Wikipedia Edit Wars

Riško Hrivnák & Paul Straberger

28.06.2024

What is an edit war?

In general, **anyone can modify Wikipedia** articles

Often, a history of an article reflects real-world differences in opinions and beliefs

Also, the level of edit activity often reflects real-world events

How can an edit war be prevented?

Different levels of protections that can be set by only by administrators, although any user may request protection

A protection excludes a certain group of users from making a certain type of change

3 revert rule: Editors must not exceed three reverts on a single page within 24 hours.

Icon	Mode
White	Pending changes protected
Silver	Semi-protected
Blue	Extended confirmed protected
Pink	Template-protected
Gold	Fully protected
Red	Interface protected
Green	Move protected
Skyblue	Create protected
Purple	Upload protected
Turquoise	Cascade protected
Black	Protected by Office

Research questions

What metrics can be used to analyze edit activity on Wikipedia articles?

What metrics can be used to analyze protections of Wikipedia articles?

How can we implement these metrics in an automated way?

What are the differences in these metrics between the selected topics and language versions?

Article selection

War in Ukraine, **War in Gaza**, **US Civil War**

We avoided articles about topics that have objective reasons for increased edit frequency

List of articles based on the articles featured on the Wikipedia portals

Data extraction and processing

Wikipedia's API services

Three types of data: **edit history of the articles**, **protection logs** and **edit count of users**

Challenges: different languages (Slavic months, Arabic script), veeeery long time (40 hours+)

Summary statistics

In total 131/371 unique articles, 573 950 edits

Edits by language

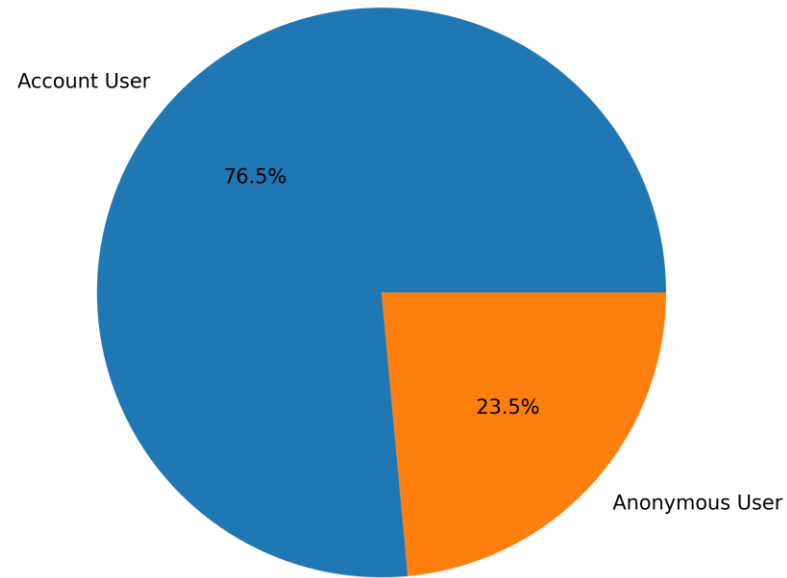| En	Ru	De	Uk	Ar
390 052	68 233	54 539	38 684	22 442

Edits by topic

Israel/Palestine	Ukraine	Civil War
159 328	197 587	217 035

User statistics

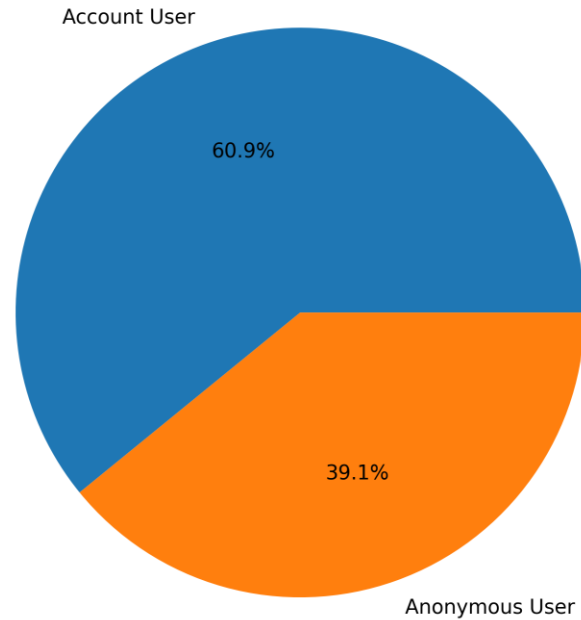Most edits are done by registered users

All Articles - Total Edits

Account User

76.5%

23.5%

Anonymous User

All Articles - Unique Users

Account User

60.9%

39.1%

Anonymous User
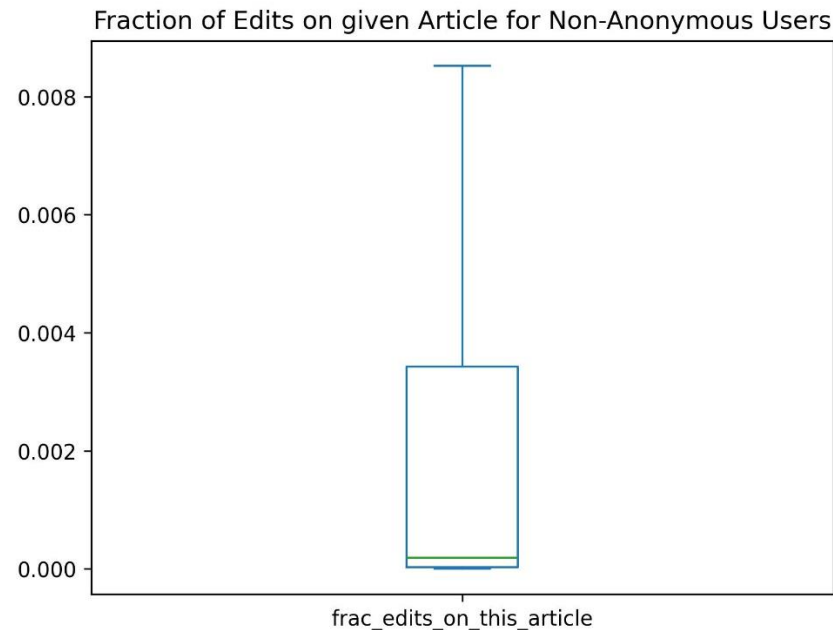
User statistics

Registered users generally edit many different articles

In our data, users had an average of 5592 edits on other articles, compared to ~3 edits on a given article

Fraction of Edits on given Article for Non-Anonymous Users

What did we do?
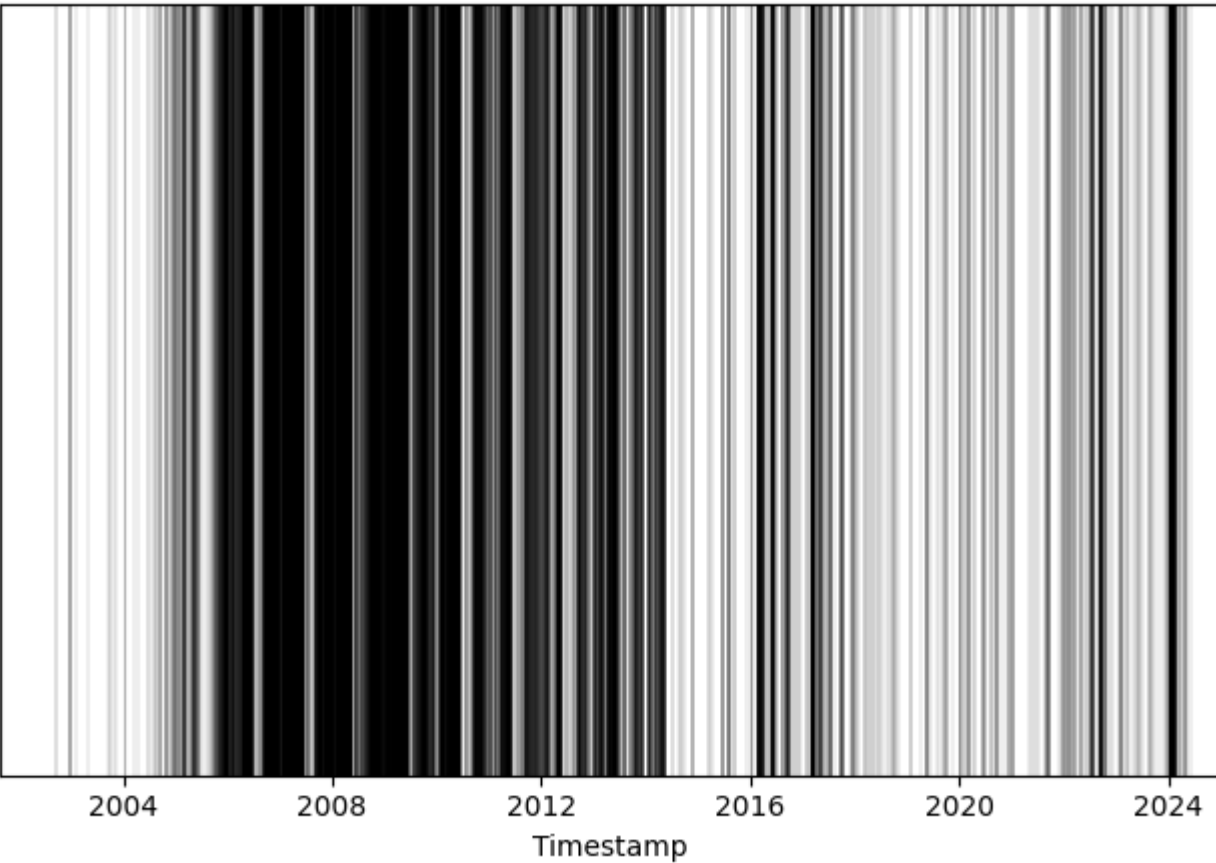
Burstiness of edit activity

Investigate edit deltas

Correlations of edit activities between articles

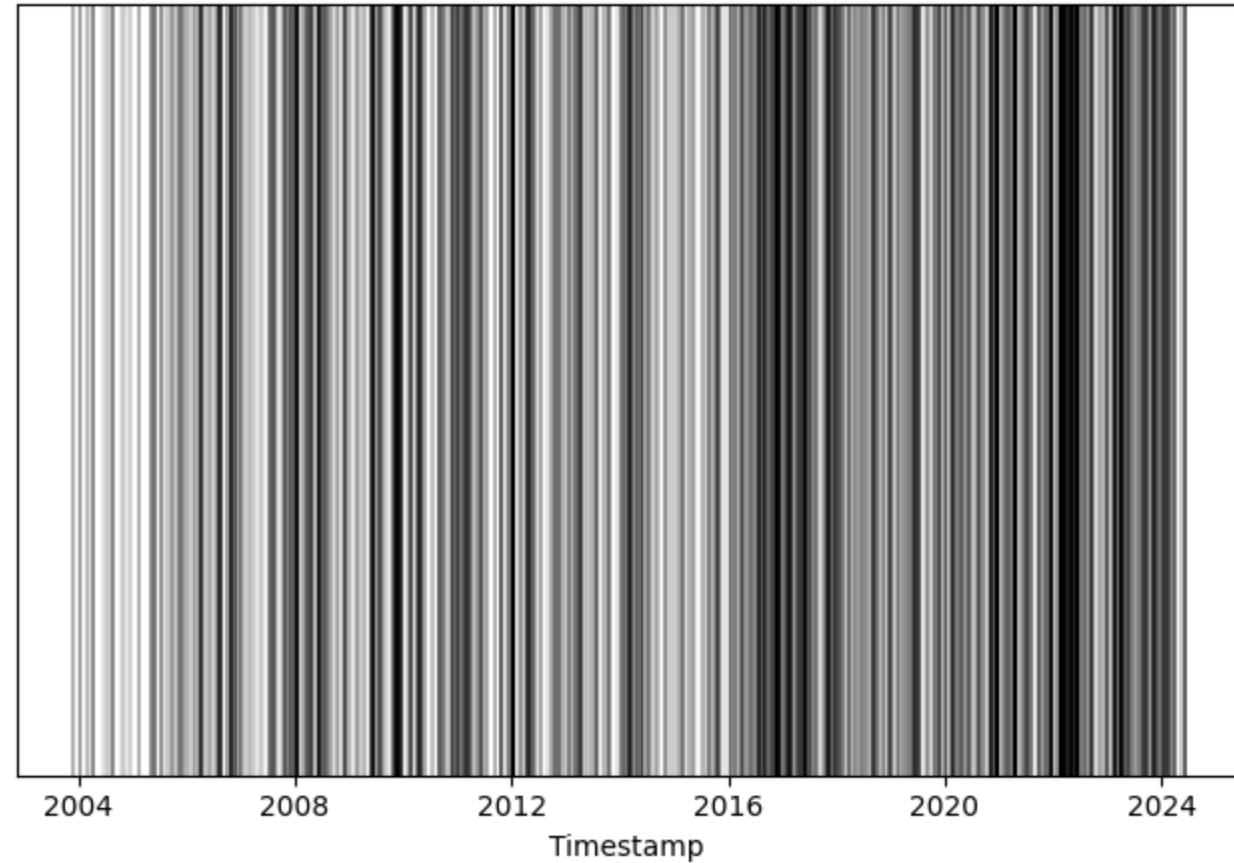Looked at evolution of protection status over time

Burstiness of edit activity

Edits in time: Battle_of_Antietam (en)

Edits in time: Ukrainian_Soviet_Socialist_Republic (en)
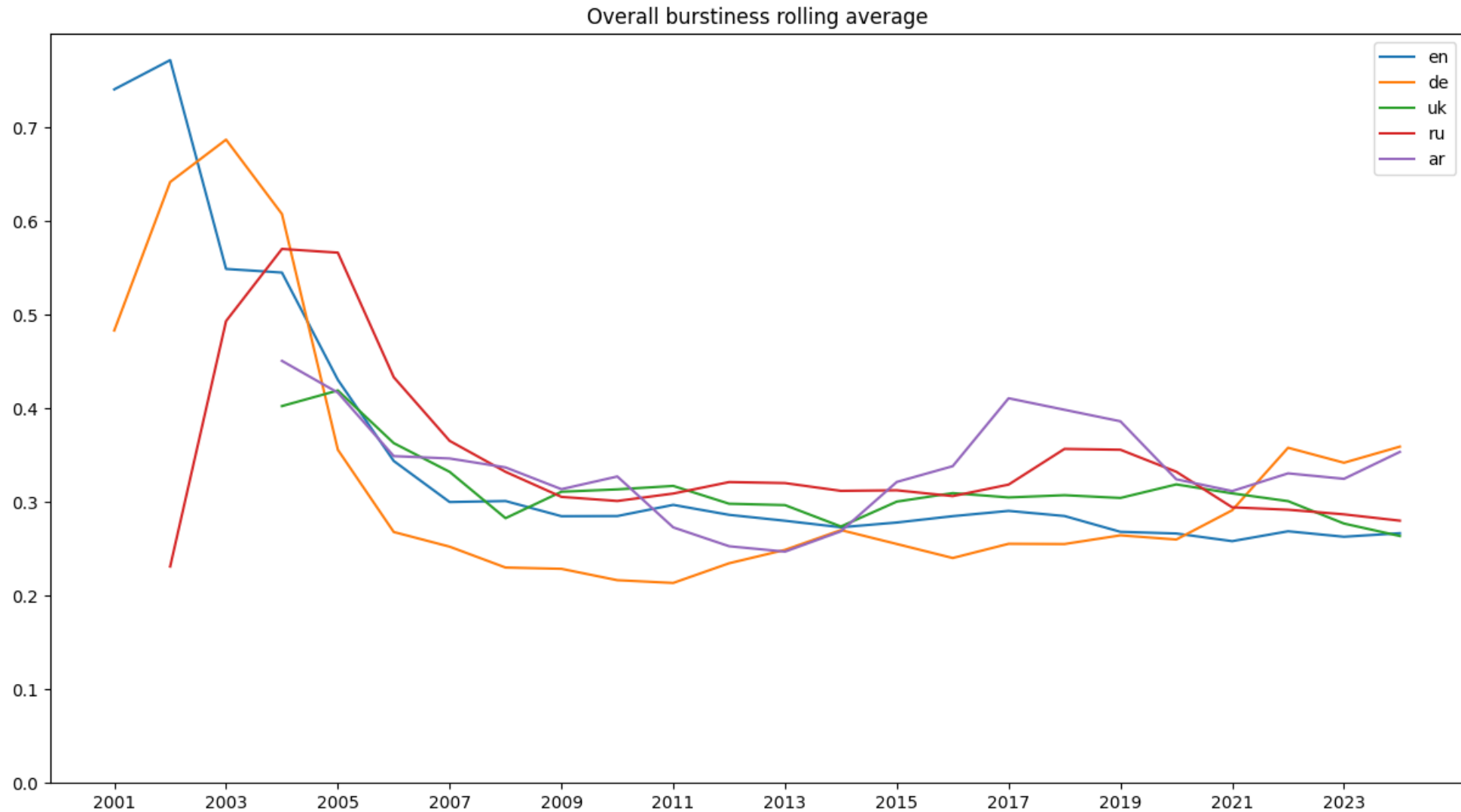
Burstiness 0.62

Burstiness 0.29

Burstiness of edit activity

Inspired by Yasseri et al. (2012)
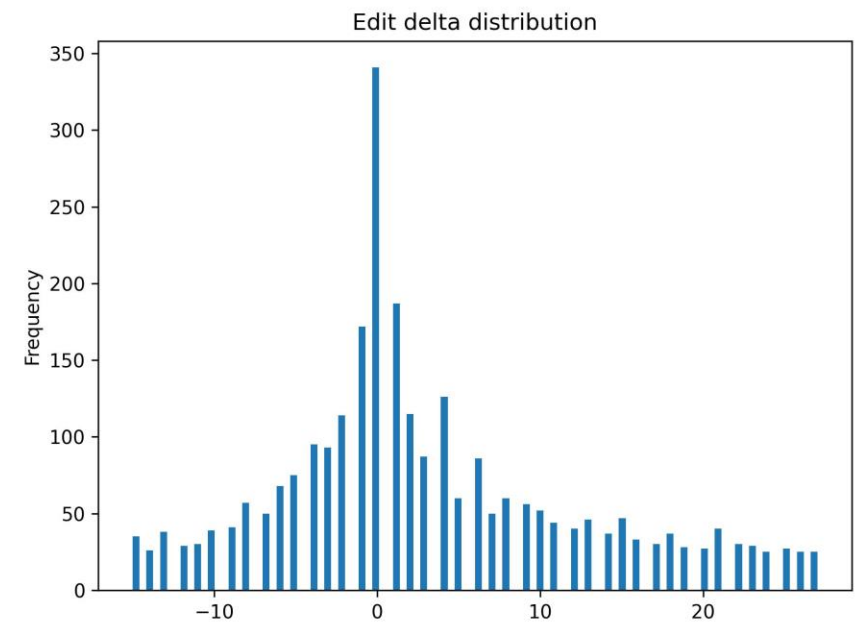
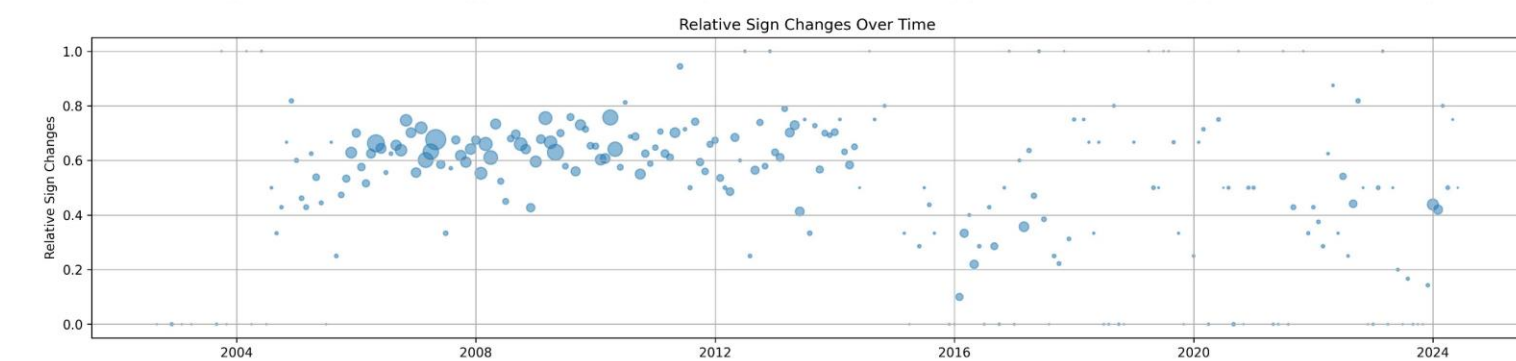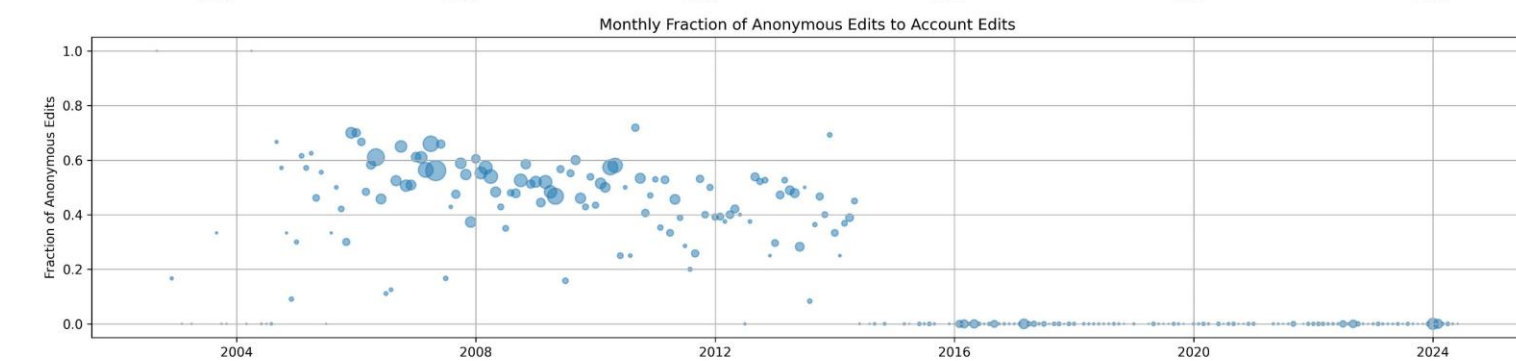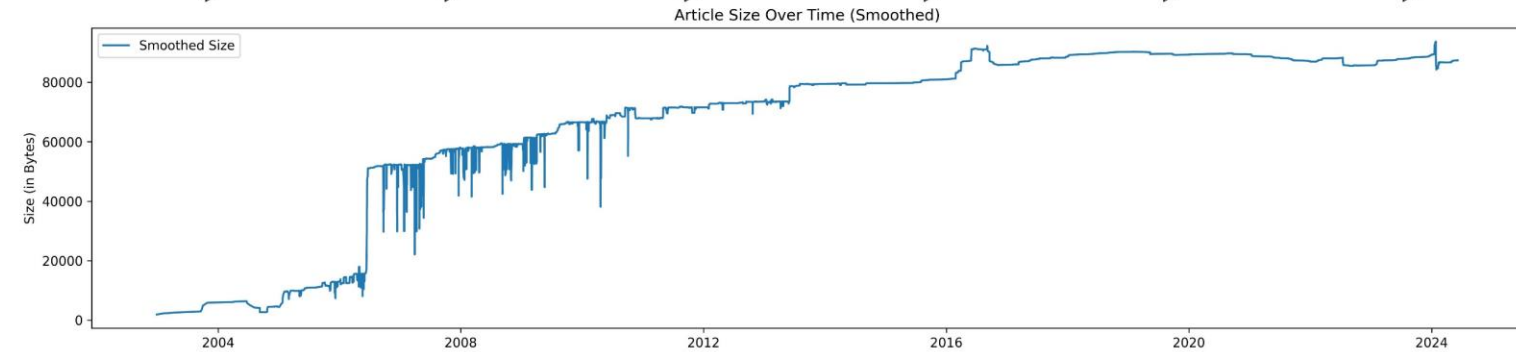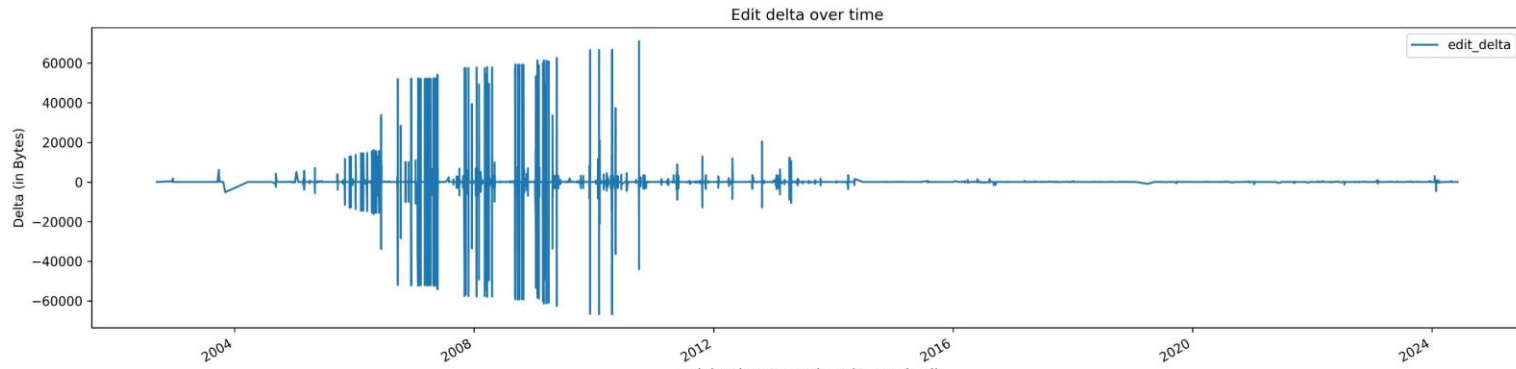$$B = \frac{\sigma_i - \mu_i}{\sigma_i + \mu_i}$$

Could help to find articles and timeframe where an edit war happened

Burstiness of edit activity

Overall burstiness rolling average

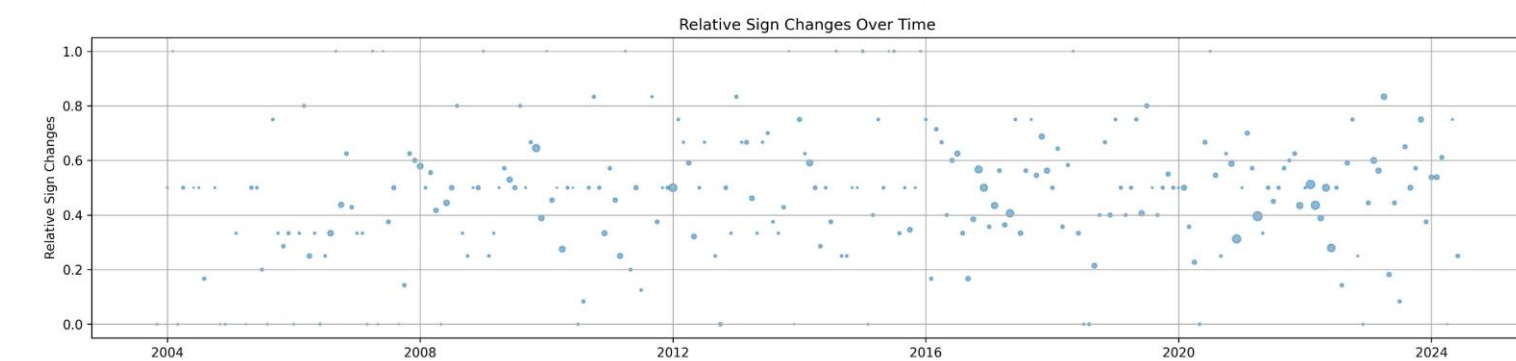# Edit Delta Analysis
– Antietam –

Edit Delta Analysis
– Ukraine –
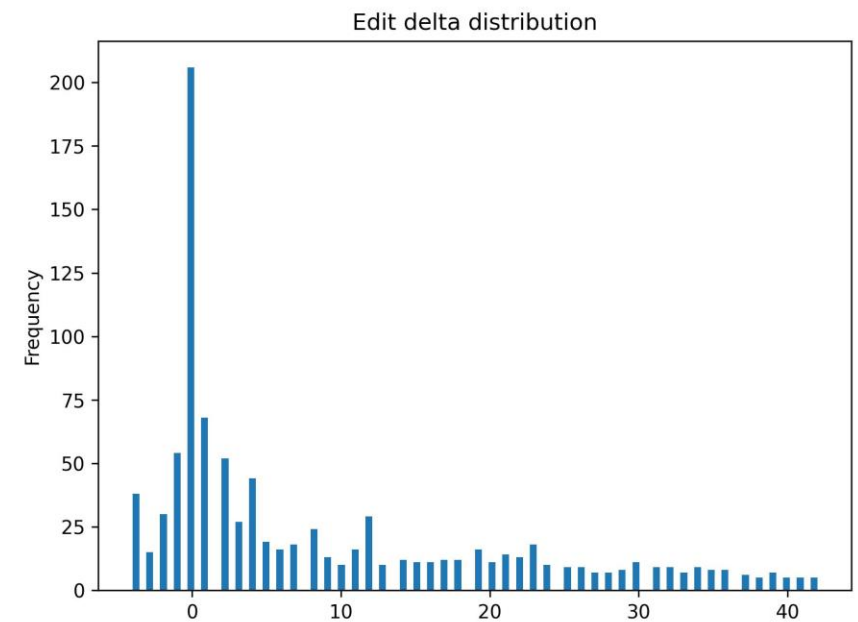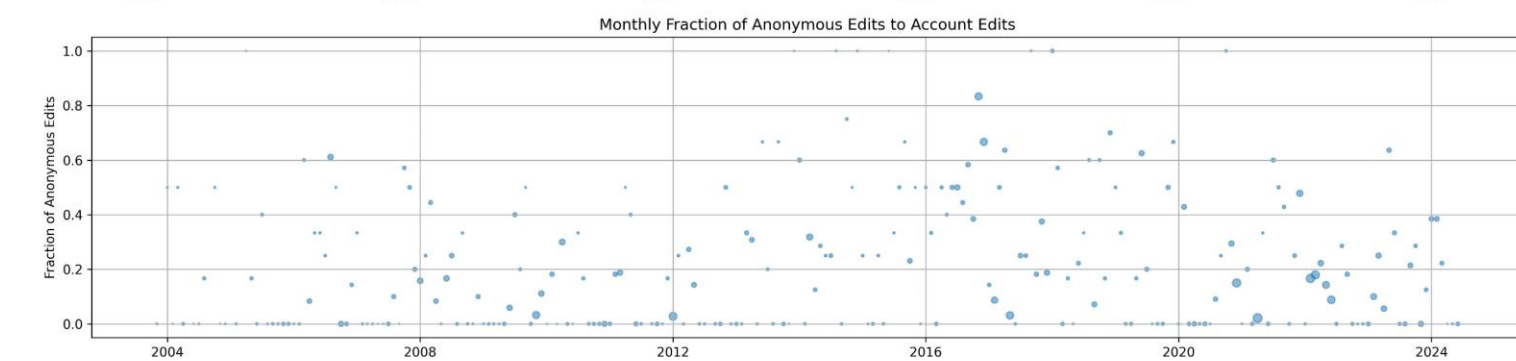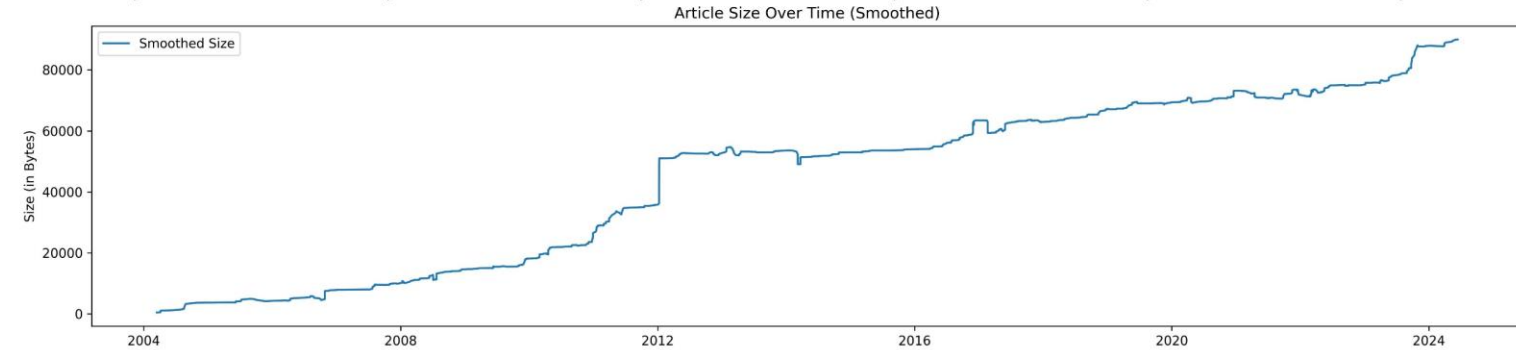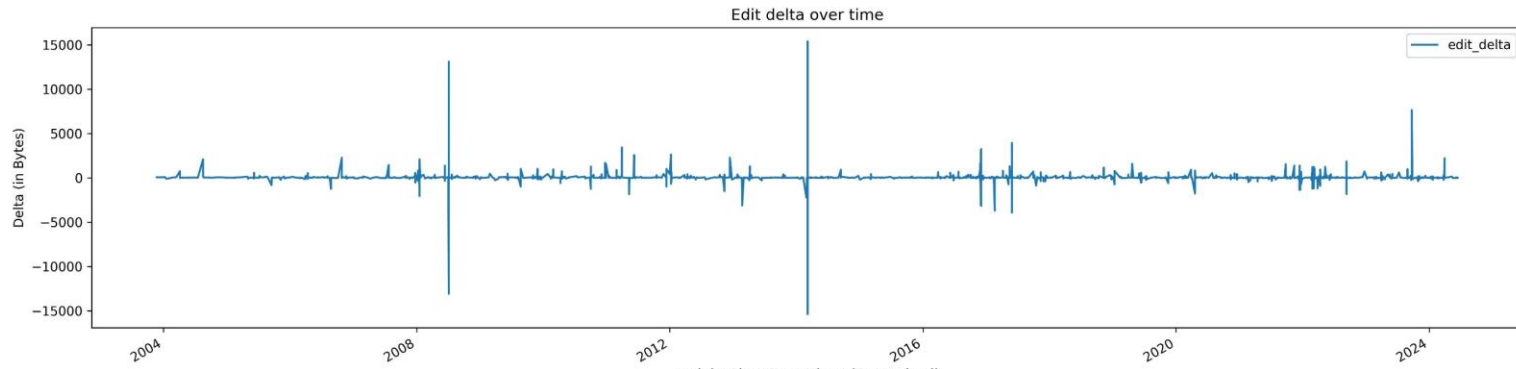
Correlations

For each pair of articles, we calculated correlation of monthly edit counts

Correlations
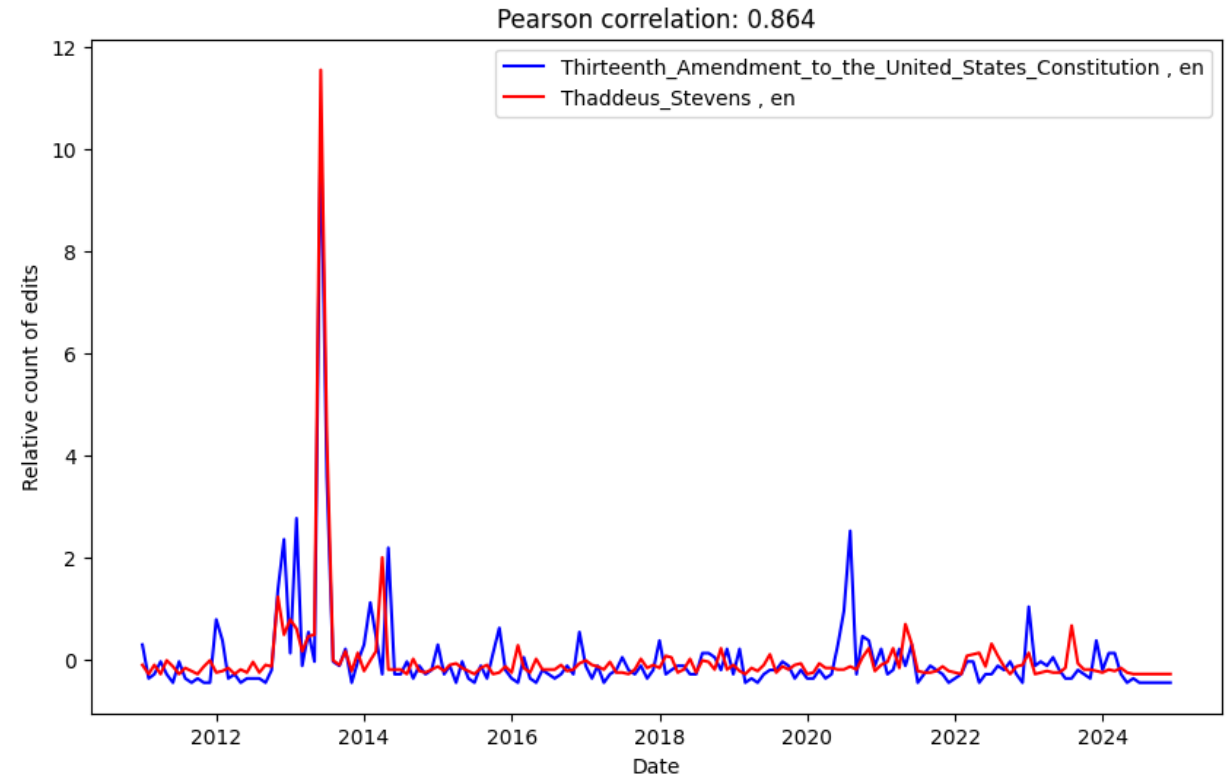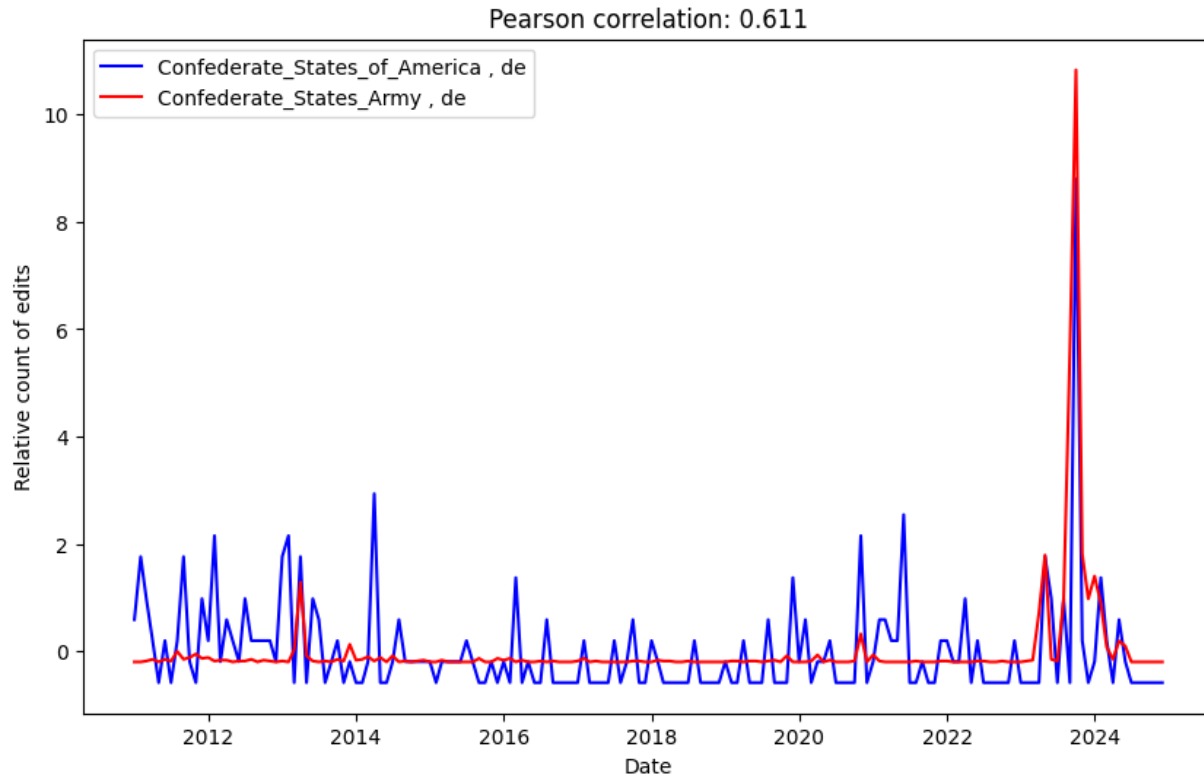
Correlations

Finding: higher share of articles belonging to the same topics in highly correlated pairs

Group	Total share (%)	Share cor >= 0.4 (%)
Nothing common	64.5	46.3
Same topic	21.4	18.3
Same topic, same language	13.3	32.0
Same article	0.6	3.4

Correlations

Limitation: small sample size, representation questionable

Correlations

Limitation: small sample size, representation questionable

Group	Total	cor >= 0.4 (%)
Nothing common	13096	81
Same topic	4325	56
Same topic, same language	2708	32
Same article	142	6

Correlations
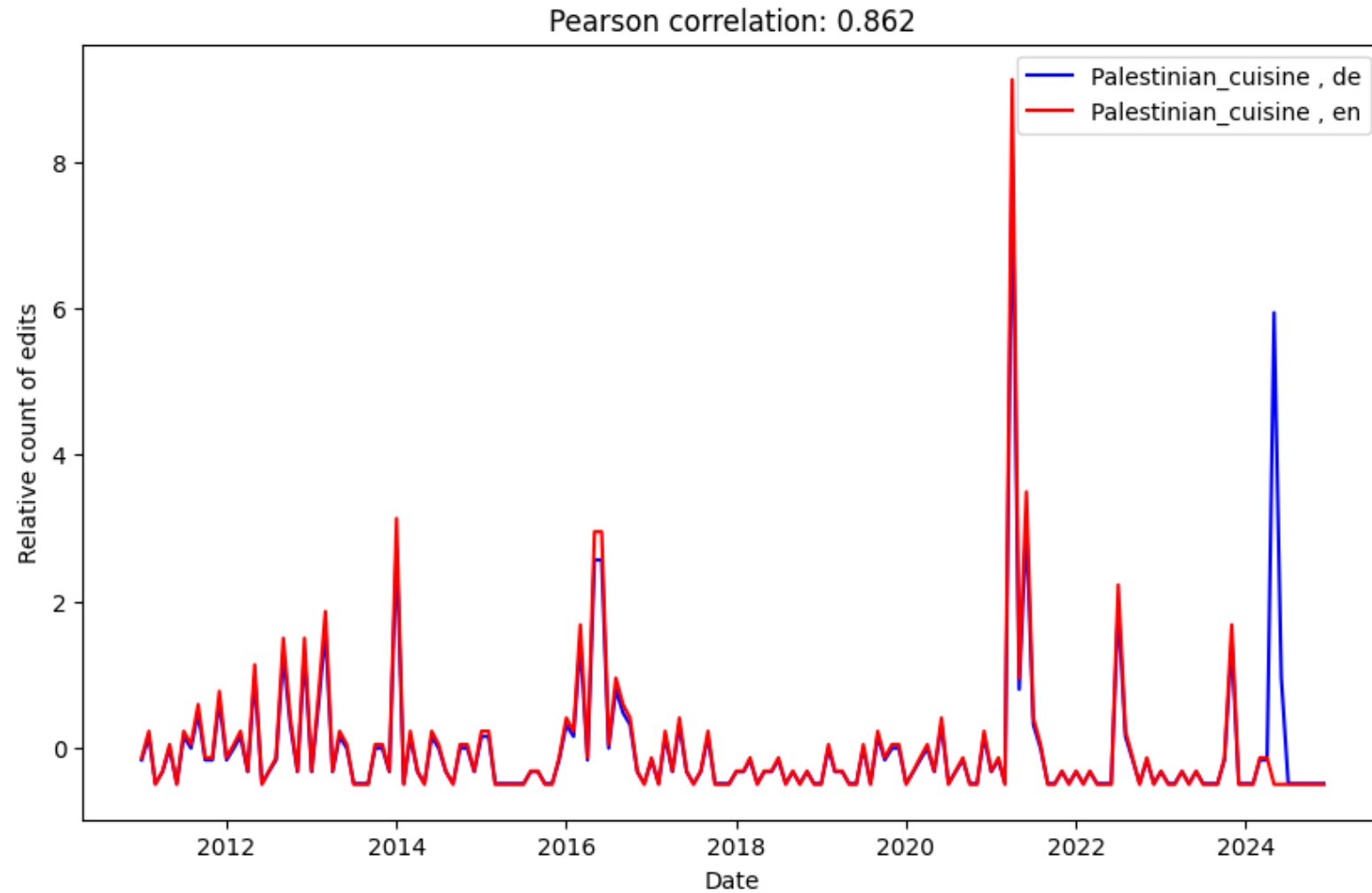
Protections over time

Long-term increasing trend

English and Ukrainian language versions protect the most

Protections are getting longer

Protections over time - English
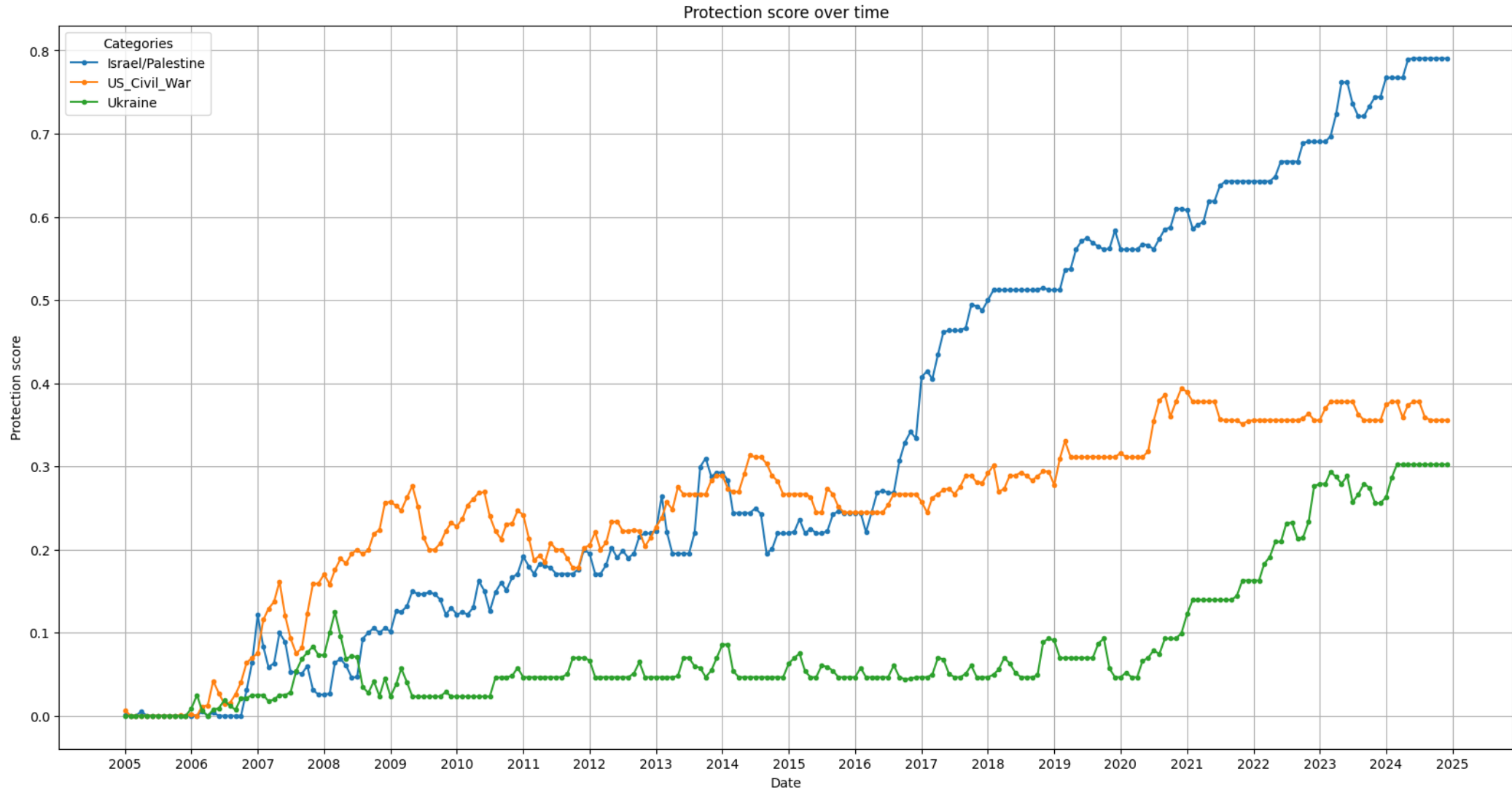
Protection score over time

Protections over time - German

Protection score over time

Protections over time - English

Protections by protection type
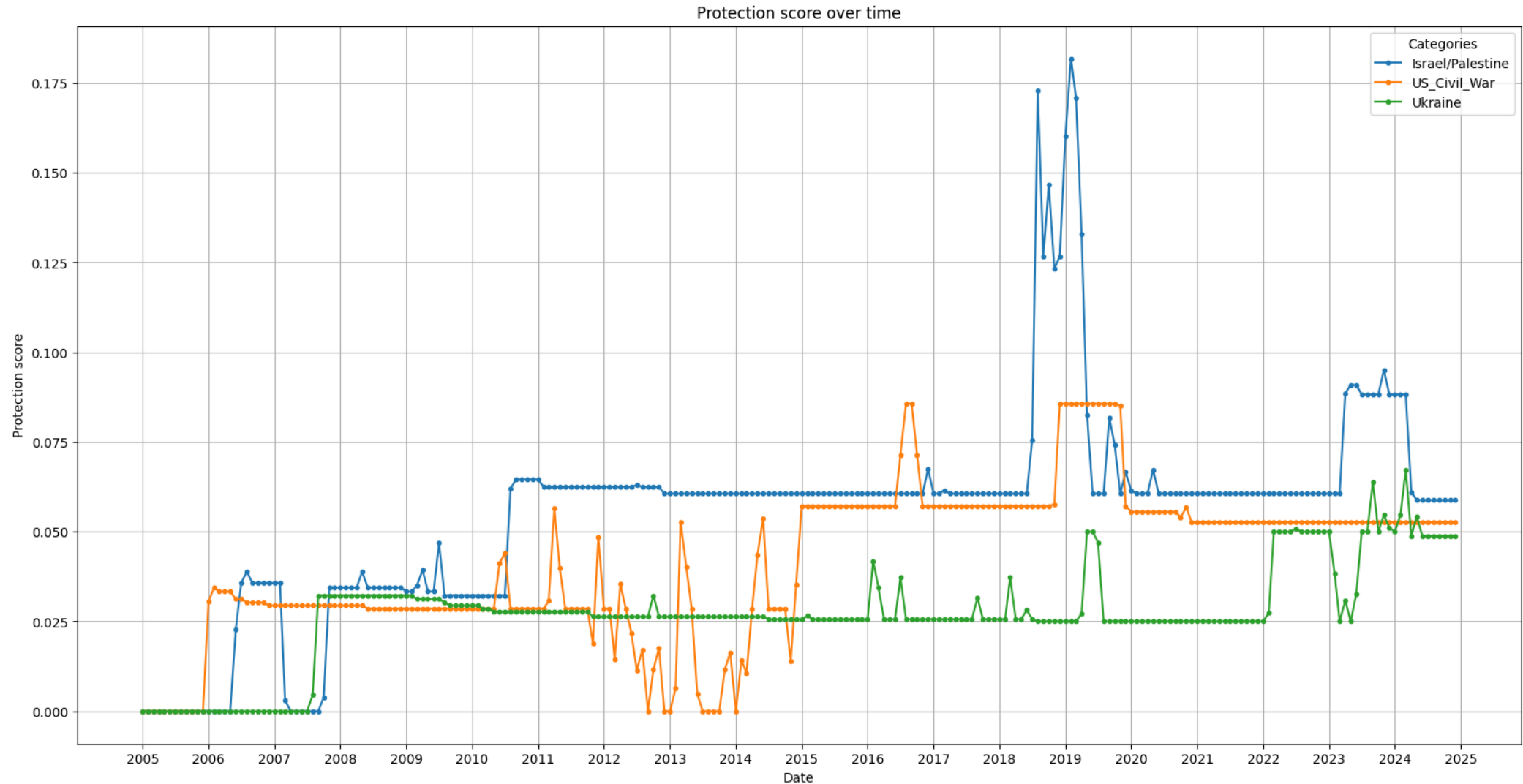
Protections over time - German
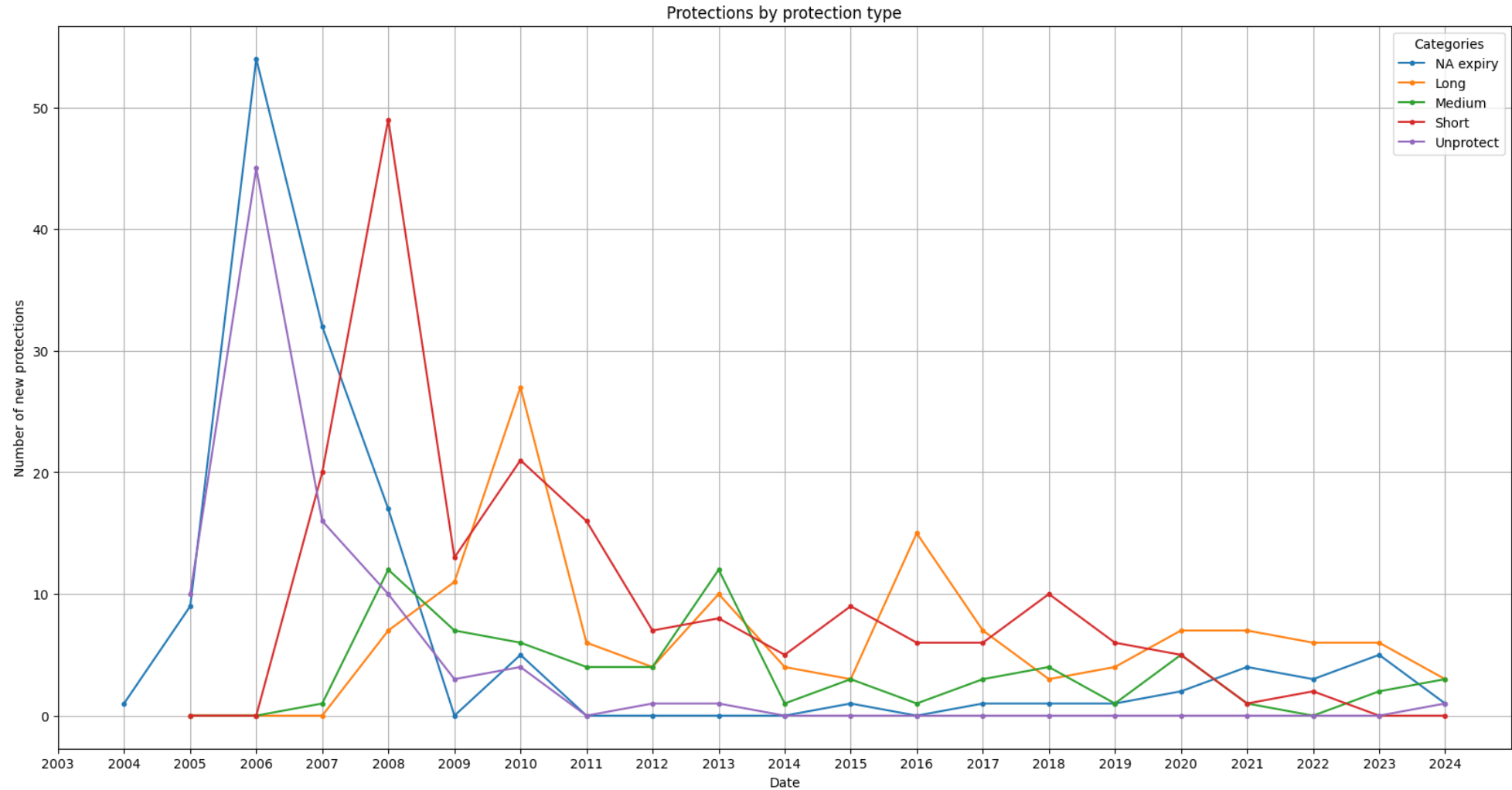
Protections by protection type
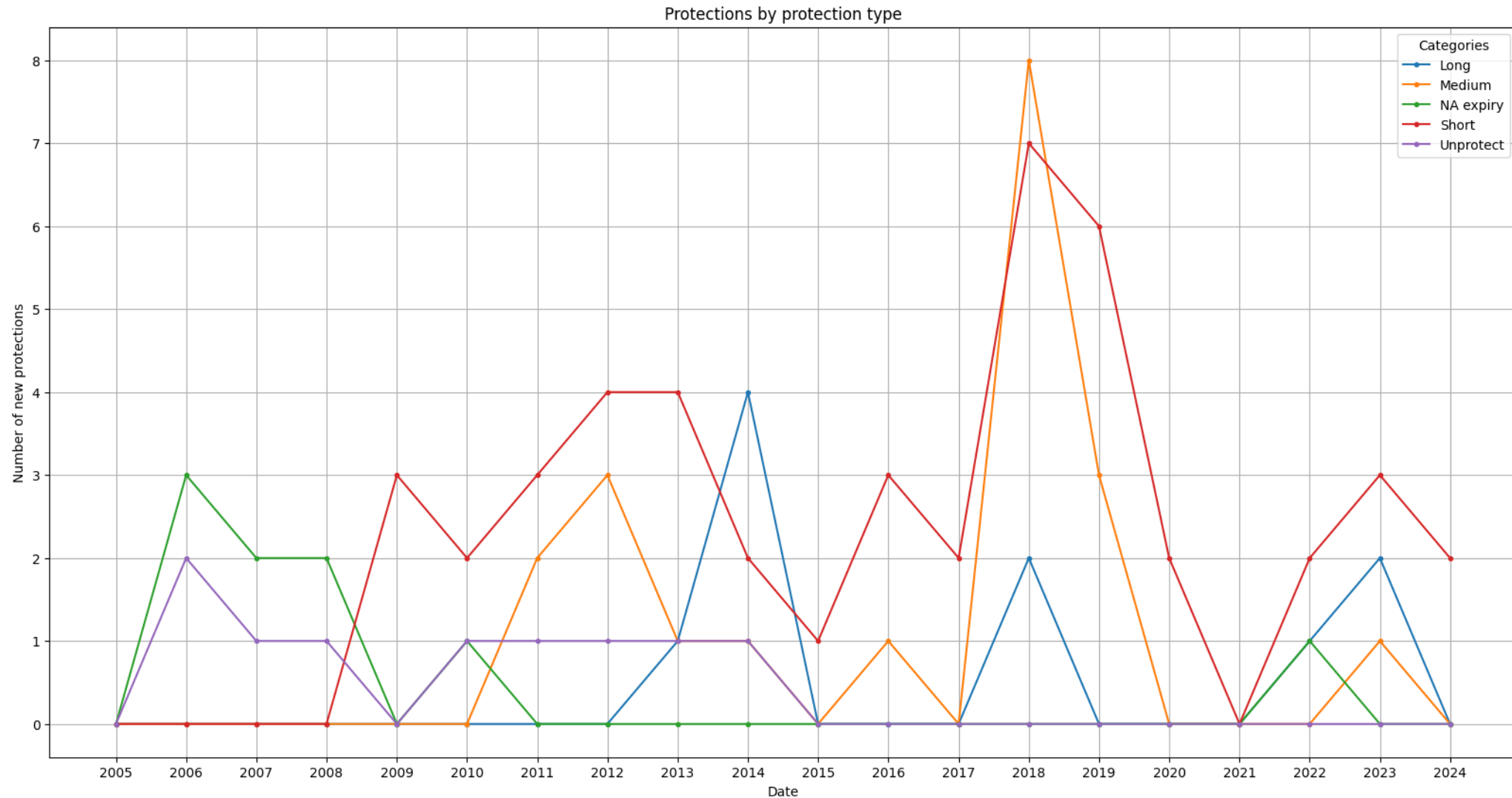
Limitations

Small sample size

Many attributes unavailable when going further in the past (reverted/reversion flags, type of protection, etc.)

Examples of the edit activity

- https://en.wikipedia.org/w/index.php?title=Battle_of_Antietam&diff=prev&oldid=521254817

- https://en.wikipedia.org/w/index.php?title=Battle_of_Antietam&diff=prev&oldid=597391114

Our contribution

We would like others to follow in our footsteps

Repository with scripts to pull and process the data

Some metrics from literature + our new ideas for metrics that can be measured and analysed also implemented in the repository

Project report including main findings from the case study

References

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,
[Online]. Available: https://doi.org/10.1007/s41109-020-00305-y (visited on 05/18/2024).

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

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

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

T. Yasseri, R. Sumi, A. Rung, A. Kornai, and J. Kertesz, "Dynamics of conflicts in Wikipedia," PloS one, vol. 7, no. 6, 2012.

K.-I. Goh and A.-L. Barabasi, "Burstiness and memory in complex systems," Europhysics Letters, vol. 81, no. 4, 2008.