.

Marco Fisichella and Andrea Ceroni. Event detection in wikipedia edit history improved by documents web based automatic assessment. *Big Data and Cognitive Computing*, 5(3), 2021. ISSN 2504-2289. doi: 10.3390/bdcc5030034. URL `https://www.mdpi.com/2504-2289/5/3/34`.

Fabian Flöck, Denny Vrandečić, and Elena Simperl. Revisiting reverts: Accurate revert detection in wikipedia. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, HT '12, page 3–12, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450313353. doi: 10.1145/2309996.2310000. URL `https://doi.org/10.1145/2309996.2310000`.

Ruth García-Gavilanes, Milena Tsvetkova, and Taha Yasseri. Dynamics and biases of online attention: the case of aircraft crashes. *Royal Society Open Science*, 3(10):160460, 2016. doi: 10.1098/rsos.160460. URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsos.160460`.

Scott A. Golder and Michael W. Macy. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333(6051):1878–1881, 2011. doi: 10.1126/science.1202775. URL `https://www.science.org/doi/abs/10.1126/science.1202775`.

James Howison, Andrea Wiggins, and Kevin Crowston. Validity issues in the use of social network analysis with digital trace data. *J. Assoc. Inf. Syst.*, 12:2, 2011.

Yuheng Hu, Shelly Farnham, and Kartik Talamadupula. Predicting user engagement on twitter with real-world events. *Proceedings of the Interna-*

tional AAAI Conference on Web and Social Media, 9(1):168–177, Aug. 2021. doi: 10.1609/icwsm.v9i1.14638. URL `https://ojs.aaai.org/index.php/ICWSM/article/view/14638`.

Johannes Kiesel, Martin Potthast, Matthias Hagen, and Benno Stein. Spatio-temporal Analysis of Reverted Wikipedia Edits. In *Eleventh International AAAI Conference on Web and Social Media (ICWSM 2017)*, pages 122–131, May 2017. URL `https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/viewFile/15699/14802`.

Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. He says, she says: Conflict and coordination in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, page 453–462, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595935939. doi: 10.1145/1240624.1240698. URL `https://doi.org/10.1145/1240624.1240698`.

Ruoran Liu, Qiudan Li, Can Wang, Lei Wang, Daniel Dajun Zeng, and Hongyuan Ma. Mining phase evolution for hot topics: A case study from multiple social media platforms. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2814–2819, 2017. doi: 10.1109/SMC.2017.8123053.

Ingmar Poese, Steve Uhlig, Dali Kaafar, Benoit Donnet, and Bamba Gueye. Ip geolocation databases: Unreliable? *Computer Communication Review*, 41:53–56, 04 2011. doi: 10.1145/1971162.1971171.

Jeffrey M Rzeszotarski and Aniket Kittur. Learning from history: predicting reverted work at the word level in wikipedia. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 437–440, 2012.

Khoi-Nguyen Tran and Peter Christen. Cross language prediction of vandalism on wikipedia using article views and revisions. In Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu, editors, *Advances in Knowledge Discovery and Data Mining*, pages 268–279, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-37456-2.