Wikipedia Edit War Analysis

Riško Hrivnák, Paul Straberger
Vienna University of Economics and Business

Supervisor: Sabrina Kirrane

July 14, 2024

Contents

1 Introduction

When relying on Wikipedia for a quick lookup or even deeper research, it can be easy to forget just how fragile the truthfulness of content hosted there can sometimes be. Wikipedia's greatest strength — its openness — is at the same time its greatest weakness. Anyone can edit Wikipedia, which ensures that a lot of eyes can validate the accuracy of information and correct potential mistakes. But it also opens the floodgates to vandalism, petty arguments over who invented which national dish and in worse cases, historical revisionism and the undermining of established facts.

Thus Wikipedia edit wars represent a battleground in the fight against misinformation, as conflicting edits impact the reliability of the world's most accessed information source. Understanding the dynamics of these conflicts is not only academically interesting, it could also be important for developing effective strategies to maintain the integrity of online knowledge repositories and combat the spread of false or misleading information. Over the past few weeks, we poured through edit logs of wikipedia and fetched countless rows of csv data in an effort to quantify what it means when wikipedians go to war.

This is a brief write-up of our work so far. We hope that others who find this repository can use this mini-report as a helpful resource.

2 Background

Here we provide a short overview of some of the literature we read while working on this project. This is by no means a comprehensive literature research; instead we want to offer a quick overview of previous research conducted and include references that might help other students continuing with this project.

Yasseri et al. [1] investigated the dynamics of editorial wars in Wikipedia, identifying a clear relationship between conflict and burstiness of activity patterns. They discovered three distinct developmental patterns for article behavior over time and found that edit wars are primarily fought by a small number of editors. Their work provides insights into the temporal characteristics of conflicts in Wikipedia.

Kiesel et al. [2] conducted a spatio-temporal analysis of Wikipedia vandalism, revealing significant differences in vandalism activities across time of day, days of the week, seasons, and countries of origin. They found that vandalism rates are typically highest during non-summer workday mornings, suggesting a link between vandalism and labor-related stress or boredom.

Kalyanasundaram et al. [3] developed an agent-based model of edit wars to study factors influencing consensus formation. Their model demonstrated that increasing the number of credible agents and those with a neutral point of view decreases the time to reach consensus, while equal proportions of agents with opposing views lead to longer durations.

Ruprechter et al. [4] examined the relationship between article quality, edit behavior, and link structure in Wikipedia. They found that conflicted articles differ substantially from others in most metrics and that controversial and edit war articles often span structural holes in the Wikipedia network. Their work highlights the potential of using edit behavior patterns and network properties to predict conflict and article quality.

Tinati et al. [5] applied an information cascade model to Wikipedia edit data, revealing insights about the structure and emergence of socio-technical phenomena. They demonstrated that constructing information cascades based on editing activity can create an alternative linking structure between articles, which is often relevant in topic and timely in relation to external global events.

3 Implemented Metrics

The metrics we implemented and analysed within our case study are based on paper by Yasseri et al. (2012) [1] or our own.

3.1 Metrics from literature

Controversiality as defined by Yasseri et al. (2012) [1] is positively influenced by two factors: number of reversions - E and the count of edits made by people involved in reversions - M_R (lower number from the reverted-reverter pair is chosen in order to give lower weight to vandalism).

Yasseri et al. (2012) [1] detect edits that constitute reversion by comparing article texts in time and looking for identical versions. In recent years, it would also be possible to search for tags *undo* or *reverted* in version history, these tags have, however, only been implemented since 2020-21. We decided not to analyse this metric in our case study due to computational complexity.

Edit frequency is the average time between two subsequent edits. An edit war is usually associated with high number of edits, which would decrease the average interval. However, we cannot apply this metric in the case study as some of the articles had much higher total counts of edits simply due to popularity and volume of existing research in the topic. (ex. Robert E. Lee vs Joseph Wheeler).

Burstiness measure tries to summarise the tendency of edit behaviour to appear in bursts of activity with short time intervals between subsequent edits separated by longer periods of infrequent activity versus edit activity appearing in regular intervals.

Yasseri et al. [1] and Goh and Barabási [6] use following formula for burstiness:

$$B = \frac{\sigma_\tau - m_\tau}{\sigma_\tau + m_\tau} \tag{1}$$

where σ_τ and m_τ stand for standard deviation and mean of time intervals between subsequent edits. Since the denominator is positive (σ_τ and m_τ are positive), the sign of burstiness depends on the enumerator:

- Case 1: $\sigma_\tau > m_\tau$ would make B positive, but not higher than 1, as $\sigma_\tau - m_\tau < \sigma_\tau + m_\tau$. The higher spread of intervals compared to mean, the closer will B be to 1.
- Case 2: $\sigma_\tau = m_\tau$ would result in $B = 0$.
- Case 3: $\sigma_\tau < m_\tau$ would make B negative, but not smaller than -1, as $|\sigma_\tau - m_\tau| < \sigma_\tau + m_\tau$. The lower spread of intervals compared to mean, the closer will B be to -1.

3.2 Edit Delta Change Frequency (EDCF)

For every edit that can be obtained from the wikipedia APIs, the current size of the article is recorded. This allows us to calculate a delta to the previous edit, indicating by how much the article grew or shrank in overall size.

Our hypothesis is that there is a higher frequency of additions followed by deletions during times of heavy discourse (as is the case during edit wars). While such addition/deletion patterns are intrinsic to any edit activity, we suspect that the frequency in which additions are immediately followed by deletions (or vice-versa) should be higher during edit wars.

Thus we compute the *Edit Delta Change Frequency (EDCF)* via the following formula:

$$\text{Relative Sign Changes} = \frac{\Sigma SignChangeEdits_t}{\Sigma Edits_t}$$

Where $\Sigma SignChangeEdits_t$ corresponds to the number of edit deltas with a different sign to the previous edit delta in a given time period t. $\Sigma Edits_t$ is the total number of edits in that time period. This metric can be defined for any arbitrary time period t, in our case we went for a monthly analysis.

3.3 Correlations in edit activity

All three topics we selected have had varying degree of attention during the past years. For example, Ukraine has been at the spotlight of world media attention during the peak Euromaidan protests and the initial Russian invasion in 2014, and the interest has been even greater after the full-scale war started in February 2022. The conflict in Israel and Palestine is low-level most of the time, but escalates every few years such as in 2009, 2014, 2021 and in 2023-24. Public discussion about the meaning and legacy of American Civil War is often triggered by incidents of racial violence or during election campaigns.

To investigate the potential connection between real-world events and edit activity on Wikipedia, we propose measuring correlation of monthly edits between pairs of articles. There are 131 unique articles or 371 unique articles if language is also considered in our sample dataset, resulting in 68 635 possible language-article pairs. The amount of combinations is large enough to have some highly correlated pairs just due to chance. Hence, to establish whether there is a relationship between edit activities on articles belonging to the same topic, we propose following method:

- Obtain monthly edit counts for each unique article of a language version,
- Normalize the monthly counts by subtracting mean and dividing by standard deviation,
- Calculate correlation between normalized time series of edits between each pair of articles in a language versions. If the time frame is different due to one of the articles being older, only consider the intersection time span,
- Split the calculated correlations into two groups: high and low correlations, threshold being 0.4,
- Split the pairs of articles in language versions into four groups : same article, but in different languages, both articles belong to same topic, both articles are in same language and same topic, and rest,
- Compare shares of article pair groups in high correlation category to the shares in low correlation category

3.4 Protection-related metrics

A protection generally limits a group of users from performing modifications. Such a measure is often placed on articles that might trigger disputes. A **type** of protection specifies the action the protection limits: editing, moving or creating a page or uploading a new version of a file [7]. A **level** of protection specifies the group of editors limited by it. For example, a *semi-protection* restricts unregistered or unconfirmed users from modifying, an *extended confirmed protection* allows editing only to registered users with more than 500 edits and account older than 30 days, and a *full protection* prevents all users who are not administrators from making changes [7]. Due to data quality related issues, we considered all kinds of protections to be equal, some of the earlier logs do not specify type or level of protection.

A protection has an expiry or it is set as indefinite. For the purposes of our case study, we classified protections as **short** (expiry in 30 days or less), **medium** (expiry in 180 days or less) and **long** (longer than 180 days). We were interested in how did preference for protection length evolve in time and whether there is a difference between topics/language versions.

We defined **protection level** of an article as $\frac{t_p}{t}$, where t_p stands for time the article was protected within time t. In our case study, we aggregated this metric based on language version or topic in order to investigate differences between categories and evolution in time.

4 Case Study

4.1 Topic and language selection

We have decided to analyze articles connected to topics which have recently sparked discussions and controversies online and offline in the public of many countries across the world. Due to interests or beliefs, there can be potential for edit war, misinformation or vandalism.

After some consideration, we decided on two topics **War in Ukraine** and **War in Gaza**. Both were still on-going while we were working on this project in May-July 2024 and had been capturing popular attention and headlines in the news outlets. Public discourse has been heavily polarizing even in countries not directly involved in the wars with disagreements on questions such as "Which party are the 'good' ones?", "Should we participate in form of sanctions or foreign aid?", "How much should we get involved?", "How should we call people who disagree with us?"

As both of the wars are still on-going, a lot of edit activity on articles connected to the topics stems from updating the information in real-time. Such articles are not the focus of the case study. The criterion is that the subject matter of the article must not be recent, but have connection to the selected recent topic. Hence, there is no objective reason for edit activity to rise besides the topic being in the spotlight.

For example, article on Ukrainian president Volodymyr Zelenskyi is expected to attract both legitimate edit activity (updates) and illegitimate activity (misinformation and vandalism). However, article on a leader of Ukrainian Cossacks Bohdan Khmelnitskyi who lived in 17th century should not attract much edit activity for the sake of updating information, and newly discovered information is not correlated with the current world events.

The controversies and conflicts related to the topics we selected are not limited to different opinions about the current world events (Is Israel justified in invading Gaza?) or opinions whether something is happening (How high are Russian causalities so far?). They spread to topics that might be hundreds of years old (Did Russians or Ukrainians invent Borscht?) or have any relation to the topic (as an example, articles on Hebrew and Ukrainian language on English Wikipedia are placed under protection as of June 2024).

We selected one more topic - **American Civil War** for our analysis. The attention to the topic might have been higher during some of the recent events (2016 and 2020 Presidential elections, Black Lives Matter protests), but the topic has been discussed at length for the past 150 years. One of the main disputes concerns the origin of the war, with Confederate apologists downplaying slavery and stressing the States' rights. The commemoration of the war is also divisive, there are statues or buildings and streets named after Confederate generals and politicians in the US South, and the conduct of Union and South leaders during the war has also been a topic of heavy discussion (Did Lincoln have right to suspend habeas corpus? Did Robert E. Lee commit war crimes?).

For **article selection**, we went to Wikipedia portals on the three topics. For American Civil War, we selected featured articles on the portal related mostly to the leaders, battles or origins of the war (slavery, abolitionist movement, states' rigths, etc.). For the War in Ukraine and War in Gaza, due to the non-recent article limitation, we avoided articles about people who are alive or ongoing events. From the Wikipedia portals on Ukraine, Israel and Palestine, we selected featured articles mostly related to culture and history.

4.2 Dataset

Our final dataset contains information on 131 articles, all of them in English, and most of them in at least another language version. Detailed breakdown is shown in the table below:

Topic	Language	Number of articles
Ukraine	English	43
Ukraine	German	41
Ukraine	Ukrainian	43
Ukraine	Russian	42
Israel/Palestine	English	43
Israel/Palestine	German	34
Israel/Palestine	Arabic	42
US Civil War	English	45
US Civil War	German	38

Table 1: Number of articles per topic and language

We were using two API services Wikipedia provides - MediaWiki API [8] and WikiMedia API [9]. During the course of implementation and testing of data extraction and transformation pipelines, we encountered numerous problems. Some were related to parameters of interest not existing before certain time, others to differences and peculiarities of some language versions [1]. We aimed to have a result that would be universal and replicable, but it is possible that future extractions will require further fine-tuning of our implementation, especially when extracting new articles from new language versions. For the purposes of our case study, we pulled all historical data connected to the article belonging to three categories.

Article edits table contains attributes timestamp, editor's user name, comment, size of article after the change and reversion/reverted flags. The reversion and reverted flags indicating if the edit was a reversion

[1]For example, many Slavic languages use genitive case for dates. April 20th in Slovak is 20. apríl (dvadsiaty apríl), but within a context, genitive is usually applied : 20. apríla (dvadsiateho apríla) - *of the twentieth April*. Unfiortunately, as of today, this liguistic feature, also present in Czech, Ukrainian, Serbian or Russian languages is not accounted for in the Python datetime library.

of a previous edit or if the edit was reverted by one of following edits are only available from 2019 and 2020, respectively. In total, we pulled and processed 564 467 unique observations. Interestingly, nine out of ten most edited articles are in English and related to US Civil War, starting with articles on Ulysses S. Grant (21 238 total edits) and Abraham Lincoln (18 354 total edits).

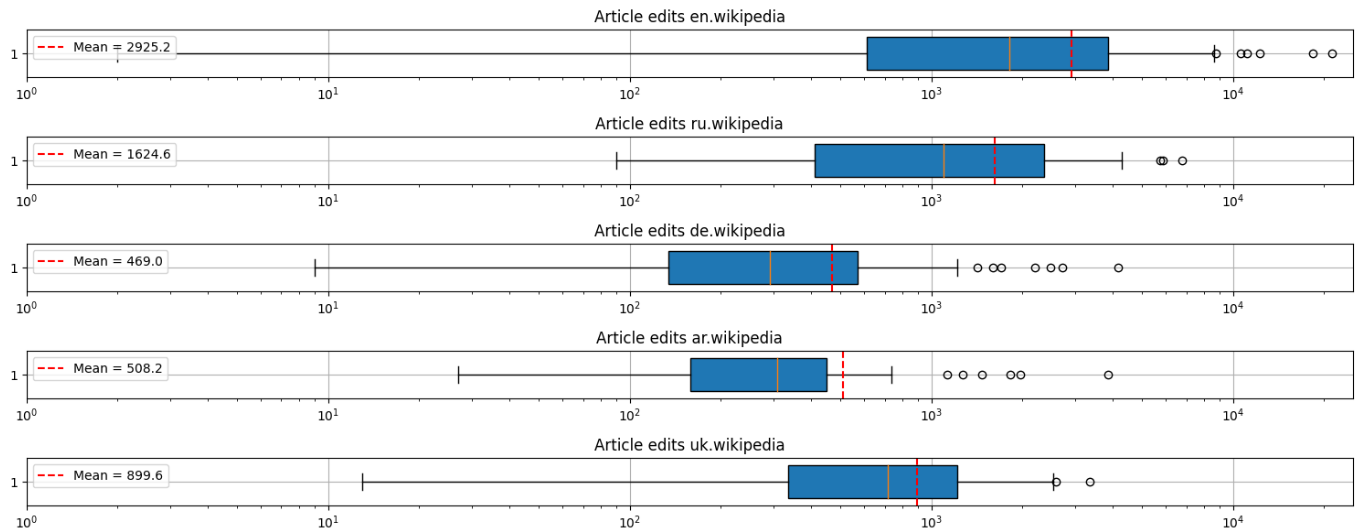

Figure 1: Boxplot of article edits count by language version

In the combined dataset of all articles, the majority of edits was conducted by registered users, however since we deliberately articles that were placed under protection this is to be expected — See Figure 2.

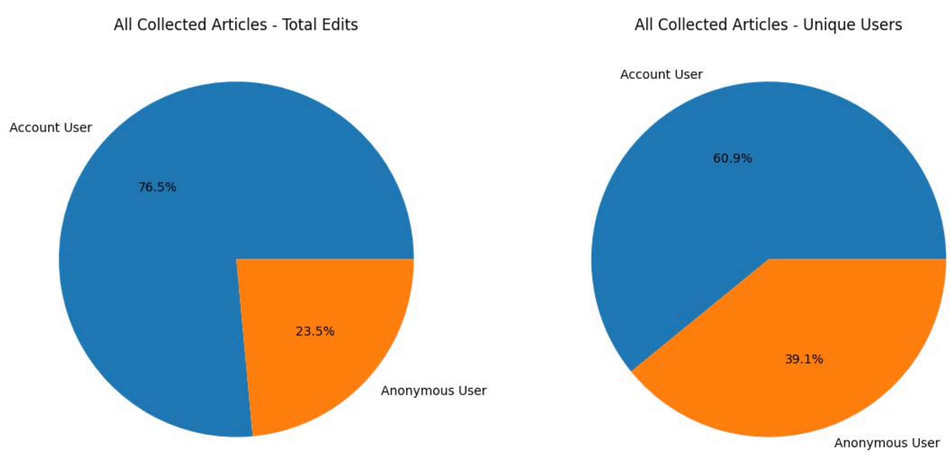

Figure 2: Piechart of user types, total edits and unique users

Protection logs of an article contain attributes timestamp, action type (protect, unprotect, modify, etc.), expiry. There are four protection types that can prevent an article from being edited, moved or renamed, created or a file from being uploaded [7].

A protection can have different levels based on who is excluded from making changes. In the strictest form, only administrators are permitted to modify, other levels may exclude the unregistered users only or users with short edit history.

4.3 Results

Due to small sample size and manual selection of articles, we do not measure significance of results. Still, there are many interesting observations we can report.

4.3.1 Protections

Articles in our dataset are becoming increasingly protected, with exception of ones written in German or Arabic. Sometimes the rapid increases of protections happen around real-world related events, for example an increase of protections of articles on US Civil War topics is visible in summer 2020, articles related to Ukraine have been protected more after 2022.

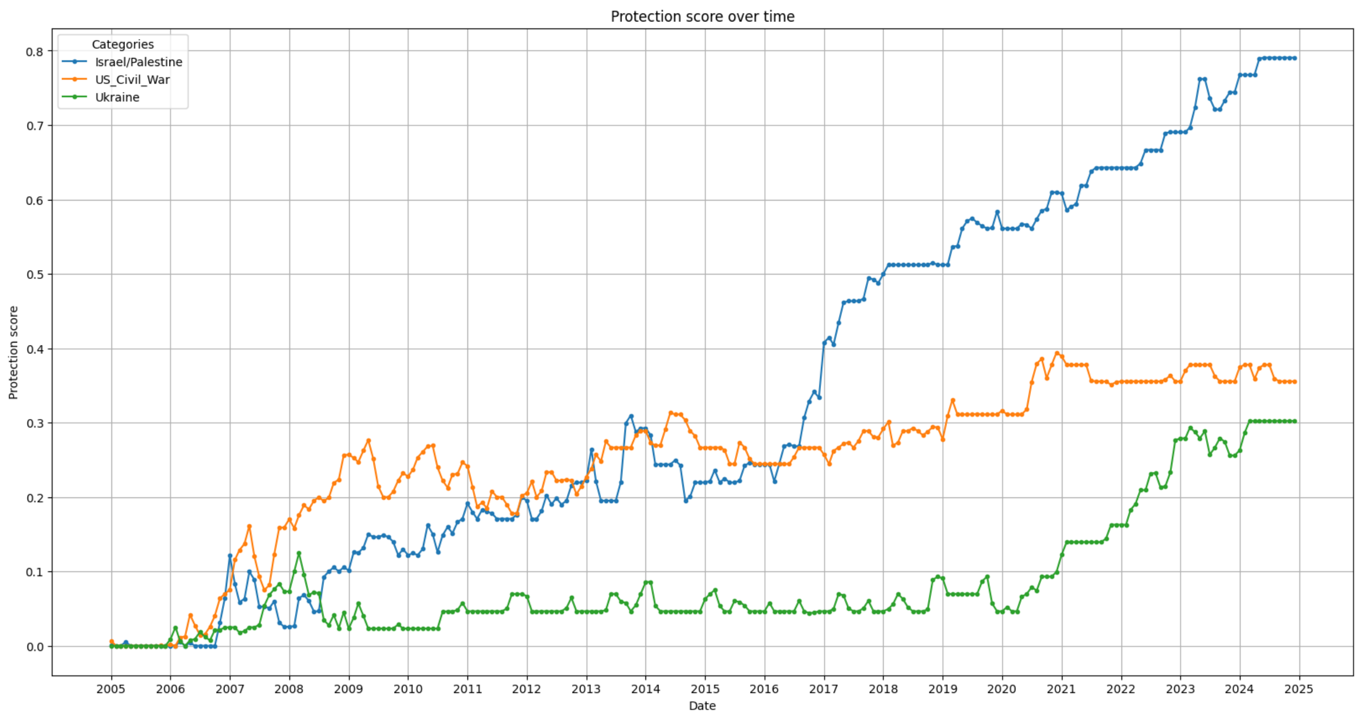

Figure 3: Protection level of English language articles in our sample over time

There is a difference between English and other language version articles when comes to length of protection. On English Wikipedia, since 2009, many articles have been put under long or permanent protection, and therefore, there has been lower need for shorter length protections, due to controversial articles being already protected. On other language versions, short protections are still a frequently used tool for moderation.

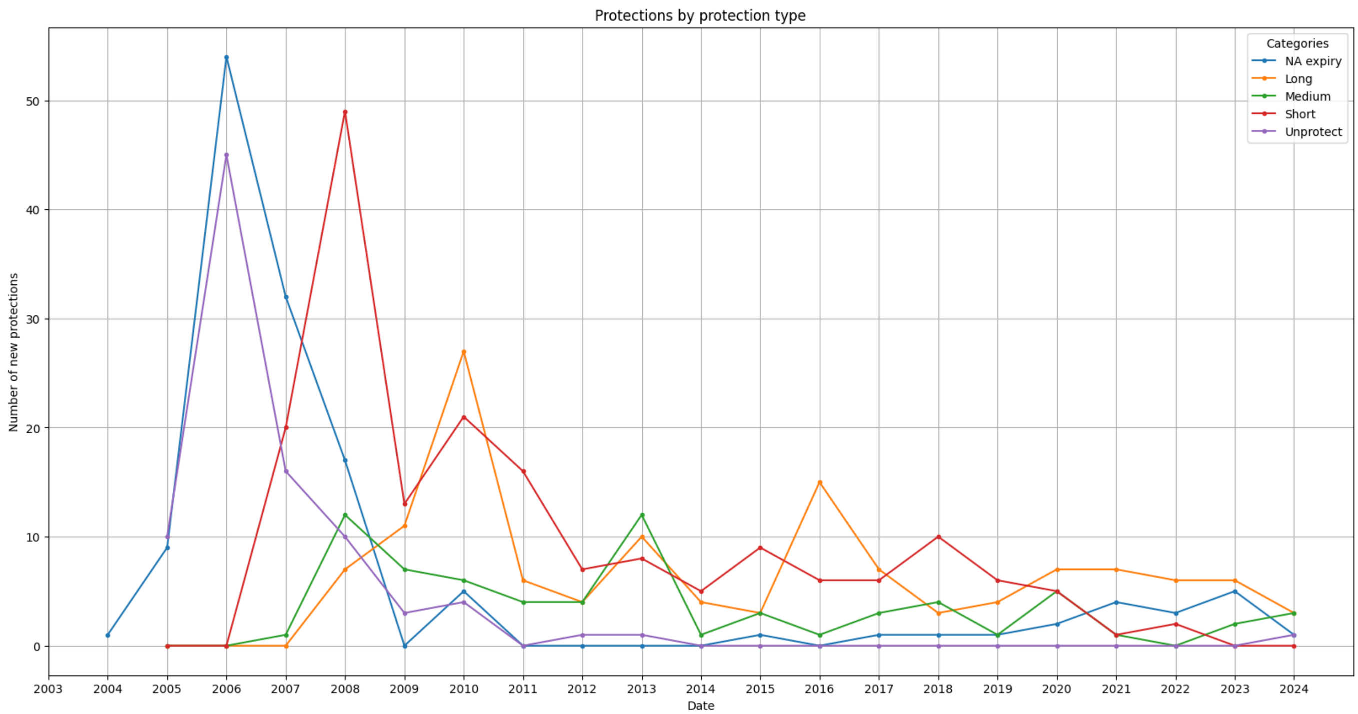

Figure 4: Lengths of protections enacted on English language articles by year

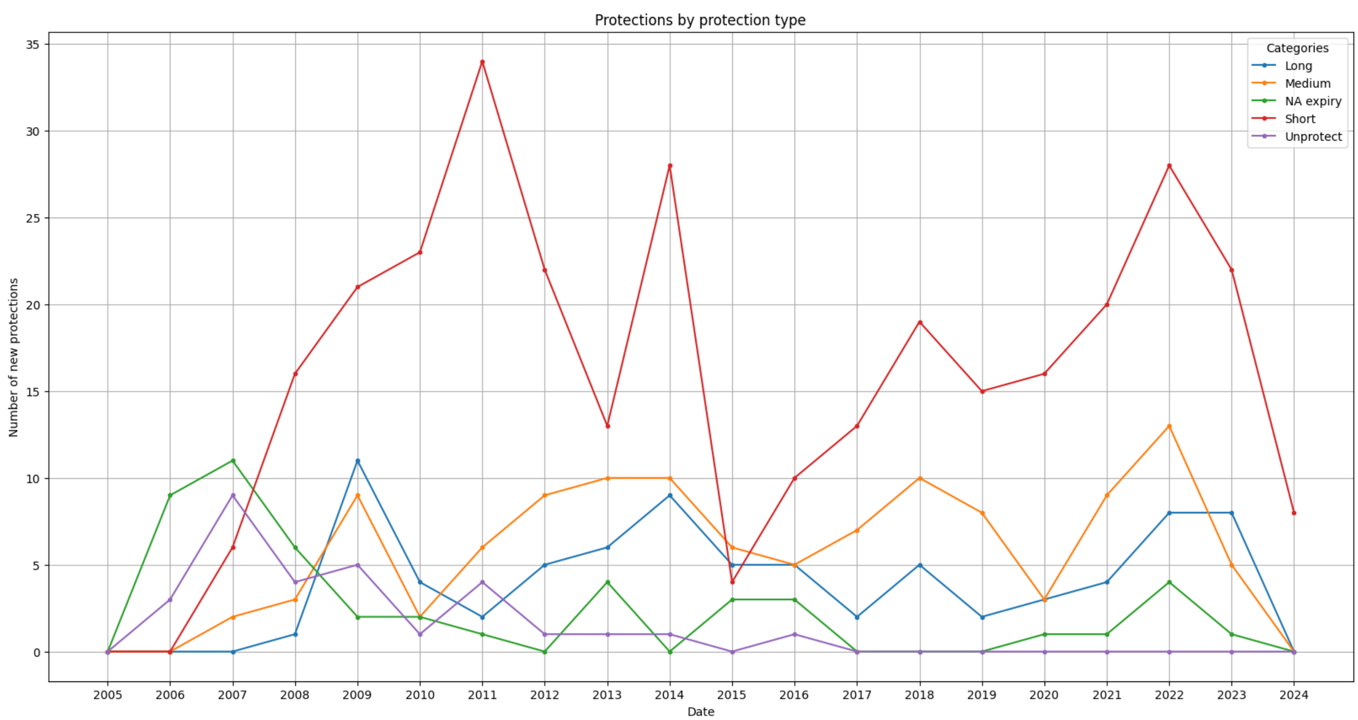

Figure 5: Lengths of protections enacted on other than English language articles by year

Trend being likely related to the increase of articles under protection over time, the **burstiness**, if calculated on yearly intervals has decreased, and then stagnated.

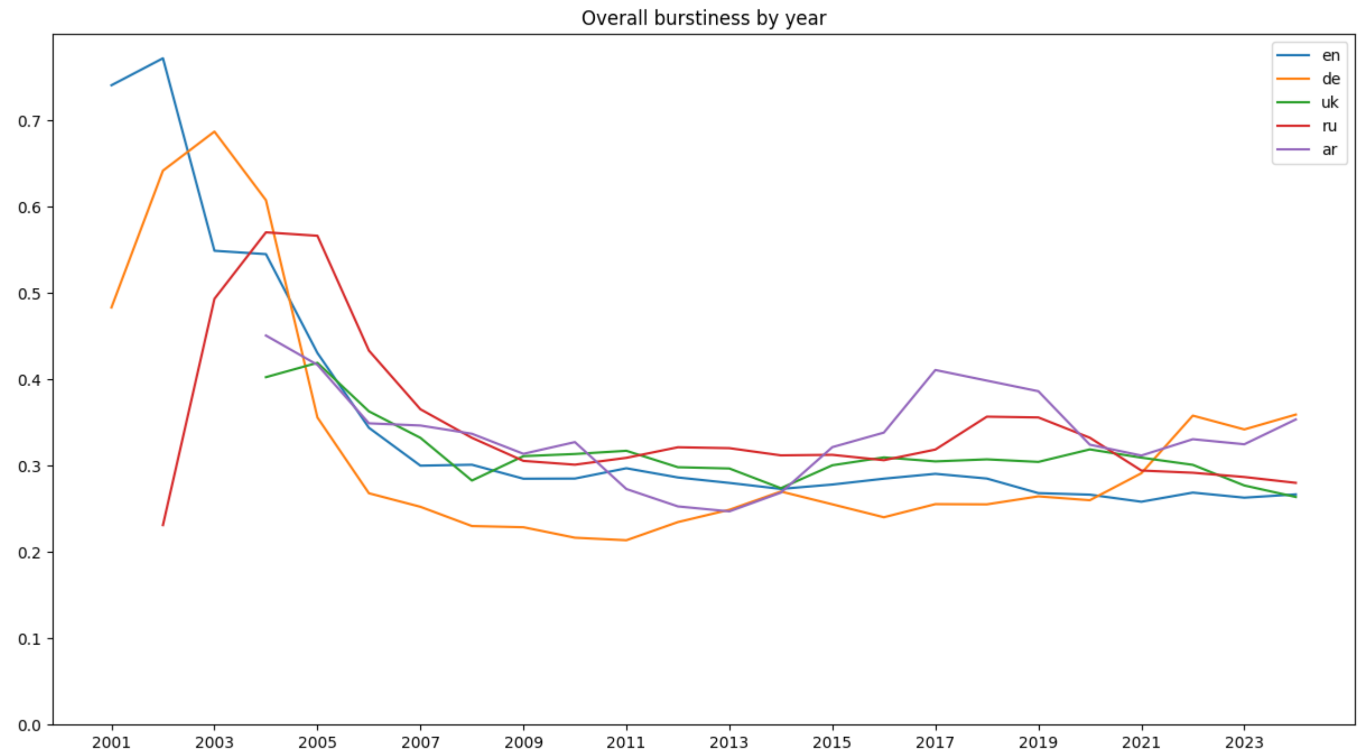

Figure 6: Burstiness in time by language version

4.3.2 Burstiness

Burstiness of articles ranged from 0.2 to 0.7, meaning all editing activity could be classified as bursty, albeit in a different degree. This range is interestingly very similar to the one observed by Yasseri et al. [1].

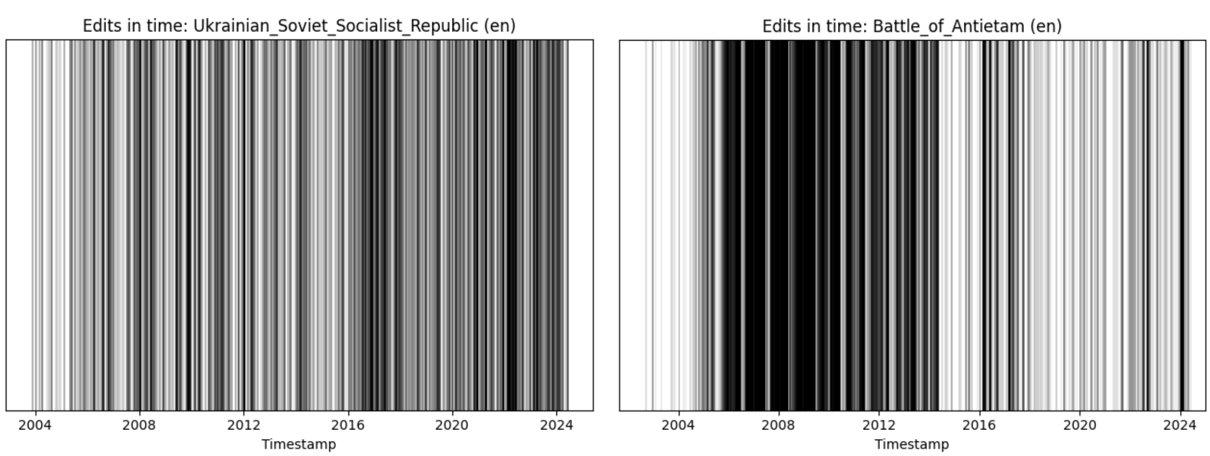

Figure 7: Each line represents an edit, article on left has burstiness 0.29, on the right 0.62

4.3.3 Edit Delta Analysis

To visualize the Edit Deltas, we continue with the examples used for Burstiness: the English language articles on the Battle of Antietam and the Ukrainian Soviet Socialist Republic.

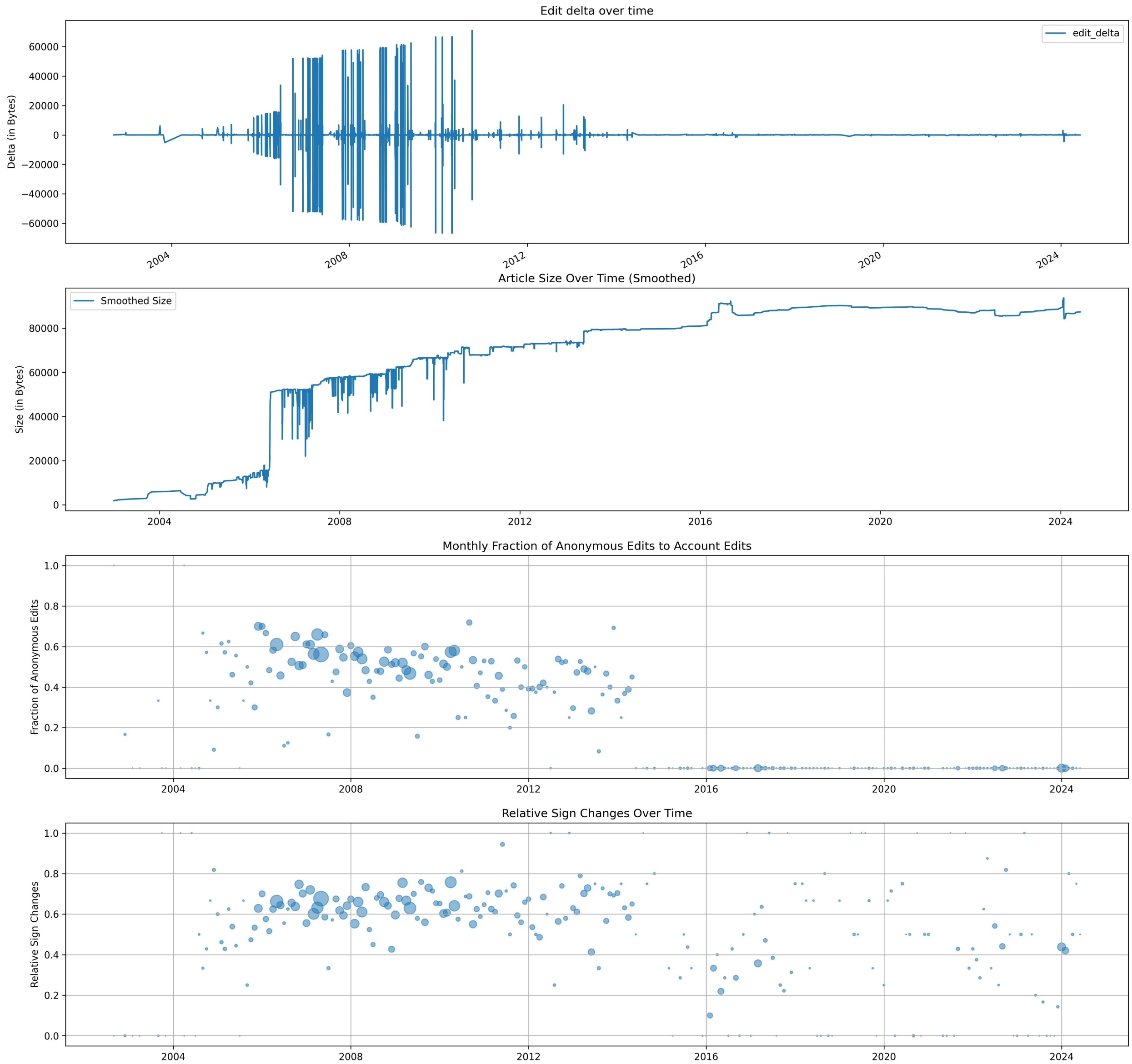

Figure 8: Edit frequency, article size, anonymous users & EDCF over time — Civil War

Figure 8 Shows the compilation of various time series plots. Plot one shows the Edit Deltas (with the y-axis indicating the magnitude of each edit. This plot closely mirrors the patterns of burstiness shown in the previous section. It can also be clearly observed in all of the time-series plots that a protection was placed under protection in 2014, when the edit spikes quickly subside.

The second plot shows the articles growth over time, illustrating the limited amount of growth experienced between 2006 and 2014 even though there was considerable edit activity.

Another visualization we were interested in is shown in the third plot, which shows the fraction of anonymous users relative to registered users in a given month. The size of the dots corresponds to the overall number of edits (as is the case in plot 4). We hypothesize that either a sort of ”brigading” is happening here, i.e. the article drew in a lot of people that usually do not edit Wikipedia and thus have no accounts — or alternatively that Wikipedians want to avoid damaging their reputation when engaging in edit-warring behavior. In any case, this fraction falls down to 0 once anonymous users are prevented from editing due to the protection put in place.

Lastly, we show the burstiness of the article, which continues the patterns of the edit war, even after the protection is put in place. This might be either a sign that edit disputes still happen on this article, but on a much smaller scale. Or it could be the case that EDCF is not a suitable metric for analyzing edit wars.

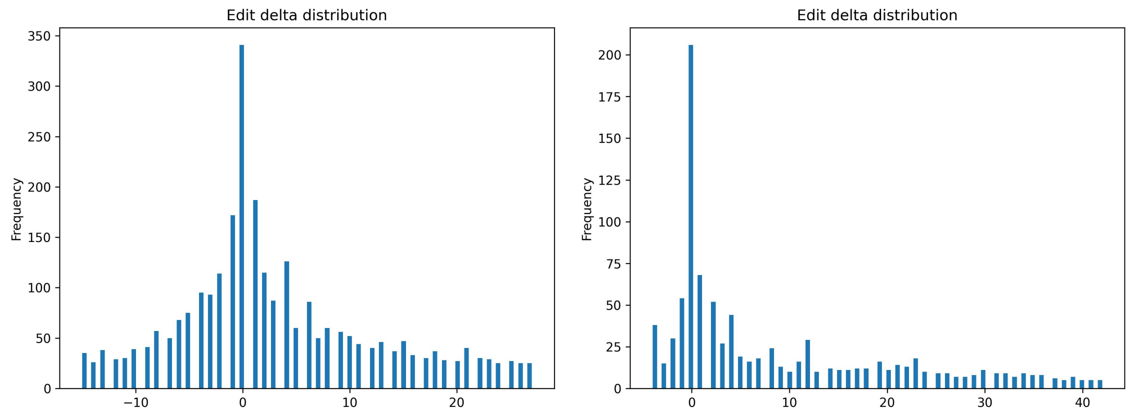

Figure 9: Distribution of edit deltas — Civil War vs. Ukraine

Figure 9 plots the distribution of edit Deltas between the more disputed Antietam articles compared to the USSR article. The distribution itself hints at different growth patterns, which can be also be seen in the timeseries plots (Figure 10) for the Ukraine article. While the Civil War articles edit distribution is almost centered around 0, the edit distribution for the USSR article is clearly right skewed — a result of stable growth during the articles lifetime.

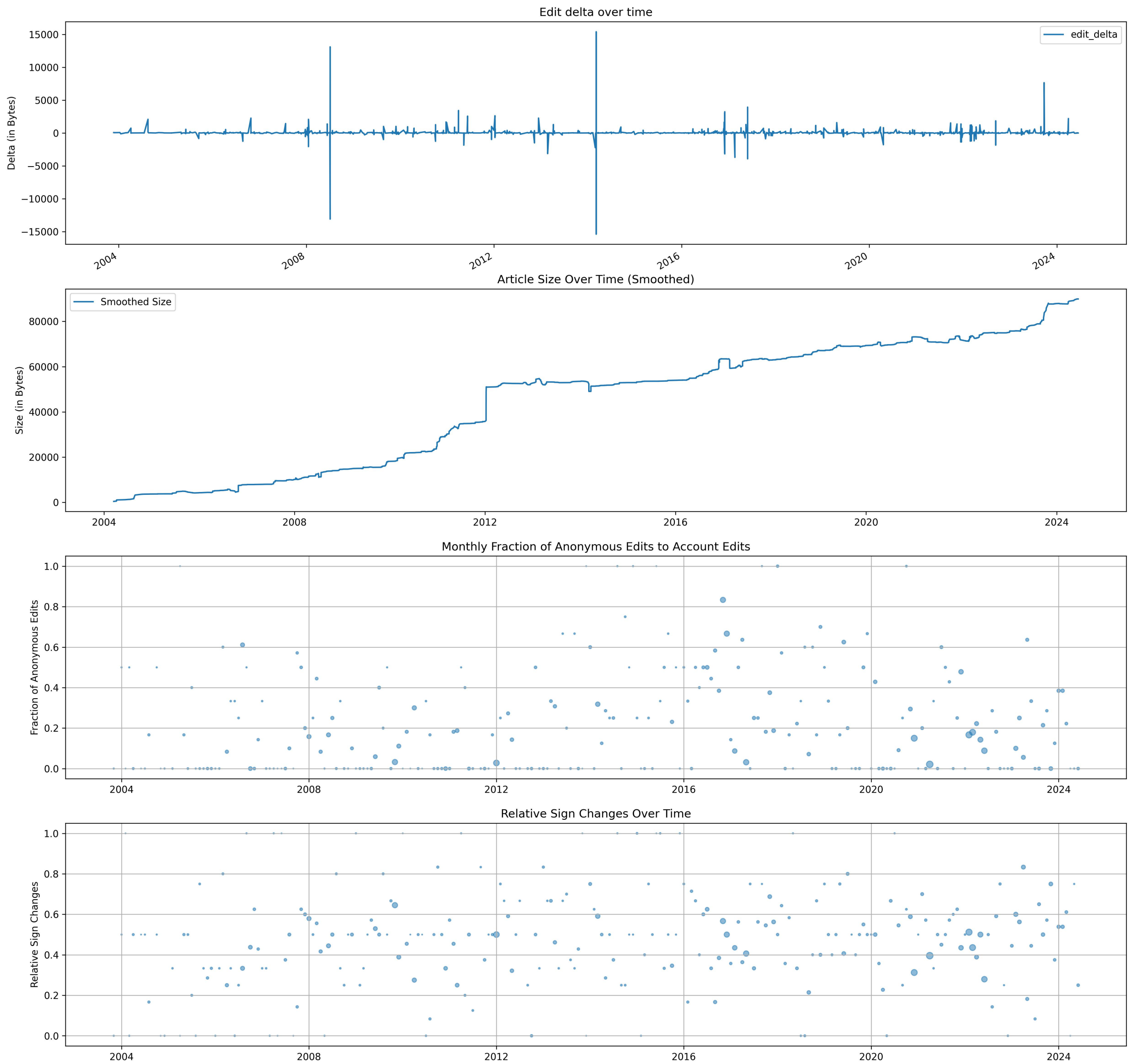

Figure 10: Edit frequency, article size, anonymous users & EDCF over time — Ukraine

4.3.4 Correlation Analysis

We also calculated **correlations** between each pair of articles.[2] Two articles can have same title, but in different languages, be in same language and related to the same topic, be in different languages and related to the same topic, or have nothing in common (perhaps just the language). The *nothing in common* group has lower share between highly correlated articles.

As already mentioned above, we cannot draw conclusions about significance of results due to low sample size. Another issue are bot translations responsible for some of the highest correlations between pairs of articles. And finally, looking at absolute count of articles by group and correlation changes the picture, too.

[2]For the results, we only considered articles that had at least 500 total edits.

Group	All pairs (%)	Highly correlated ($\geq$ 0.4) pairs (%)
Nothing common	64.5	46.3
Same topic	21.4	18.3
Same topic, same language	13.3	32.0
Same article	0.7	3.4

Table 2: Share of article pairs' groups

Group	All pairs	Highly correlated ($\geq$ 0.4) pairs
Nothing common	13096	81
Same topic	4352	32
Same topic, same language	2708	56
Same article	142	6

Table 3: Count of article pairs' groups

5 Conclusion

We see our main **contribution** in following:

- We created a repository with functions to **pull and process** relevant information using Wikipedia API services.
- We **expanded upon metrics** from paper by Yasseri et al. (2012) [1] by also analyzing protections, dynamics of edit activity and correlation between article pairs.
- We **implemented** the metrics in notebooks in the public repository.
- We propose a **methodology to select articles** for analysis of article selection when analysing potential conflict related to current topics.

While very interesting and full of rabbit holes to explore, our case study's main limitation is small sample size. For that reason, we decided not to assess or discuss statistical significance of our results. We believe there is a great potential in an expanded analysis that calculates and evaluates metrics from this project on a significantly larger dataset.

Besides that, we also see potential in improving some of the metrics. The **correlation** measure could also take into account activity in the preceding and following time periods, which could involve measuring cross-correlation at specified lags or redefining time periods based on which edit counts are aggregated. Or, when calculating **burstiness**, edits could be weighed based on the total byte size change.

Follow-up research could also differentiate more in edit and protection data. Due to their unavailability for older periods, we decided to drop attributes such as revert/reversion flag for edits or protection level and type for protections. However, it is possible to conduct analysis including these attributes also with our dataset if limited for the past 4-5 years.

We had a lot of fun working on this project and hope that maybe some other student group will pick up our work. We would like to thank our wonderful supervisor Sabrina Kirrane for providing the space for us to be creative and conduct this project at our own pace.

6 References

References

[1] T. Yasseri, R. Sumi, A. Rung, A. Kornai, and J. Kertész, "Dynamics of conflicts in wikipedia," *PloS one*, vol. 7, no. 6, 2012.

[2] J. Kiesel, M. Potthast, M. Hagen, and B. Stein, "Spatio-temporal analysis of reverted wikipedia edits," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 11, 2017, pp. 122–131.

[3] A. Kalyanasundaram, W. Wei, K. M. Carley, and J. D. Herbsleb, "An agent-based model of edit wars in Wikipedia: How and when is consensus reached," in *2015 Winter Simulation Conference (WSC)*, IEEE, 2015, pp. 276–287.

[4] T. Ruprechter, T. Santos, and D. Helic, "Relating Wikipedia article quality to edit behavior and link structure," *Applied Network Science*, vol. 5, no. 1, p. 61, Sep. 9, 2020, ISSN: 2364-8228. DOI: `10.1007/s41109-020-00305-y`. [Online]. Available: `https://doi.org/10.1007/s41109-020-00305-y` (visited on 05/18/2024).

[5] R. Tinati, M. Luczak-Roesch, and W. Hall, "Finding structure in Wikipedia edit activity: An information cascade approach," in *Proceedings of the 25th international conference companion on world wide web*, 2016, pp. 1007–1012.

[6] K.-I. Goh and A.-L. Barabási, "Burstiness and memory in complex systems," *Europhysics Letters*, vol. 81, no. 4, p. 48 002, 2008.

[7] WikiMedia, *Protection policy*. [Online]. Available: `https://en.wikipedia.org/wiki/Wikipedia:Protection_policy` (visited on 05/27/2024).

[8] MediaWiki, *MediaWiki API*. [Online]. Available: `https://www.mediawiki.org/wiki/API:Main_page` (visited on 05/26/2024).

[9] WikiMedia, *WikiMedia API*. [Online]. Available: `https://api.wikimedia.org/wiki/Main_Page` (visited on 05/26/2024).