Quantifying the Effects of Real-World Events on Wikipedia

Thesis Defense | 29th August 2023

Student
Ali Saqallah

Supervisor
Dr. Johannes Kiesel

Referees
Prof. Dr. Benno Stein
Prof. Dr. Volker Rodehorst

Introduction

Wikipedia's Significance

- Crucial information resource for people
- Importance extends to large language models like ChatGPT
- Beyond human use, vital for machine verification of factual accuracy

Scope of Wikipedia's Content

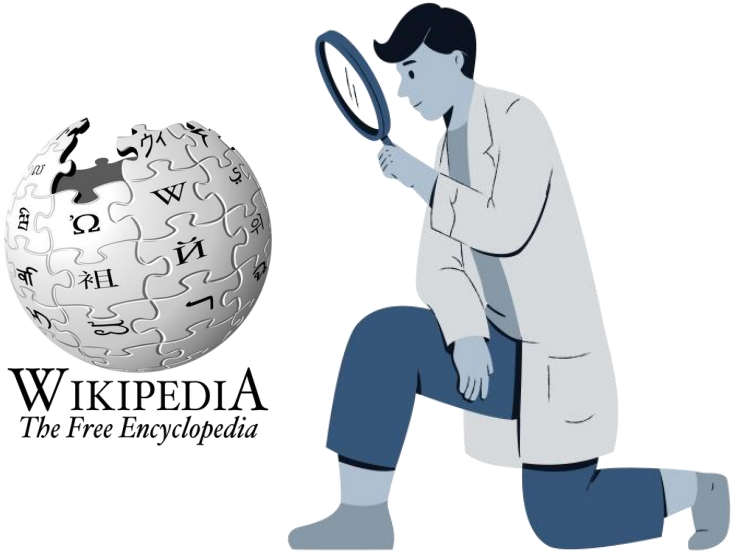
- Encompasses static knowledge (natural sciences, biographies)
- Also covers evolving information about ongoing events and developments

Establishing Trust in Wikipedia

- Understanding evolution of information and editing behavior
- Influence of external factors on editing behavior

Thesis Focus

- Investigating real-world events' (e.g., armed conflicts, elections, disasters) impact on Wikipedia editing behavior
- Quantification of these effects



Introduction

Thesis Objective:

Quantitatively analyze the influence of real-world events on English Wikipedia editing behavior.

Sub-Objectives:

01 Investigating Event Impact on overall editing activity:

Understand how real-world events affect overall editing activity on English Wikipedia.

02 Exploring Editing Behavior Patterns:

Identify patterns in editing behavior around real-world events, including editing bursts, revisions, and changes in focus.

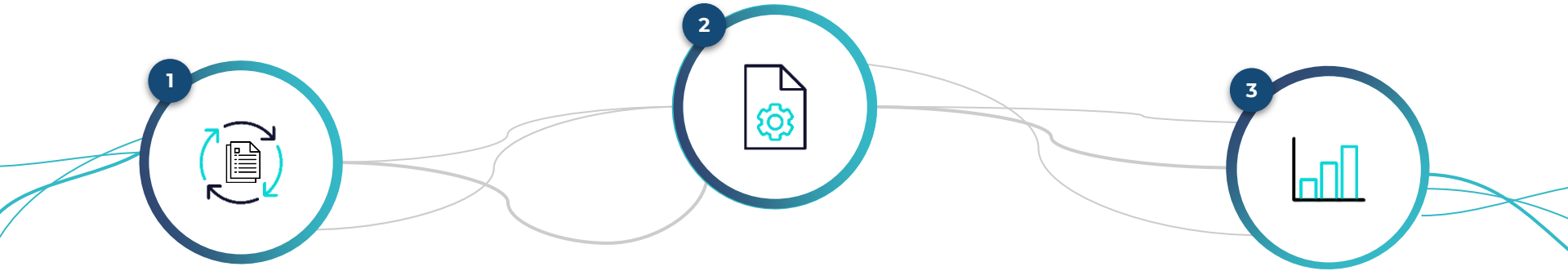
03 Variations Across Event Categories:

Explore if event types (e.g., armed conflicts, elections, disasters) lead to distinct editing behavior patterns.

04 Wikipedia's Response via Article Protection:

Analyze how Wikipedia protects articles related to events to prevent vandalism and misinformation.

Introduction



Contributions

Reproduction of Previous Study

- Successfully replicated Kiesel et al.'s (2017) "Spatio-temporal Analysis of Reverted Wikipedia Edits" findings with newer Wikipedia data dump.
- Verified earlier research's robustness, establishing a strong foundation for analysis.

Development of Analysis Methodology

- Introduced a novel analysis approach to measure real-world events' impact on Wikipedia.
- Scalable and systematic methodology applicable to various Wikipedia language editions.

Quantification of Event Impact

- Analyzed 15 events across 5 categories.
- Enhance understanding of how events shape Wikipedia editing dynamics.

REPRODUCTION OF PREVIOUS STUDY

"Spatio-temporal Analysis of Reverted Wikipedia Edits" by Kiesel et al.'s [2017]

Reproduction of Previous Study

"Spatio-temporal Analysis of Reverted Wikipedia Edits" by Kiesel et al.'s [2017].

The paper analyzes vandalism-reverted edits in Wikipedia with regard to the time and the country it originates from.

Objective :

01 Methodology Validation:
- Validate the applicability of methodologies in subsequent analysis.

02 Software Functionality Testing:
- Evaluate the functionality of their open source software to obtain relevant editing data for subsequent analysis.

03 Consistency Over Time Assessment:
- Analyze editing behavior consistency over 6-7 years.
- Determine if editing patterns have changed.
- Explore the potential longevity of findings.

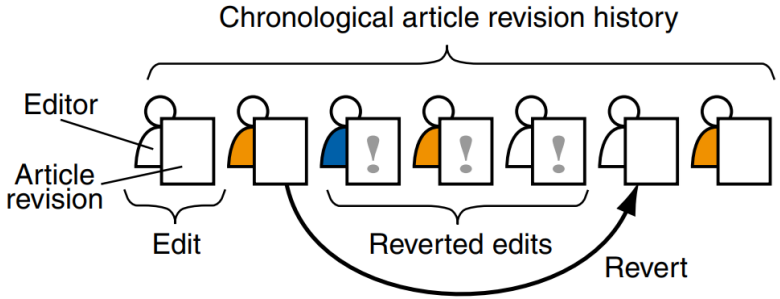


Figure taken from [Kiesel et al., 2017]. Illustration of a Wikipedia article revision history.

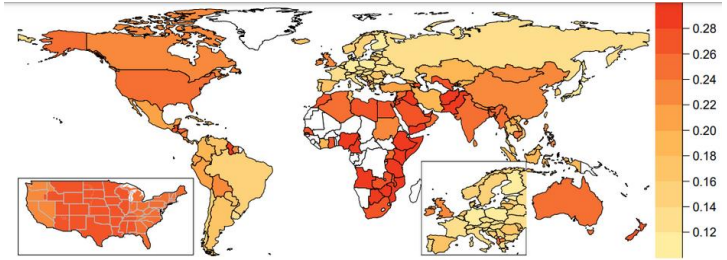
Reproduction of Previous Study | Conclusion

Successful Replication:

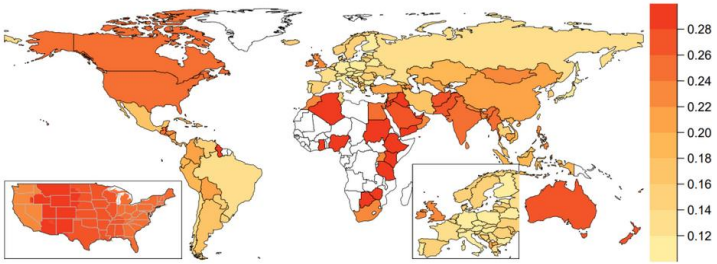
- Validated original findings, with minor deviations attributed to updated dataset.
- Open-source software remains functional.
- Methods effective for subsequent chapters in our thesis.

Consistency Over Time:

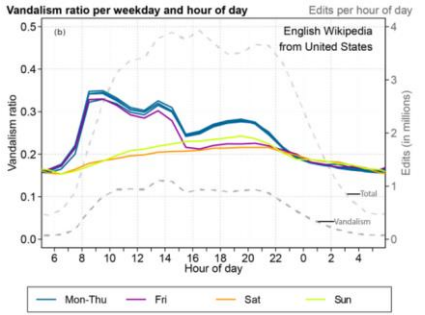
- Editing behavior on Wikipedia remains relatively consistent
- Patterns and trends identified in original study hold true with recent dataset.



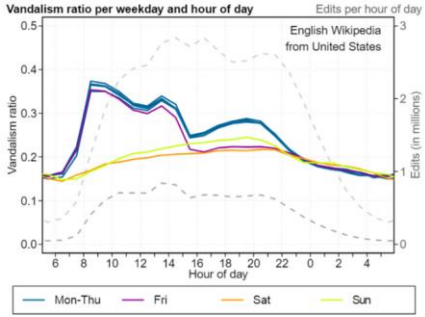
(a) From our reproduction study.



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



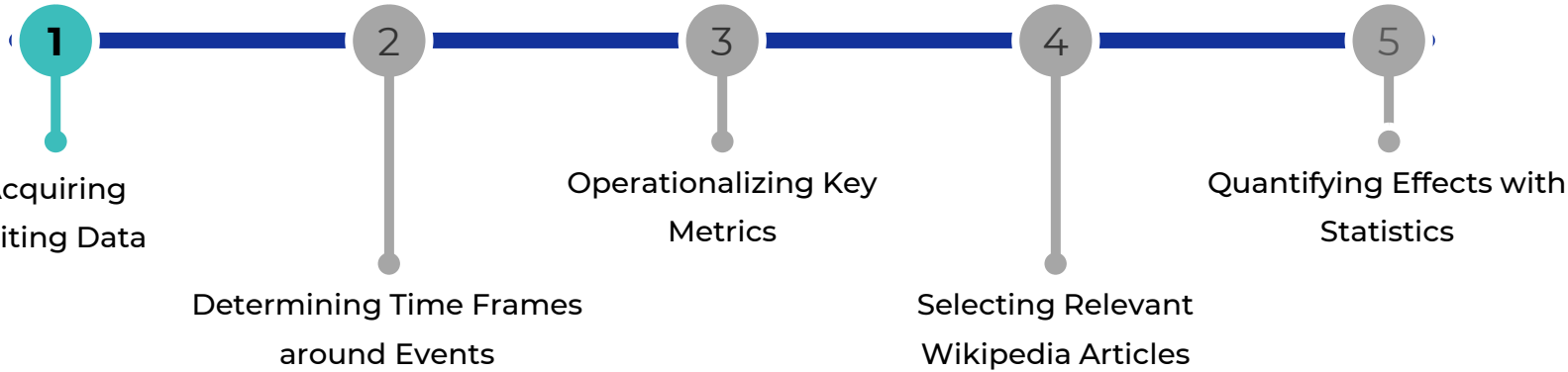
(a) From our reproduction study.



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

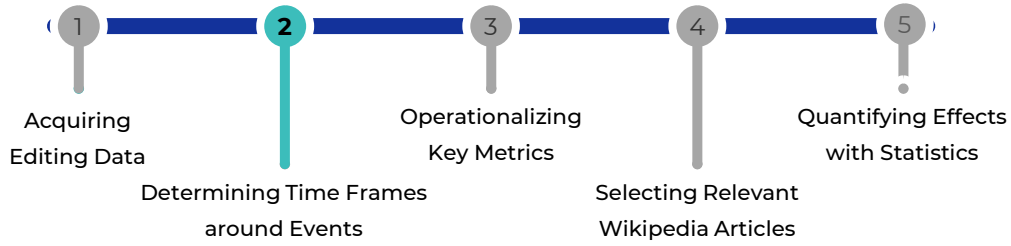
DEVELOPMENT OF ANALYSIS METHODOLOGY

Development of Analysis Methodology



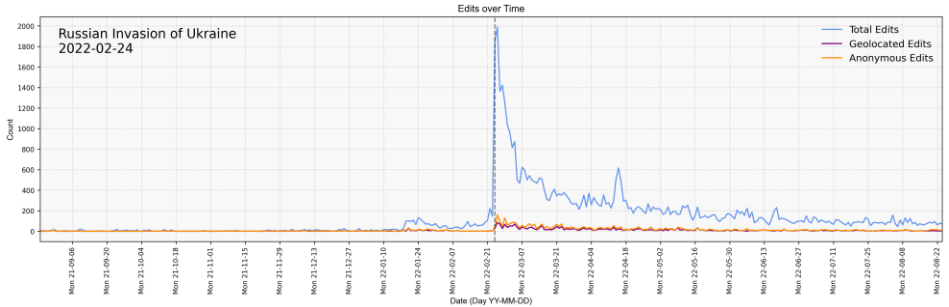
Relevant editing data obtained from the reproduction of Kiesel et al.'s [2017] paper "Spatio-temporal Analysis of Reverted Wikipedia Edits," using authors' open-source software.

Development of Analysis Methodology



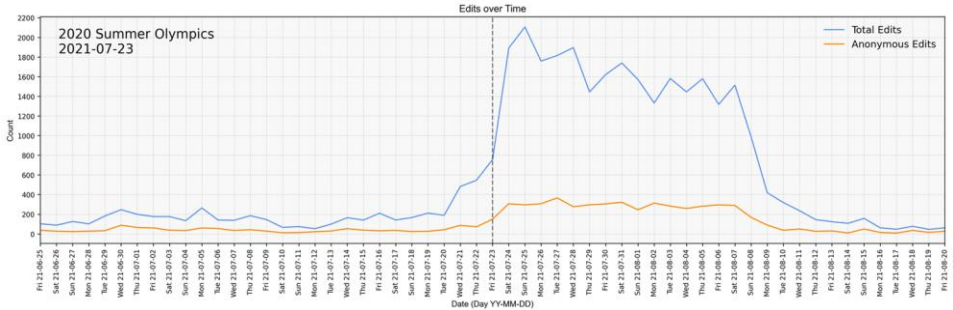
12-month period

- Graphical interpretation.
- Includes 6 months before and 6 months after events.
- Assess distinct patterns beyond immediate pre/post-event periods.

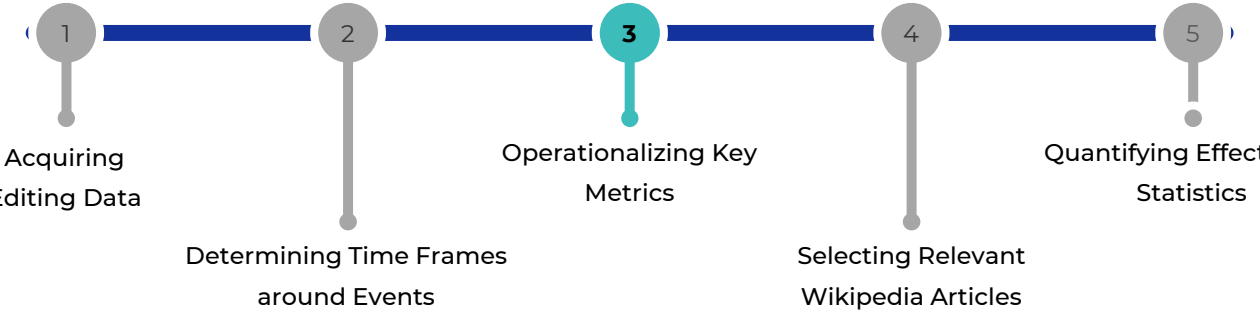


8-week period

- Analyze 4 weeks before and 4 weeks after the event (Quantifying editing behavior changes before and after the event).
- Comparing to identify shifts and trends.



Development of Analysis Methodology



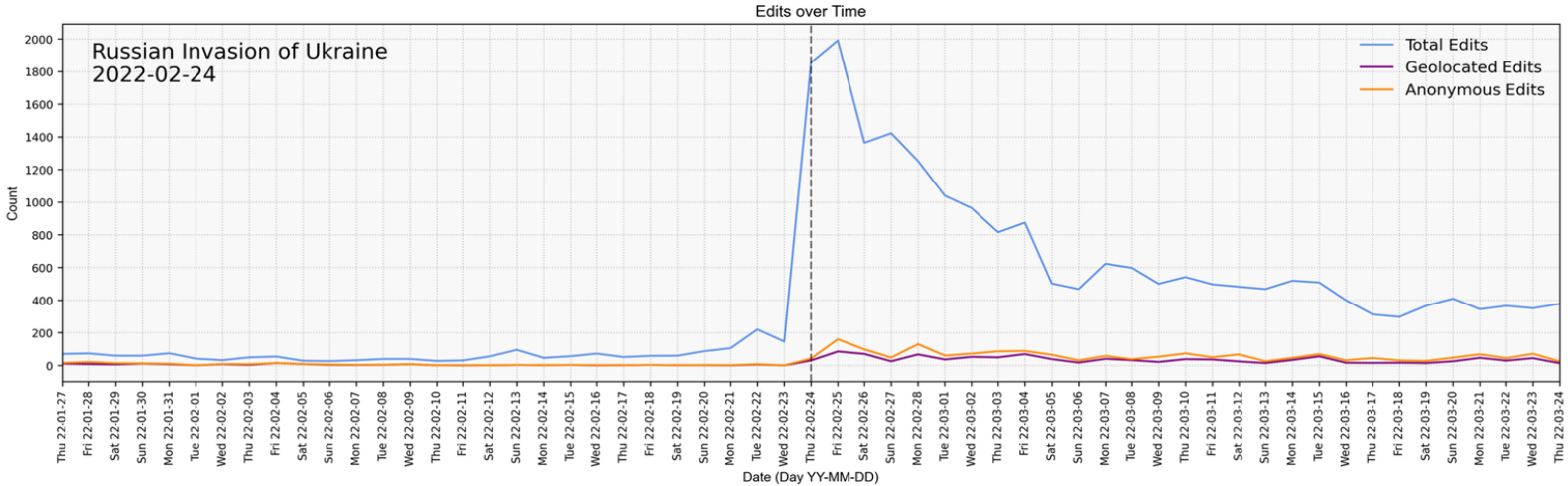
What can we measure?

- Total Edits
- Reverted Edits
- Vandalism-Reverted Edits
- Top Articles and their Contributions
- Number of Protected Articles

Development of Analysis Methodology

Total Edits
Measure of overall editing activity and engagement.

1. Acquiring Editing Data
2. Determining Time Frames around Events
3. Operationalizing Key Metrics
4. Selecting Relevant Wikipedia Articles
5. Quantifying Effects with Statistics

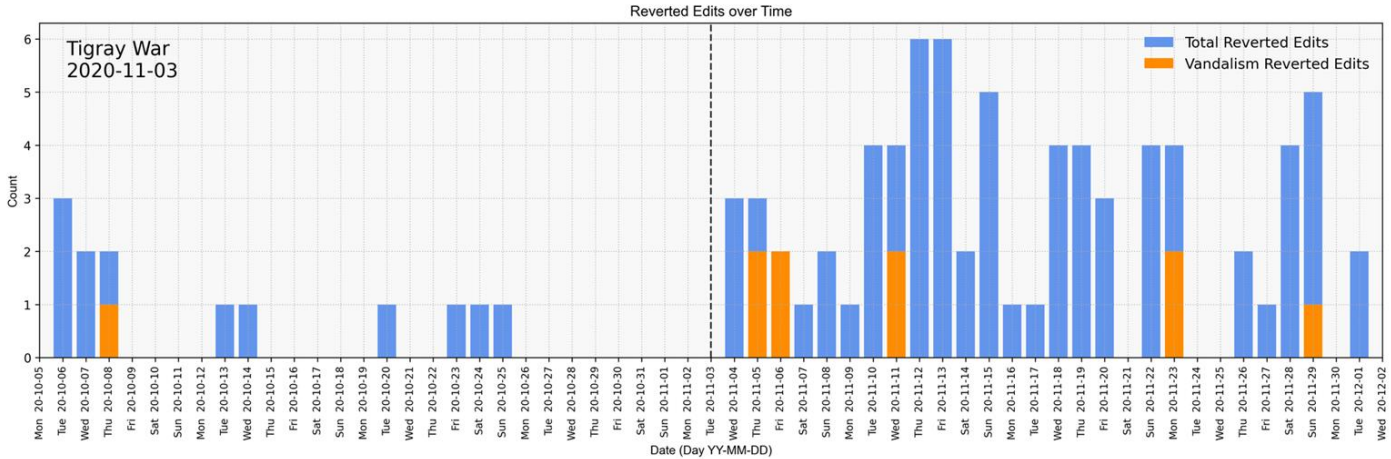


Development of Analysis Methodology

Reverted Edits
Indication of conflicts or disagreements among editors.

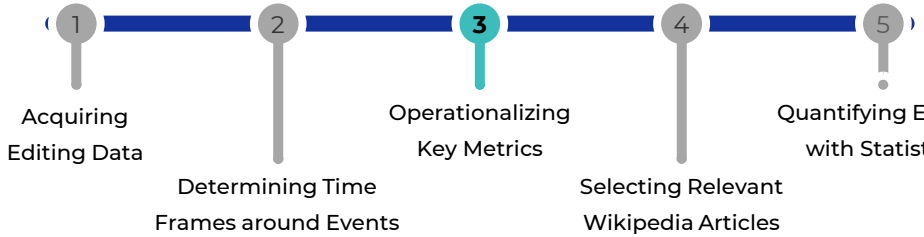
Vandalism-Reverted Edits
Highlighting malicious attempts to manipulate content.

1 Acquiring Editing Data
2 Determining Time Frames around Events
3 Operationalizing Key Metrics
4 Selecting Relevant Wikipedia Articles
5 Quantifying Effects with Statistics



Development of Analysis Methodology

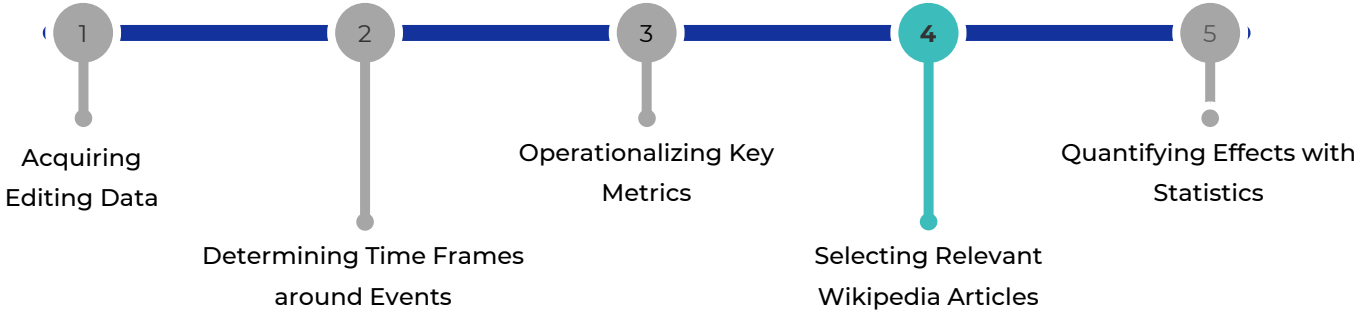
Top Articles and their Contributions
Identifying topics that attract the most attention.



Number of Protected Articles
Understanding Wikipedia editors' response to events.

Top 10 Articles by Total Edits	Edits				Reverts		Vand.		Protec.	
	∑	%	←	→	←	→	←	→	🔒	@
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	×	
∑	17,011	76%	1,625	15,386	26	35	8	1	5	4

Development of Analysis Methodology



Definition of Relevant Article

- A "relevant article" pertains directly to the event under study.
- Provides specific information, details, and perspectives about the event.

4 Approaches Explored:

Main Article:

Limits editing behavior comparison before and after the event (Main articles often lack pre-event editing data).

Main Category:

Many unrelated articles; missing relevant ones.

Directly Related Category :

Inconsistent, lacks direct connections.

Title Based Search:

- Iterative process for relevance.
- Ensures high relevance to the event.
- Consistent.

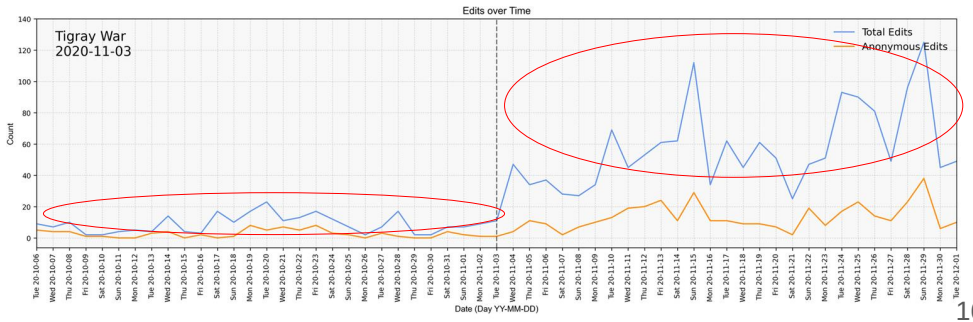
Development of Analysis Methodology

1. Acquiring Editing Data
2. Determining Time Frames around Events
3. Operationalizing Key Metrics
4. Selecting Relevant Wikipedia Articles
5. Quantifying Effects with Statistics

Quantifying Effects by:

- Conducting comparative analysis of key metrics before and after events.
- Computing relative change and supporting it using Cohen's d.
- Examining top 10 contributing articles and their protection status.
- Comparing events within categories and across different categories.

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



QUANTIFICATION OF EVENT IMPACT

Quantification of Event Impact

Selection of Event Categories:

Objective is to study the impact of different event types on editing behavior.

Category	Event	Expected editing behavior
Armed Conflicts and Wars	Russian invasion of Ukraine 2021 Israel–Palestine crisis Tigray War	Evoke strong emotions, result in extensive article updates.
Elections	2020 United States presidential election 2021 German federal election 2018 Bangladeshi general election	Trigger high engagement due to political significance.
Natural Disasters	Hurricane Harvey 2018 Sulawesi earthquake and tsunami 2018 Kerala floods	Prompt editing surges for information sharing and documentation.
Sports and Entertainment Events	2020 Summer Olympics Super Bowl LV 94th Academy Awards	Drive increased updating due to broad interest.
Legal and Legislative Events	Same sex marriage legislation in the United States Legalization of cannabis in Canada General Data Protection Regulation	Spark discussions and revisions reflecting new developments.

Quantification of Event Impact

01 Overall Editing Activity

02 Reverted Edits

03 Vandalism-Reverted Edits

04 Article Protection and Vandalism

05 Editorial Biases

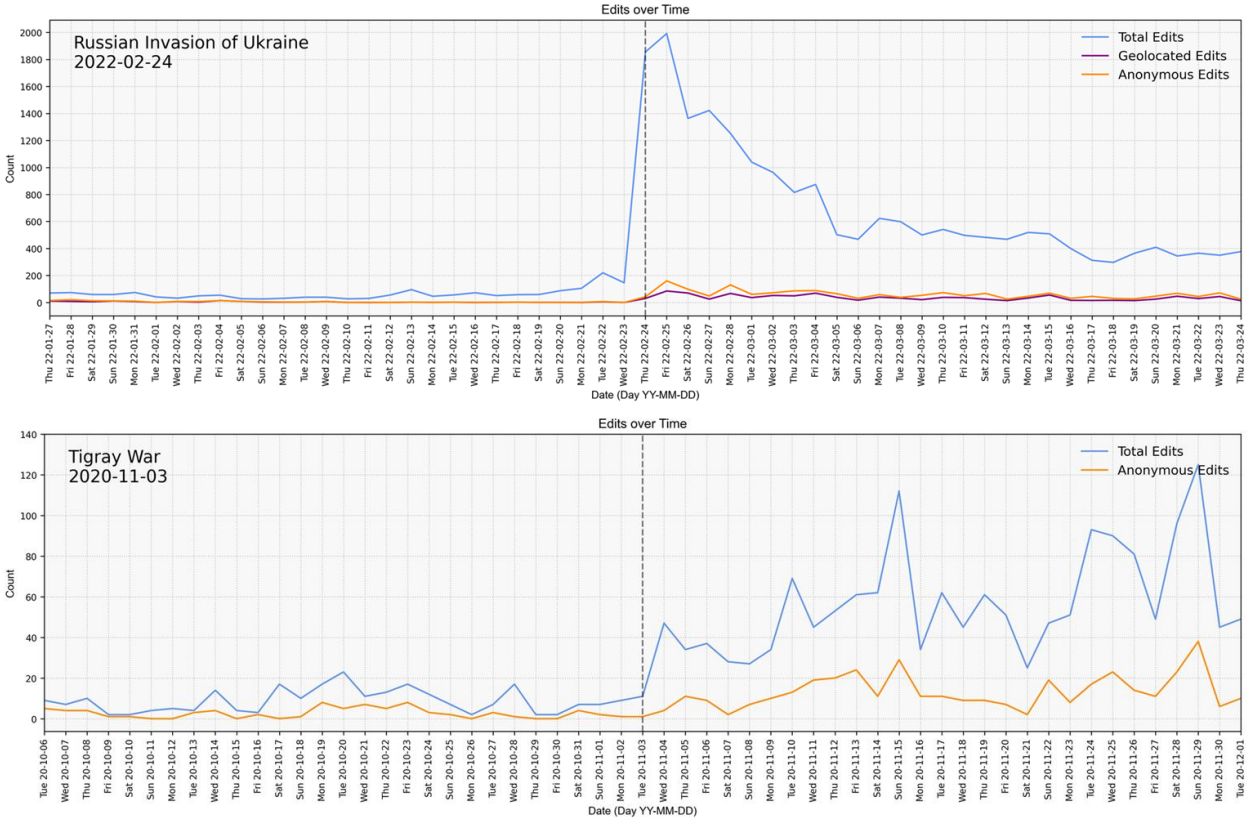
06 Registered Users Contributions

Category/Event	Top 10 articles					Cohen's d		
	$\sum$	←	→	🔒	@	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

Quantification of Event Impact | Results

01 Overall Editing Activity

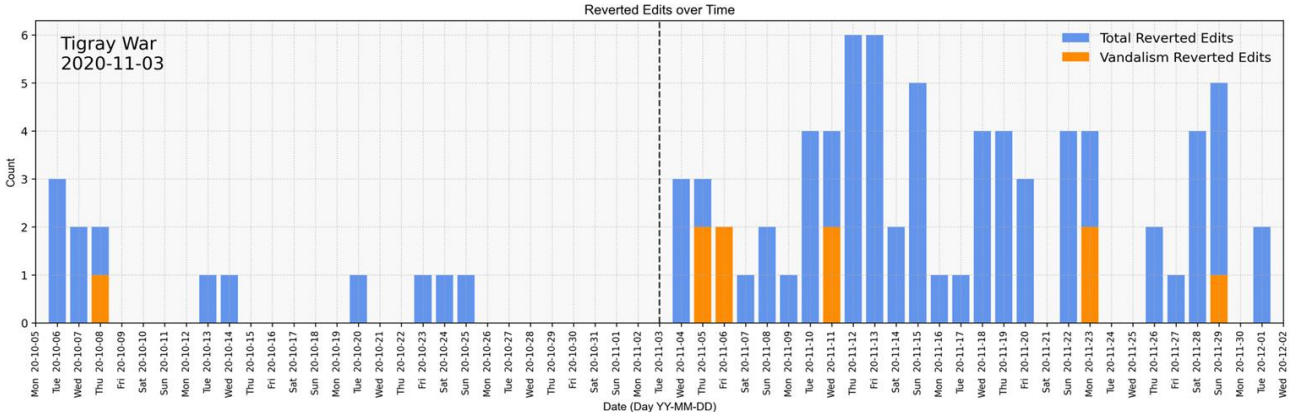
- Increased editing activity observed after events across all categories.
- Russian invasion of Ukraine event had highest total edits due to global impact.
- Tigray War exhibited highest effect size (d = 2.42***), indicating sustained activity.



Quantification of Event Impact | Results

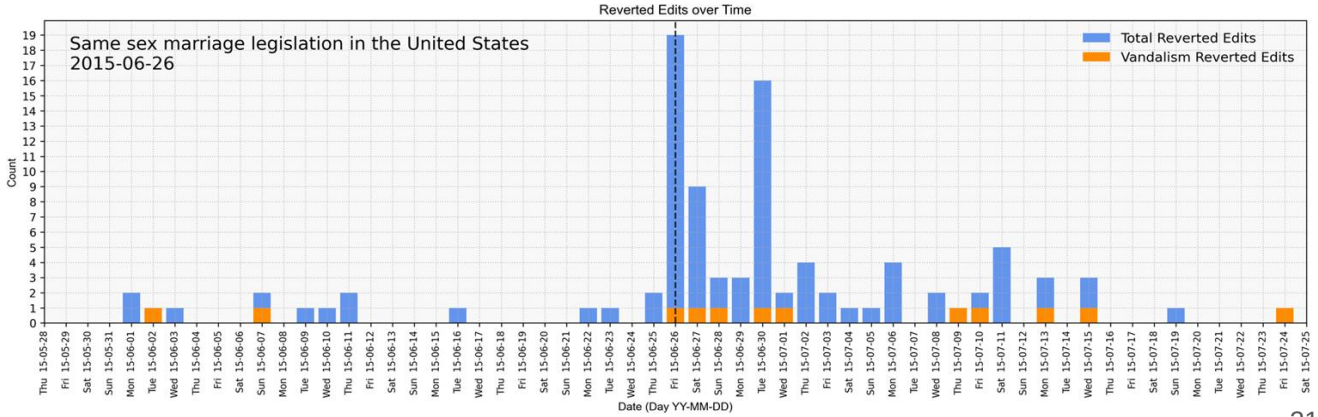
02 Reverted Edits

- Tigray War had highest effect size on reverted edits (d = 1.45***).
- Russian invasion of Ukraine and Israel-Palestine crisis had lower reverting rates.



03 Vandalism-Reverted Edits

- Same-sex marriage legislation event had highest effect size for vandalism-reverted edits (d = 0.70*).
- Controversial topics likely contributed to increased vandalism-reverted edits.



Quantification of Event Impact | Results

04 Article Protection and Vandalism

- Armed conflict-related articles more protected due to vulnerability.
- Same-sex marriage article still prone to vandalism despite protection.

Category/Event	Top 10 articles					Cohen's *d*		
	∑	←	→	🔒	@	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

Top 10 Articles by Total Edits	Edits				Reverts		Vand.		Protec.	
	∑	%	←	→	←	→	←	→	🔒	@
Same-sex marriage	459	21%	141	318	5	48	2	5	✓	×

Quantification of Event Impact | Results

05 Editorial Biases

- United States events received higher editing activity compared to similar events from other countries.

06 Registered Users Contributions

- Majority of edits during events were by registered users.
- Lowest registered edits: 71% during 94th Academy Awards event.
- Highest registered edits: 96% during 2021 Israel-Palestine crisis event.
- Registered users demonstrate consistent dedication to updating and enhancing Wikipedia's content.

Category/Event	Top 10 articles		
	$\sum$	←	→
Armed Conflicts and Wars			
Russian invasion of Ukraine	17,011	1,625	15,386
2021 Israel-Palestine crisis	3,023	101	2,922
Tigray War	1,783	213	1,570
Elections			
2020 United States presidential election	5,290	1,885	3,405
2021 German federal election	1,321	293	1,028
2018 Bangladeshi general election	350	133	217
Natural Disasters			
Hurricane Harvey	5,200	744	4,456
2018 Sulawesi earthquake and tsunami	699	16	683
2018 Kerala floods	851	95	756
Sports and Entertainment Events			
2020 Summer Olympics	14,020	1,806	12,214
Super Bowl LV	1,912	629	1,283
94th Academy Awards	4,410	1,429	2,981
Legal and Legislative Events			
Same-sex marriage legislation in the United States	1,791	452	1,339
Legalization of cannabis in Canada	623	231	392
General Data Protection Regulation	441	189	252

Quantification of Event Impact | Conclusion

Contributions

- Reproduction of Previous Study
- Development of Analysis Methodology
- Quantification of Event Impact

Notable findings

Editing Surge:

- Lowest relative change in edits: 30% during General Data Protection.
- Highest relative change in edits: 4,169% during 2018 Sulawesi earthquake and tsunami.
- Russian invasion of Ukraine saw highest total edits
- Tigray War showed highest effect size on edits

Vandalism and Article Protection:

- Armed conflict-related articles more protected due to vulnerability.
- Some articles still prone to vandalism despite protection.

Editorial Biases and Engagement

- US events received more attention than others
- Higher dedication and engagement from registered users (Registered users contributed majority of edits).

Study Limitations

English Wikipedia Focus:

- May not fully represent other language editions.

Categorization of Events:

- General categories captured significant events
- Potential for finer-grained categorization

Future Directions

- Explore impact on different language editions
- Investigate finer-grained event categories