D2.5 Computational analysis report

Final version

Authors: David Abella and David Laniado,
Big Data and Data Science,
Eurecat - Technology Centre of Catalonia

February 2026

Partners:

Supporter:

Table of content

List of acronyms and abbreviations

API: Application programming interface
DAWR: Deleted to Added Words Ratio
DSA: Digital Services Act
ECR: European Conservatives and Reformists
EPE: European Parliament Elections
EPP: European People's Party
ESN: Europe of Sovereign Nations
EU: European Union
EU Party: European political party
Greens: Greens/European Free Alliance
GUE: The Left group in the European Parliament – GUE/NGL
ID: Identity and Democracy
MEP: Member of the European Parliament
PCR: Policy Citation Ratio
NI: Non-Inscrits
PDF: Probability Distribution Function
PFE: Patriots for Europe
NPOV: Neutral Point Of View
RE: Renew Europe
S&D: Progressive Alliance of Socialists and Democrats in the European Parliament
TAR: Talk to Article Ratio
VLOP: Very large Online Platform

Introduction

In this deliverable, we describe the computational analysis performed in the project. This analysis complements the legal analysis described in deliverables D2.3 and D2.4 with insights from the data describing the actual editorial activity and dynamics of the Wikipedia communities during the 2024 European Parliament Election. The aim of this work is therefore to explore the practices and dynamics of Wikipedia communities during this particular election as they emerge from their digital traces, in particular in relation to the DSA (Digital Services Act) and the obligations it introduces for VLOPs (Very Large Online Platforms) with respect to so-called disinformation.

The open nature of the Wikipedia project has long attracted criticism and skepticism, until encountering increasing acceptance even by academics (Jemielniak and Aibar, 2016). Indeed, the fact that most pages can be edited at any moment by anybody just by clicking on the "edit" tab and modifying the content at will would at first raise concerns about the reliability of its content. According to a famous Wikipedia joke, "the problem with Wikipedia is that it only works in practice. In theory, it can never work." This apparent miracle is made possible, among other elements, by a complex set of norms and policies that allow each community to self-organise, granting special rights to users who have specific roles such as administrators, and setting guidelines and mechanisms such as consensus seeking and arbitration committees for dispute resolution. The set of norms that constitute the basis for this coordinated work has been growing over the years, since the creation of the English Wikipedia in 2001, to an increasingly sophisticated system which can vary across language communities, and which has been called an emergent bureaucracy (Butler et al, 2008).

Deliverable D2.3 presents an analysis of community norms and practices from a legal perspective, in particular in relation to elections and the DSA. To operationalise the study of policy usage in practice by the communities in their daily work with a computational analysis, we developed a method to identify policy citations in edit summaries (where an editor describes or explains the motivation for their edit) and comments in talk pages (pages devoted to discussion about the content of a Wikipedia article). We filtered policy citations to focus on the ones related to content and behaviour (avoiding policies about articles format, template and editorial issues related to style), and then, to focus more specifically on the ones directly related to content moderation and so-called disinformation, we applied a further filter relying on manual inspection of the policies based on the work in deliverable D2.3. This is, to the best of our knowledge, the first study that develops a computational multilingual method to extensively trace and inspect the invocation of community norms by editors.

The work described in this deliverable builds on the dataset and preliminary analysis described in D2.2. We started from the dataset described in D2.2, maintaining the same set of pages related to the 2024 European Parliament election, i.e. pages associated with political parties and candidates from all the countries involved, and pages describing the elections, and the same set of language editions, i.e. 31 languages official or relevant in the European Union.

We enriched the dataset in different directions. First, we added user rights metadata, which can be used to infer its status and role and are specially relevant to inspect community practices with respect to information reliability. The second is adding page views information, to go beyond metrics of activity and also account for attention dynamics, considering not only the thousands of daily editors but also the millions of daily readers of Wikipedia. Finally, we included gender as a variable associated with all biographies of politicians in our dataset, extracted from the Wikipedia page itself. This last addition followed a suggestion received in a community engagement event in which we disseminated the preliminary results of this work to representatives of European Wikipedia communities. The data and code generated in this work will be available on an open repository (Abella and Laniado, 2025b).

The document is organised as follows: in the next section, we describe how we enriched the dataset presented in deliverable D2.2; then in the central part of the deliverable we have three sections for three different kinds of analysis: language agnostic analysis, which includes analyses based only on interaction data and not considering the text, therefore independent of the language; multi-language analysis, which accounts for the text of the pages and edits, and therefore deals with natural language processing; and policy analysis, which specifically focuses on how community policies are mentioned by editors. Each of these three sections devoted to analysis include a first subsection describing the methodology. Finally, in the last section, we draw some conclusions and directions for future work.

Dataset enrichment

As a starting point for this work, we use the multi-language dataset described in previous deliverable D2.2, which includes a set of pages related to the 2024 European Parliament election, such as the pages of political parties and candidates. The dataset comprises 10,249 articles and 4,662 talk pages across 31 languages (covering all 24 official EU languages plus major regional and widely spoken languages). For each page, the dataset includes interaction data and metadata including related country and European political party.

We refer to deliverable D2.2 for more details about the dataset and the criteria for including articles and selecting language editions.

Temporal window

As a first addition to the previous dataset, we expanded the temporal window considered. In order to include not only the electoral campaign but all the period relevant to the elections, we decided to consider the time window ranging from the date when the election dates were announced (23/05/2023) to the date of the formation of the European Commission (01/12/2024).

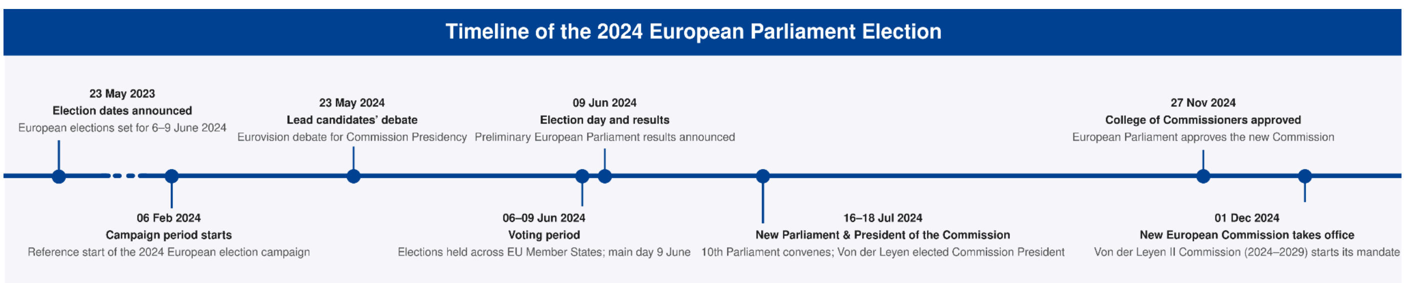

Figure 1. Timeline of the 2024 European Parliament election. Timeline of relevant dates regarding the election, from the election announcement to the start of the new elected commission.

Gender extraction for candidate biographical pages

Based on feedback received from the communities while disseminating our preliminary results at the Big Fat Brussels meeting[1], in October 2025 in Brussels with representatives of the European Wikipedia communities, we decided to add the gender dimension to our dataset. As our dataset is based on a collection of articles, of which a large part are biographies of politicians, the gender is associated with each of these biographical articles and can be extracted computationally from Wikipedia pages.

[1] See official page for more information: https://meta.wikimedia.org/wiki/Wikimedia_Europe/Advocacy/Big_Fat_Brussels_Meeting_X.

This addition allows us to study whether and how the gender of a politician may affect editing dynamics, vandalism, etc, on the corresponding articles. For example, one might wonder whether biographies of female politicians receive more or less vandalism, or compare the most frequent words added and deleted in female vs male biographies. Indeed, a recent study by Sala et al. (2025) focused on the same elections that are object of our analysis showed that women candidates on the X microblogging platform[2] tended to receive the same amount of toxic messages as men candidates, but with higher intensity, and with more messages containing sexist slurs or sexually explicit content.

We extracted gender information about each article from Wikidata, and associated, with the corresponding article and talk page in our dataset, one of these four categories:

- **M: Male** → for biographies of persons identified with male gender.
- **F: Female** → for biographies of persons identified with female gender.
- **NB: Non Binary** → biographies of persons identified with non-binary gender.
- **A: Agender** → for all pages that do not represent a person, but a political party, the parliament or the election itself, etc.

Given the very limited presence of non-binary gender persons in the dataset, we generally omit them from the results shown in this deliverable. In several cases, to emphasise differences between male and female biographies we omit "agender" pages.

Page views extraction

Since 2015, Wikimedia makes available hourly page view data for its projects through an API[3]. We extracted page-view data with daily time granularity for all the articles in our dataset, and for the period defined around the 2024 European Parliament election.

User level definition

In order to be able to distinguish administrators and other special categories of users, we extracted from the Wikipedia API user rights information, indicating for every user present in our dataset to which categories they belong according to their user rights.

As there are many different editorial rights, and it would be too complex to account for all of them, we built on previous work by Arazy et al. (2015) to aggregate special rights categories within a reduced number of user levels. This categorization was developed with a focus on the English Wikipedia, but, due to the similarity of user rights and organisation across communities, we used it for all language wikipedias considered in this work.

[2] https://x.com

[3] https://doc.wikimedia.org/generated-data-platform/aqs/analytics-api/concepts/page views.html

Table 1 shows the levels defined by Arazy et al. (2015). In our dataset, just 5 of these levels were present:

- **Level 0 - Unregistered users:** We went one step forward and considered here anonymous and bot users, to differentiate them from users at different levels.
- **Level 1 - Registered / Manually Registered:** Newly registered users and manually registered users, which had to bypass some small restrictions.
- **Level 2 - Tech Admin / Border Patrol / QA:** This categorization includes Technical administrators, responsible about files; users fighting vandalism and ensuring content quality and users who developed automated tools for the community.
- **Level 3 - Administrators:** Highly involved users responsible for the community administration, taking a role of governance and representation.
- **Level 4 - Security Force (Oversight/CheckUser):** Users who work to keep users with bad intentions out of the Wikipedia project, and combat intentional manipulation of content.

Level on Org. Chart	Role	Description	Access Privileges
Level 0	*Unregistered Users*	Non-community members	*
Level 1	*Registered Users*	Newly registered users	user
			autoconfirmed
	Manually Registered Users	New users who had to be manually registered to bypass some restrictions	confirmed
			IPpblock-exempt
Level 2	*Technical Administration*	Privileged users responsible for the administration of the technical aspects (e.g. user accounts, files)	filemover
			accountcreator
	Border Patrol	Users responsible for fighting vandalism by reverting malicious edits	rollback
	Quality Assurance	Privileged users responsible for patrolling Wikipedia and ensuring content quality	reviewer
			autoreviewer
	QA Technicians	Users who develop automated tools (i.e. edit filters) to assist quality assurance work	abusefilter
Level 3	*Administrators*	Highly involved users that are responsible for the social administration of the English Wikipedia community	sysop
			bureaucrat
Level 4	*Security Force*	Highly trusted users who are working to keep malicious users out and combat intentional manipulations of content	oversight
			checkuser
Level 5	*Directors*	Key users responsible for oversight of the Wikimedia organization	steward
			importer & transwiki
	Privacy Commissioner	High-ranking users who investigate complaints about violations of privacy policy	ombudsman
Level 6	*Benevolent Dictator*	Jimmy Wales; responsible for defining high-level policies and norms and for overall direction of the community	founder

Table 1: Wikipedia editors' levels as defined by Arazy et al. (2015).

Text revisions extraction

While for the language-agnostic analysis we only need metadata, interaction and page-view data, if we want to go deep into the editorial content and the policy citations, we need the text modified after each revision. For both articles and talk pages, we identified the words added and deleted in each revision with respect to the previous one.

We used the Wikipedia API to retrieve the text of each revision included within the considered time window, for all the articles and language editions in the dataset.

Language agnostic analysis

In this section we focus on metrics that do not depend on the language, and are based on interaction dynamics and page views.

Methodology

Vandalism filter

For most of the analyses, it is useful to be able to separate meaningful contributions from vandalism, which is defined by Wikipedians as "editing (or other behaviour) deliberately intended to obstruct or defeat the project's purpose"[4]. Given the context of our project, it is interesting to be able to quantify and characterise vandalism across language editions in the context of the European elections; therefore we will devote this section to explain our filtering methodology, while for the rest of the analyses we will generally filter out vandalism.

To identify vandal edits, and their corresponding reverts (restoration to the version of the article previous to the vandal edit), we followed the methodology introduced at Borra et al (2015). We detect an edit as vandal when at least one of these conditions is met:

- The comment of a revert contains the word 'vandal', 'vandalism' (or the corresponding translation according to the Wiki's language).
- The user name making the revert belongs to one of the known anti-vandalism bots.
- An IP-edit is reverted within 60 seconds.
- The automatic edit summary (WP:AES)[5] indicates that the content of a page was blanked or replaced by unrelated text such as curse words.

When an edit is flagged as vandal, if it is reverted, we will filter all edits occurring between the vandal edit and the reverting edit, since they may correspond to more vandal edits or to editions that will not be affecting to the article, since the revert will restore the version of the article previous to the vandal edit. For more details about this methodology, refer to Borra et al (2015).

Talk to Article Ratio (TAR) metric

As we introduced above, editorial activity on each encyclopedic entry takes place on two parallel spaces: the article page, and the associated talk page. In the article page, editors do directly edit the content that will be immediately visible to the article readers, adding a short "edit comment" or "edit summary" to describe or motivate their edit; the talk page is an associated page devoted to explicit discussion, communication and organisation among editors (Kaltenbrunner and Laniado, 2012).

[4] See: https://en.wikipedia.org/wiki/Wikipedia:Vandalism
[5] See: https://en.wikipedia.org/wiki/Help:Automatic_edit_summaries

To investigate the relationship between these two activity streams, we use the Talk-to-Article Ratio (TAR) which we define as the ratio between talk and article revisions for a given article (Abella and Laniado, 2025a).

Results

We now describe the results obtained for different language editions. Only in some cases the results shown are based on all the language editions considered, i.e. in the cross-language analyses and for some specific results where we aggregate data from all languages together. Apart from these cases, results are computed separately for each language, keeping the independent organisation between Wikipedia language editions. For maintaining this document readable and compact, we only show each result for one or a few selected language editions and we focus mostly on the English and other major language editions, with high amounts of interaction data. Results for all language editions will be available on the online repository (Abella and Laniado, 2025b).

Page creation and coverage

We first have a look at the temporal dynamics of page creation. In Figure 2 we can see the evolution of the overall number of articles and talk pages in our dataset across time, which exhibits a constant growth with some peaks especially after the elections. The two plots on the top show the growth of the cumulative number of articles and talk pages, while the plot on the bottom shows the number of new pages created per day, highlighting the peaks for both articles (blue line) and associated talk pages (red line). The data for each language edition considered separately tend to follow the same pattern.

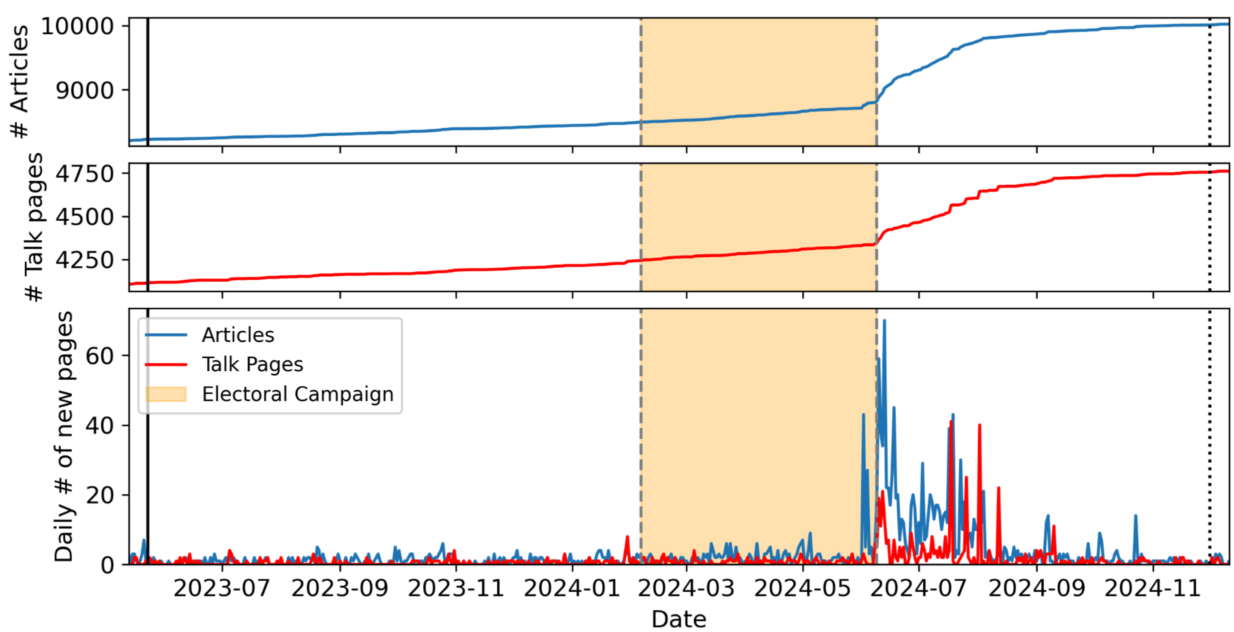

Figure 2. Page growth through the electoral period for all language editions considered. Number of articles (upper) and talk pages (centre) from our selected list created at a certain date. Daily number of new pages during our selected period (lower). The orange shaded region delimited by grey dashed lines is the electoral campaign, the solid line is the Election announcement (23/05/2023) and the dotted line highlights the EU Commission formation (01/12/2024), which delimit the period in which we perform the analysis.

We find the highest peaks for article creation just after the elections, which can be partly explained by the fact that many candidates did not have a dedicated page before the elections, and after being elected deserved one, getting to meet the "notability" criteria of different Wikipedia communities. We find another important peak in article creation a few days before the elections, which may be explained by some candidates and parties getting more attention by the media and public opinion as the election approached.

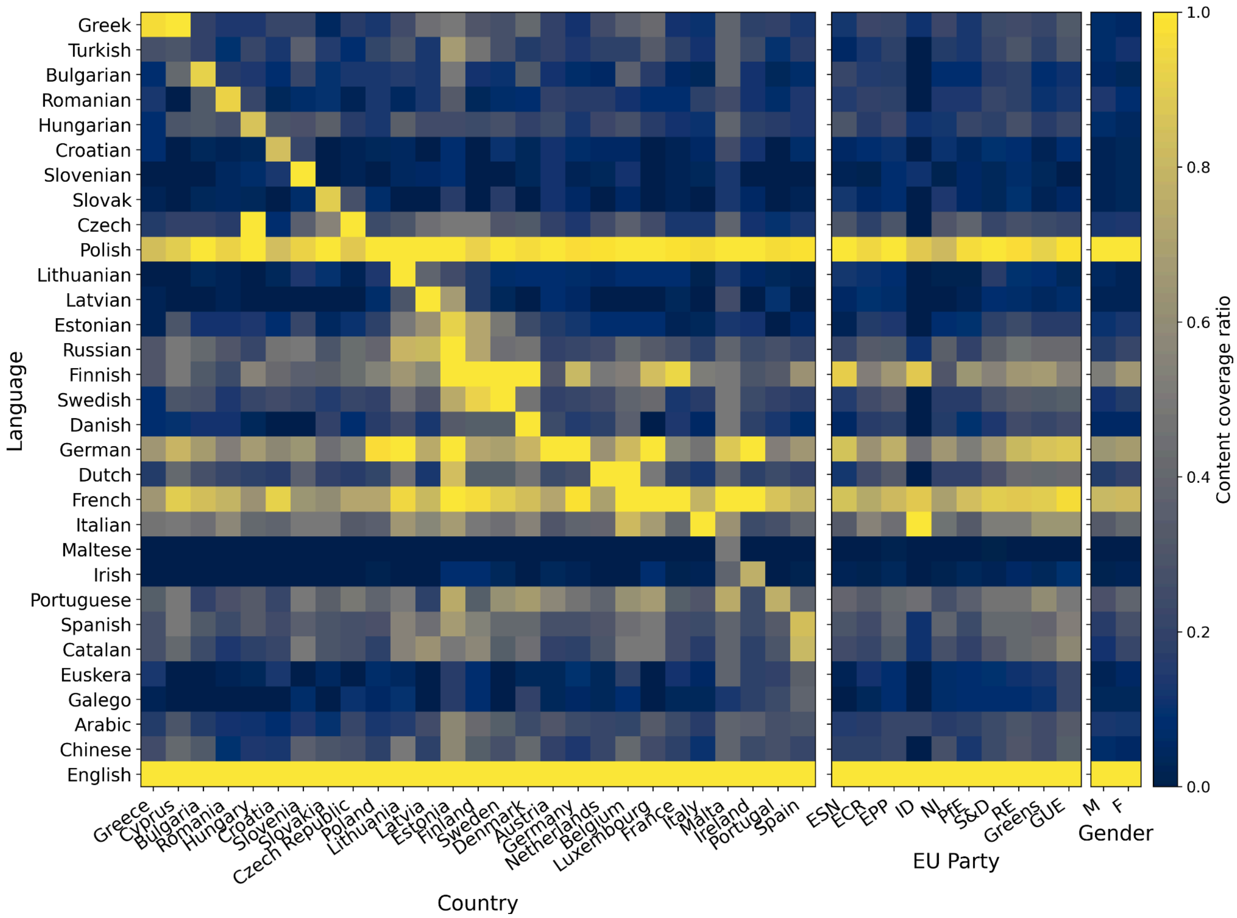

Figure 3. Cross-language content coverage across different countries, European parties and genders. Ratio between the number of the pages that each language covers and the total number of pages with content about the different countries (left), EU parties (centre) and gender (right).

The temporal dynamics of page creation also reveal the community's responsiveness to electoral events. The pattern where post-election peaks exceed pre-election activity suggests that

Wikipedia's notability criteria are often met only after candidates achieve electoral success, rather than during their campaigns. This reflects the encyclopedia's emphasis on documented significance rather than anticipatory coverage of potential political figures.

It should be noticed that these plots include page creation dynamics across language editions, i.e. it does not only include, for example, pages about Bulgarian politicians on the Bulgarian Wikipedia, but also on any other language edition included in the dataset (if they exist in the other languages). We will now have a deeper look at this aspect.

Figure 3 shows how each language edition covers pages related to each country, EU Party and gender. As it could be expected we observe a highlighted pseudo-diagonal indicating that each language edition tends to have a high coverage of the country or countries related to it. The ordering of languages on the vertical axis and of countries on the horizontal axis is made to reflect these relations, via the pseudo-diagonal.

Beyond the unsurprising tendency of each language community to focus on its related cultural context (Miquel-Ribé & Laniado, 2018) and the fact that some language editions, such as English and Polish, tend to cover most content about all countries, we observe the reflection of other interesting ties between cultures and countries: Scandinavian countries tend to cover one another's content, with Finnish having also high coverage on France, and Danish covering mostly content related to Denmark.

Coverage of EU Parties varies across language editions and it seems to reflect the prominence of certain parties in certain countries, as well as the adherence of national parties to different EU parties. Regarding gender, no strong differences are observed in terms of coverage between biographies of men and women politicians. "Agender" pages are not shown in this and in most of the following graphs in order to highlight the differences between male and female politicians. These pages tend to have higher coverage as they are pages related to political parties and elections that tend to more easily meet notability criteria and attract public attention.

Activity metrics

In the previous section we have focused on page creation and coverage, considering only the existence and creation of pages, but not the activity on those pages. We now take a deeper look at activity patterns and dynamics.

As explained above, each article page may have a talk page associated, and activity takes place in form of revisions of either kind of page. As a general rule, we refer to "revisions" or "editorial activity" to indicate both revisions to article and talk pages, while we use the term "edit" or "article revisions" to refer to revisions to article pages, and "comments" or "talk page revisions" to refer to revisions to talk pages.

Figure 4 shows activity patterns among language editions and countries, in terms of the proportion of revisions of each language edition devoted to each country. We observe a much more highlighted diagonal with respect to Figure 3, indicating that while certain language communities may have pages covering other countries' politics, the activity on these pages tends to be low, as editors tend to concentrate most activity on their own country as observed in previous literature (Miquel-Ribé et al, 2021). This said, it is interesting to notice some strong persistent cross-language ties, such as higher activity in Swedish on Finland, in Hungarian on Slovakia, or in Turkish on Germany, Poland and France.

In terms of gender, we notice higher activity on pages related to male than female politicians especially in some Baltic countries; comparing these results with those from Figure 3 it is interesting to notice that this tendency is not related to coverage, i.e. pages on female politicians exist in these language editions, but they attract less activity.

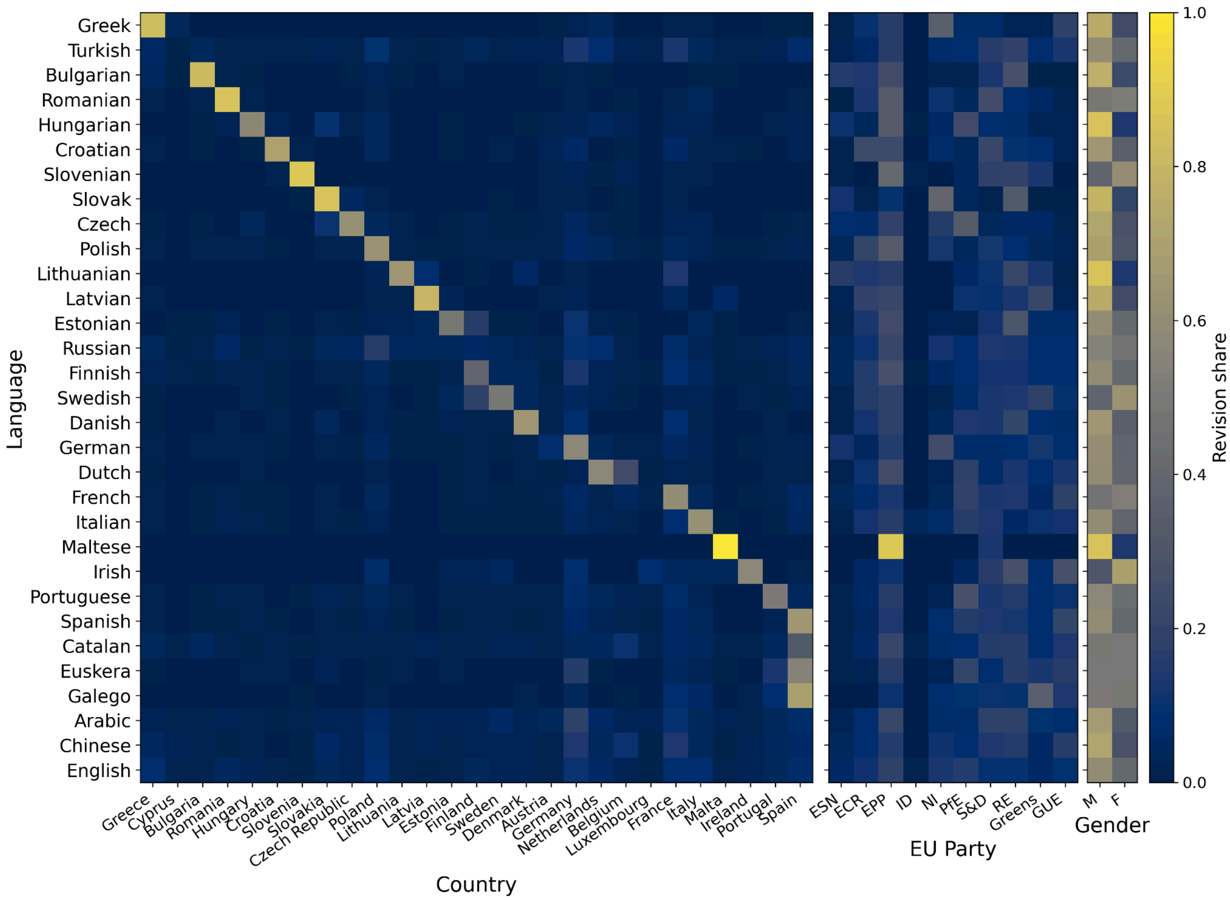

Figure 4. Revision share across different countries, European parties and genders. Proportion of revisions in each language edition on content about the different countries (left), EU parties (centre) and gender (right) and the total number of edits in the language edition.

User metrics

While until here we have just considered overall activity in each community, disregarding the different kinds of editors involved, we now take a deeper look with respect to this aspect, in order to get an idea of the relative importance and role of different user levels in each community.

In Figure 5 we show the distribution of revisions by user, highlighting user types (on the left) and tiers (on the right). Regarding user types, anonymous users appear as one only user having a very high number of contributions; this is due to how we treat the data, aggregating all anonymous contributions in one only user, because of the difficulty of distinguishing different users identified only through their IP address. As expected, in both user types and user levels, we observe a fat-tailed distribution of revisions per user, a result already reported for entire Wikipedia editions. Here we see that in an electoral period, this law holds, even for different user levels, where we observe the same distribution decay with different cut-off, depending on the size of the groups. In the pie charts we can see the proportion of revisions made by each type of user profile. We observe a high proportion of contributions made by tech admins and border patrols in the English Wikipedia, and by Security Force in the Polish Wikipedia.

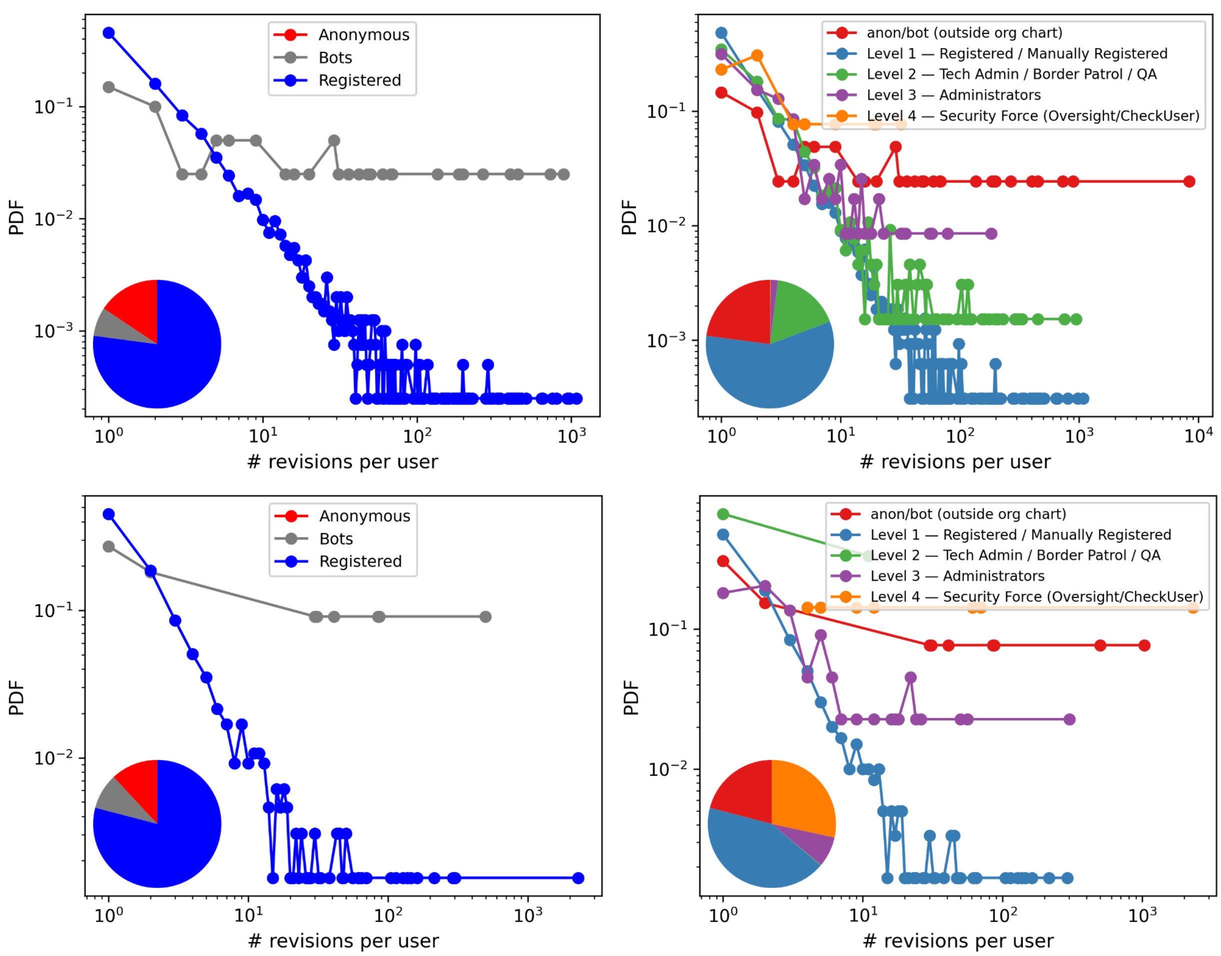

Figure 5. Distribution of activity per user by user type and tier. Distribution of revisions per user by user type (left) and tier (right) for the **English** (top) and **Polish** (bottom) Wikipedia language edition (in log-log scale). Each plot shows a pie chart with the proportion of editorial activity per user type (left) and user level (right). Anonymous users cannot be tracked individually, so the distribution of revisions per user cannot be plotted.

Figure 6 reveals interesting relationships between user diversity and editorial activity across different privilege levels in the English Wikipedia. The left panel shows that pages with higher total editorial activity tend to attract more distinct editors, following an approximately power-law relationship, which suggests that popular topics benefit from broader community participation. The central and right panels examine how activity from higher-level users (Level 2 and Level 3) correlates with anonymous or bot contributions.

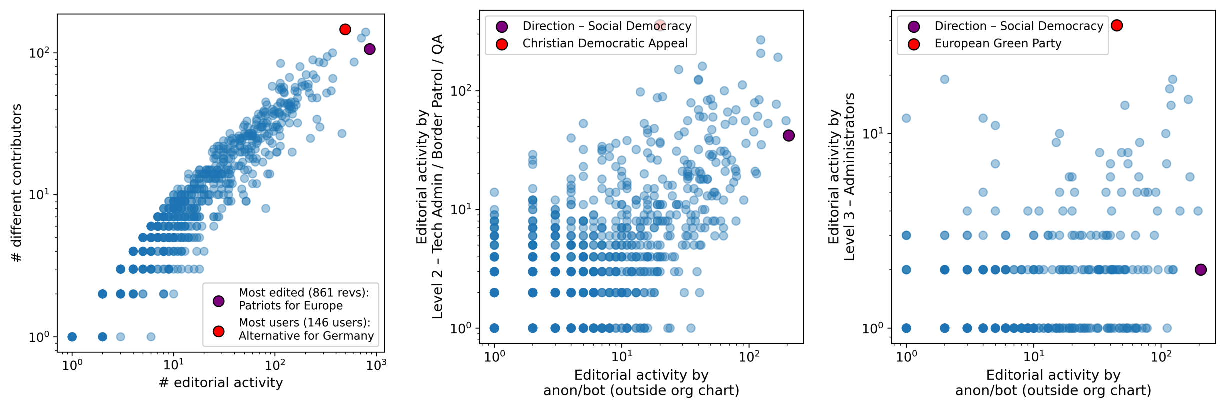

Figure 6. Relation between user level and editorial activity per page. Scatter plots representing the relationship between users heterogeneity and the activity on different levels at the English Wikipedia during the electoral period. Specifically, we show the number of distinct editors as a function of the number of revisions (left), and the number of edits made by Level 2 editors (centre) and Level 3 editors (right) as a function of the number of edits made by anonymous users or bots. In each plot, each point is a Wikipedia page. The pages with highest Y-variable and X-variable are highlighted in red and purple, respectively.

Notably, we observe that pages with substantial anonymous editing activity also tend to receive more attention from Level 2 users (technical administrators and quality assurance), likely reflecting increased monitoring and quality control needs when articles receive more unregistered contributions. The relationship is less pronounced for Level 3 users (administrators), whose involvement appears more evenly distributed across pages regardless of anonymous activity levels, possibly because their role focuses more on policy enforcement and dispute resolution than on direct response to anonymous edits.

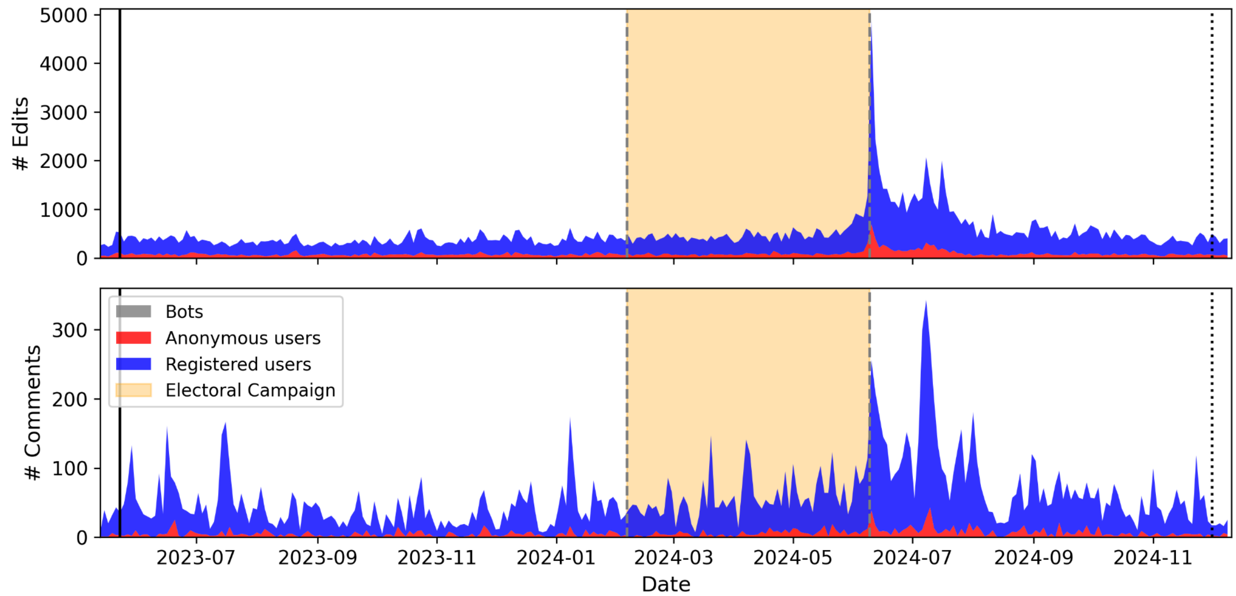

Figure 7. Editorial activity through the electoral period. Number of edits (upper) and comments (lower) in the pages during the electoral period for all language editions. The share of editorial activity is shown in colours: anonymous users in red, registered users in blue and bots in grey**.** The orange shaded region delimited by grey dashed lines is the electoral campaign, the solid line is the Election announcement (23/05/2023) and the dotted line highlights the EU Commission formation (01/12/2024), which delimit the period in which we perform the analysis.

Figure 7 illustrates the temporal distribution of editorial contributions across the entire dataset, aggregating all language editions. The dominance of registered users (blue) in both article edits and talk page comments underscores the central role of the committed Wikipedia community in maintaining content during the electoral period. Anonymous users (red) contribute a relatively small but consistent share, with their activity concentrated in article edits rather than talk pages, which aligns with the tendency of unregistered users to make direct content modifications rather than engaging in meta-discussions about article development. Bot contributions (grey) remain minimal throughout the period, appearing primarily in article edits where automated maintenance tasks are more common.

The temporal patterns show increased activity during and immediately after the electoral campaign, with the most substantial spikes occurring post-election. This timing suggests that the community's primary work involved updating articles to reflect election outcomes, documenting the new parliamentary composition, and resolving any content disputes that emerged during the campaign period. The relatively stable proportions across user types throughout the timeline indicate that the community maintained its characteristic collaborative structure even during this high-attention period, rather than experiencing significant shifts in the composition of contributors.

Vandalism

We examine vandalism, starting from the observation of the proportion of vandalism edits by language edition; Spanish is the community with the highest proportion of vandalism edits, which reaches almost the 20% of total edits; followed by Romanian, Slovak, Bulgarian and Italian which also have values ranging between 10% and 15%. We find the lowest ratios in smaller Wikipedia editions.

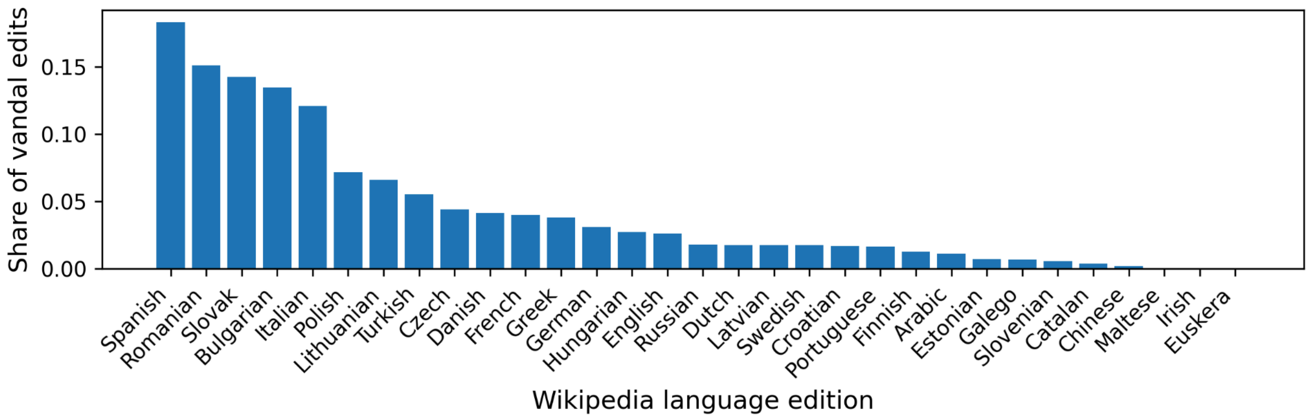

Figure 8. Share of vandal edits across different Wikipedia Language editions. Ranking of the proportion of vandal edits detected (with respect to the total edits) for the language editions analysed during the election.

Figure 9 examines the relationship between vandalism and overall editorial activity across four major language editions, with each point representing an individual Wikipedia page. The general positive correlation between vandalism and non-vandalism edits suggests that more active pages attract both more legitimate contributions and more vandalistic behaviour, a pattern consistent with the "tragedy of the commons" where high-visibility resources face greater abuse.

However, the relationship is not strictly proportional, with considerable scatter indicating that page-level factors beyond mere activity levels influence vandalism susceptibility. The gender-coded colouring reveals the relation between gender and vandalism during this particular electoral period. This suggests that institutional articles, not just biographical ones, can become targets during politically charged periods. The distribution of female politicians' pages (purple) shows they appear throughout the vandalism spectrum rather than clustering systematically at higher or lower levels, indicating that gender alone does not determine vandalism activity.

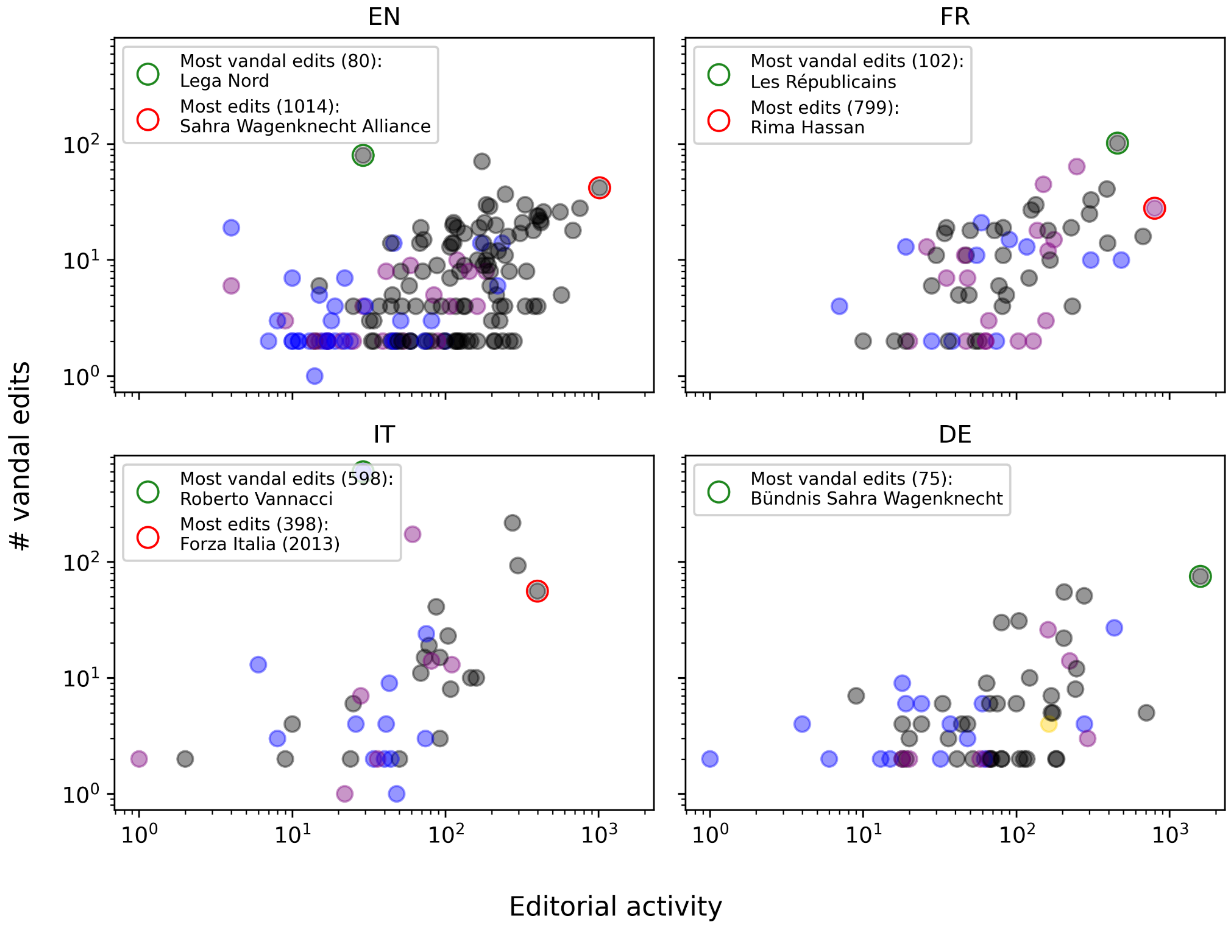

Figure 9. Vandalism in relation to editorial activity across the pages. Number of vandal edits vandal vs non-vandal edits per article in English (top left), French (top right), Italian (lower left) and German (lower right) Wikipedia. In each plot, each point is a Wikipedia page, coloured according to gender: male in blue, female in purple, non-binary in yellow and agender in black. The pages with highest Y-variable and X-variable are highlighted in red and green, respectively.

The breakdown of vandalism by content characteristics in Figure 10 provides deeper insight into which topics attracted disruptive editing during the electoral period. In the English Wikipedia (left panels), "agender" pages, representing political parties, electoral systems, and the elections themselves, receive disproportionately high vandalism rates compared to biographical pages. This pattern is particularly pronounced for specific EU party groups, with PfE (Patriots for Europe) and EPP (European People's Party) showing elevated vandalism levels. The high vandalism targeting Italy-related content, followed by Spain, Greece, Poland, and Portugal, may reflect the particular electoral dynamics and controversies in these countries during 2024.

The French Wikipedia (right panels) presents a contrasting pattern with higher vandalism rates targeting female politicians compared to male politicians, particularly affecting women from ECR (European Conservatives and Reformists) and GUE (The Left) parties. This gender disparity in vandalism prevalence raises important questions about whether female politicians face different forms of online hostility, even on platforms with community governance like Wikipedia. The pattern aligns with broader research findings on gendered online harassment (Sala et al., 2025), though the mechanisms may differ given Wikipedia's transparent editing history and community oversight.

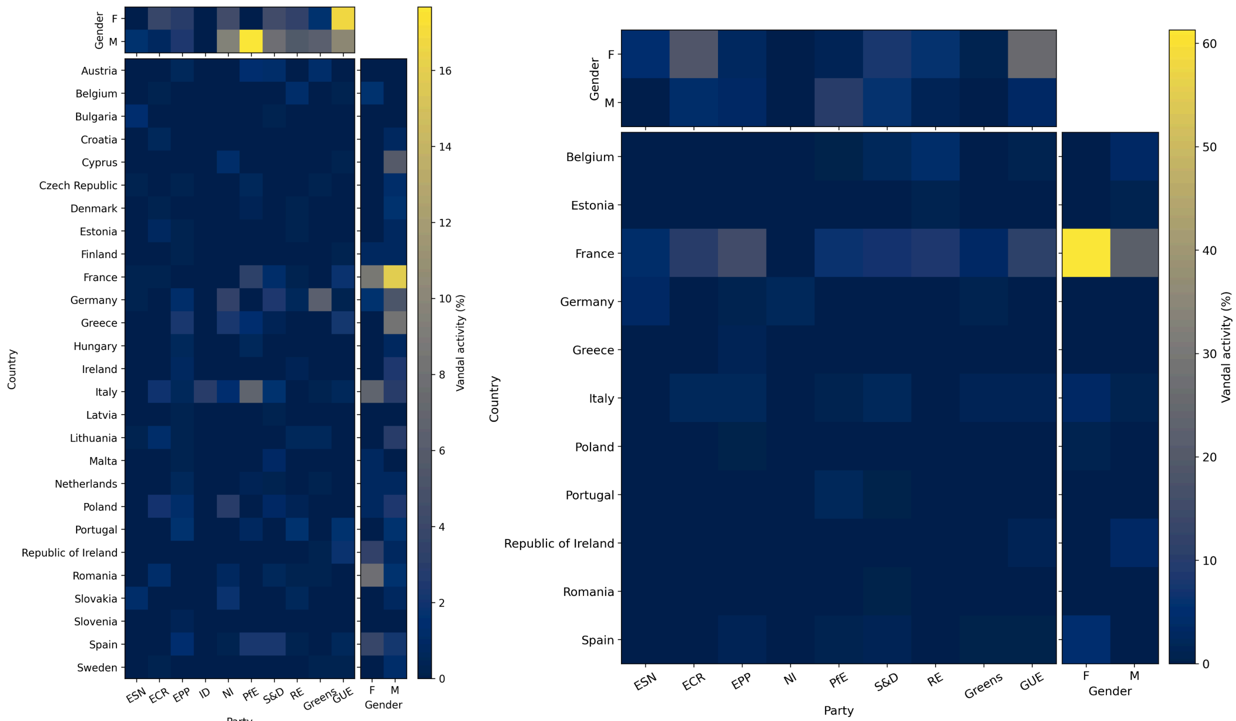

Figure 10: Proportion of vandalism across the different content. Percentage of vandal edits per country, European party and gender for the **English** (left) and **French** (right) Wikipedia during the electoral period.

The variation in vandalism patterns across both countries and political parties suggests that vandalism during electoral periods is not random but reflects the specific political tensions, controversies, and public attention surrounding different political actors and institutions. Pages about politically controversial or newly prominent entities appear more vulnerable, as do those representing ideological positions that provoke strong reactions from segments of the public.

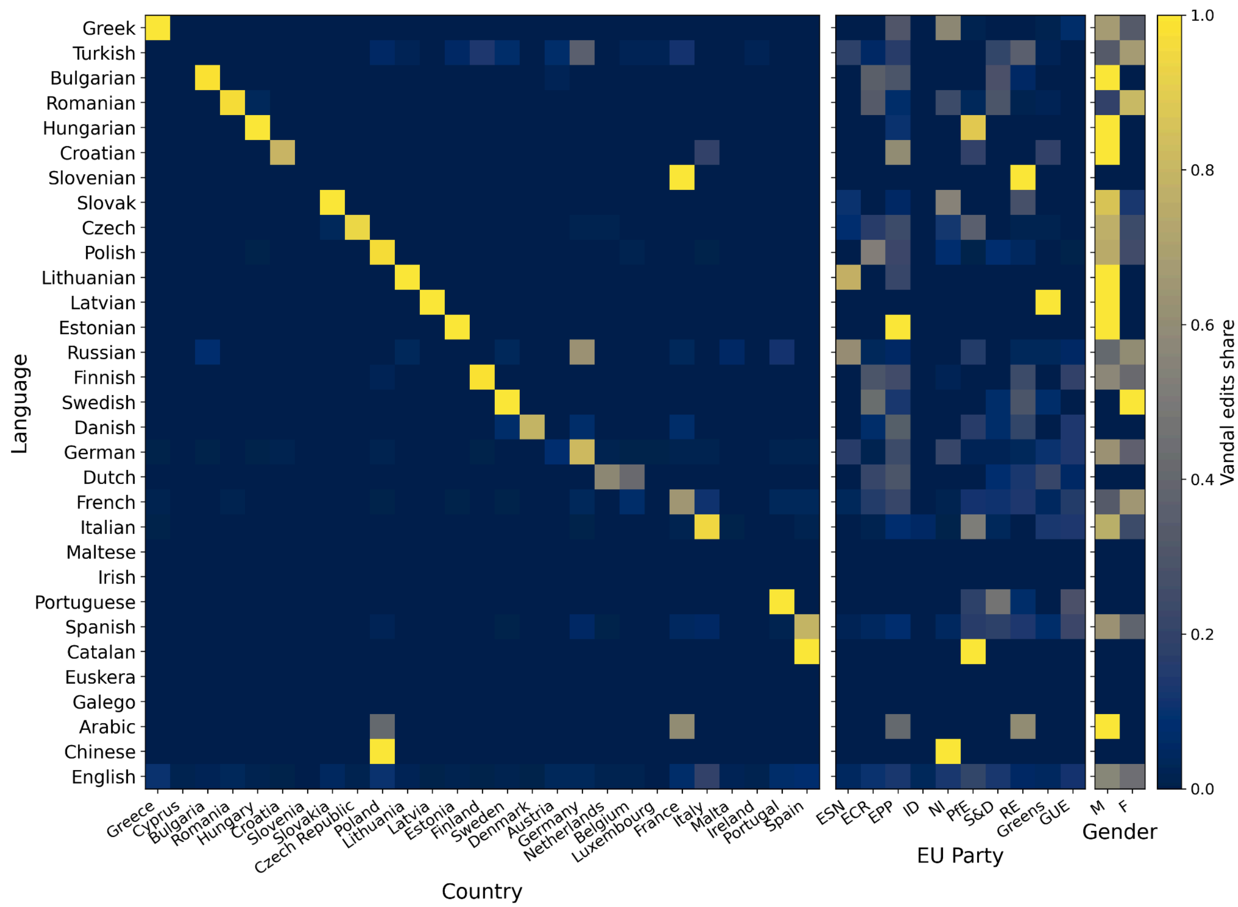

Figure 11: Vandal edits share across different countries, European parties and genders. Proportion of vandal revisions in each language edition on content about the different countries (left), EU parties (centre) and gender (right) and the total number of edits in the language edition.

Some countries, especially bigger countries, receive a certain amount of vandalism in other languages; this is the case for Germany in Russian and Turkish; France in Slovenian and Arabic; Poland in Chinese and Arabic.

Interestingly, vandalism tends to focus on different EU parties depending on the language edition. Regarding gender, "agender" pages receive the highest level of vandalism in general, although this is not shown to highlight gender differences between male and female politicians. The graph shows that in most countries men MEPs tend to receive more vandalism than women; this is particularly accentuated in Baltic countries and in several Eastern countries, which can only in part be explained by the higher presence of male MEPs in these countries. We find more vandalism directed towards women in Swedish, Romanian, French, Turkish and Russian.

Talk to Article Ratio (TAR) metric

The Talk-to-Article Ratio (TAR) metric provides a window into the collaborative and deliberative processes underlying Wikipedia's content creation, and allows us to compare the amount of deliberation with the editorial activity around a certain topic (which may serve as an indicator of how controversial a page is for the Wikipedia editorial community). In Figure 12 we observe the TAR as a function of the editorial activity for the pages related to the electoral process, for the German, Czech, Italian and Greek Wikipedia editions. As explained above, the TAR indicator is defined as the ratio between revisions in talk pages vs in article pages; therefore, by definition, lower values of TAR are only possible with higher values of the denominator, which implies a higher number of revisions; this explains why the bottom left quadrant in the scatterplots is empty, as lower values of TAR (y axis) are only possible with higher values of editorial activity (x axis). Apart from this effect, we observe, especially in the German edition, a positive relation between the 2 variables shown: the higher the overall number of revisions, the higher the TAR. This result, that might seem counterintuitive a priori since the TAR is inversely proportional to the revisions in the article page (edits) by definition, is showing that editorial activity comes with debate around it, taking place in the talk page (comments). In some cases, the debate activity in talk pages is so high around a topic that it may surpass the number of edits in the main article (as it occurs for Alternative for Germany and The Left pages in the German Wikipedia, that have a value of TAR greater than 1).

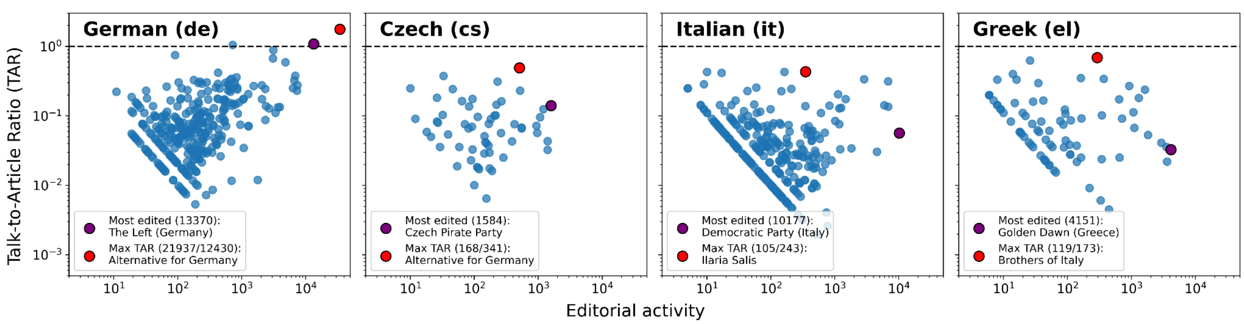

Figure 12: Talk-to-article ratio (TAR) across different Wikipedia editions. TAR for the pages in the whole dataset as a function of the editorial activity for 4 different language editions: German, Czech, Italian and Greek. In each plot, each point is a Wikipedia page. The pages with the highest TAR and Editorial activity are highlighted in red and purple, respectively.

Moreover, from the highlighted pages (with the highest TAR), we observe certain controversies that extend across Wikipedia editions. For example, in the Czech Wikipedia, the highest TAR is exhibited by the page Alternative for Germany, which corresponds to a national party that is not present in the Czech Republic. Similarly, in the Greek Wikipedia, the highest TAR is exhibited by Brothers of Italy, which is an Italian national party. Therefore, even though the contents and revisions share of each language edition mainly focus on their corresponding countries (as we have seen in Figures 3 and 4) and usually the debate takes place in that corresponding language edition (and the equivalent pages in other languages are translations with low debate), there are particular controversies and debates that extend beyond borders, leading to the observed patterns.

Figure 13 explores how TAR relates to various page characteristics across our dataset. The left panel shows the equivalent to Fig. 12 for the English Wikipedia, where we tend to observe a similar positive relationship. We also observe several articles, mainly biographies, with high TAR in correspondence with low editorial activity; in these cases the high value of TAR is due to the low number of overall edits (the TAR denominator) rather than to a high volume of activity in the talk page (the TAR numerator), indicating pages with limited activity; this is also the case of the page having the highest TAR. The central panel reveals a particularly low relationship between vandalism and TAR: pages experiencing more vandalism do not necessarily show elevated discussion-to-article ratios. This pattern makes intuitive sense, as vandalism on Wikipedia is solved fast with immediate reversion, with no need in principle for a subsequent talk page discussion about this. High-TAR pages in this context often represent battlegrounds where the community must explicitly negotiate content decisions rather than relying on implicit consensus.

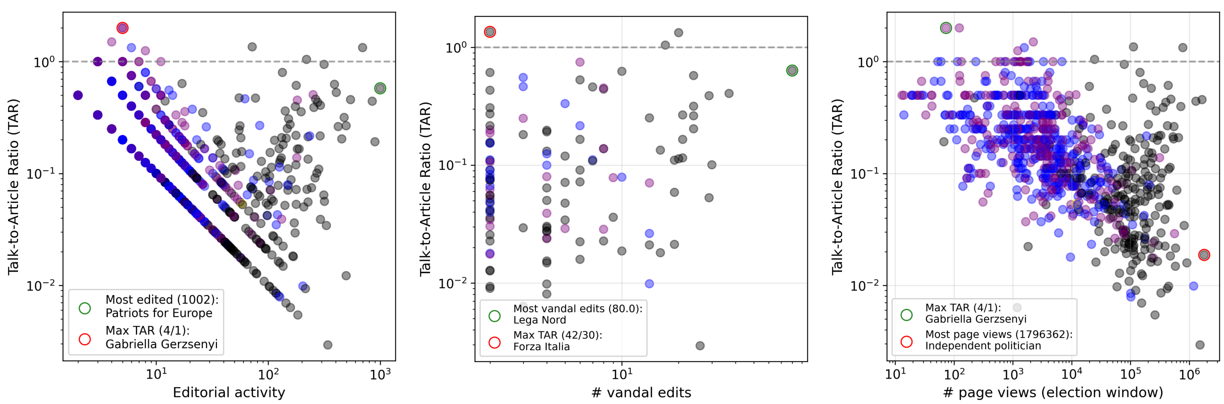

Figure 13. Talk-to-article ratio across different pages and relation with other variables. TAR for the pages in the dataset as a function of the editorial activity (left), the vandal activity (centre) and the page views in the electoral period (right). In each plot, each point is a Wikipedia page, coloured according to gender: male in blue, female in purple, non-binary in yellow and agender in black. The pages with highest Y-variable and X-variable are highlighted in red and green, respectively.

The right panel examines TAR in relation to page views, showing that high-attention articles tend to have moderate TAR values, while some less-viewed pages display very high ratios. This suggests two distinct scenarios: popular articles that generate both editing activity and discussion proportionally, and lower-profile articles where discussion predominates over direct editing, possibly due to uncertainty about content, protection policies limiting who can edit, or small groups of dedicated editors carefully deliberating over limited content. Moreover, page views split the biographies (with less attention) and the "agender" pages (institutional articles about parties and elections), with broader attention. This allows us to observe an inverse proportional relation between TAR and page views in biographies, which is not intuitive a priori.

The gender-coded points reveal that "agender" pages (institutional articles about parties and elections, depicted in grey) often are uniformly distributed across TAR values. Biographical pages show more variable TAR, with some high-profile politicians' articles generating intensive discussion relative to their size, likely reflecting controversies or disputes about appropriate content framing.

Attention dynamics

While until here we have focused on editors' activity, we now consider readership patterns by inspecting page views data. Previous literature found some degree of non-alignment between readers' and editors' attention, with the most read articles not necessarily corresponding to the most edited ones (Lehmann et al., 2014).

First, in Figure 14 we can see the overall number of views per language edition in our dataset, in logarithmic scale. In Figure 15 we see the temporal evolution of page views for a selection of language editions. While some events, like the election day, produce spikes in all the languages observed, some noticeable differences exist, with different patterns across languages and some noticeable peaks only in some language editions, showing national-level patterns in the attention of Wikipedia readers (which is a proxy for internet users).

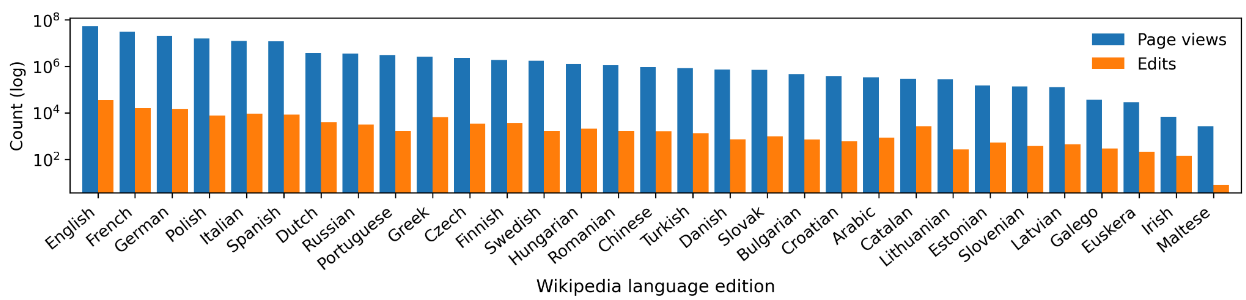

Figure 14. Attention ranking during the electoral period. Ranking of the total number of views in the pages selected in each Wikipedia language edition during the selected electoral period, from 23/05/2023 to 01/12/2024. The number of page views is plotted in log-scale due to the different orders of magnitude.

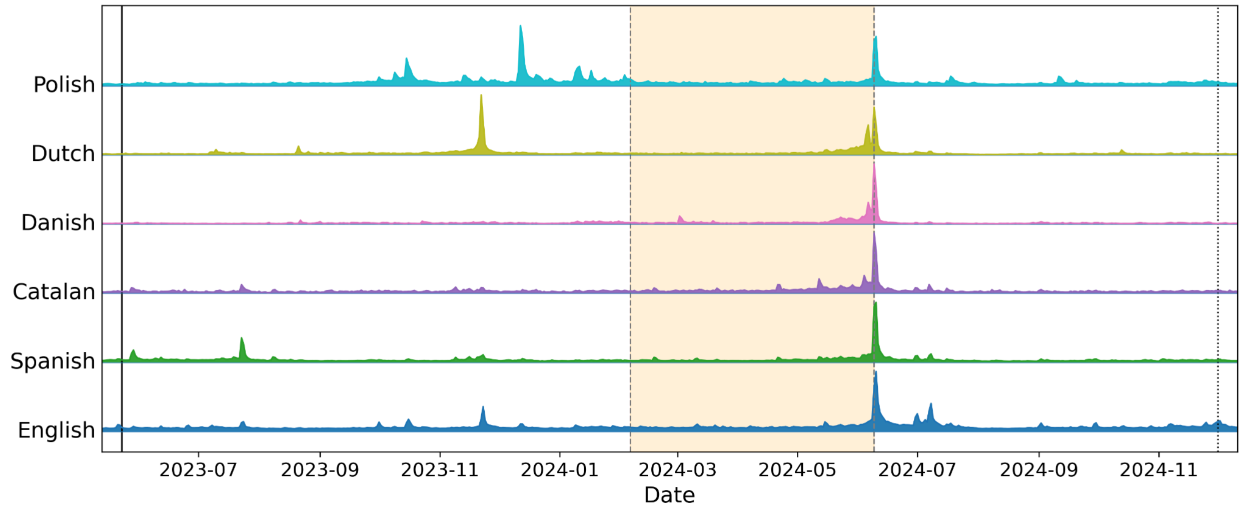

Figure 15. Daily attention through the electoral period. Page views time series for different Wikipedia language editions: Polish, Dutch, Danish, Catalan, Spanish and English. The orange shaded region delimited by grey dashed lines is the electoral campaign, the solid line is the Election announcement (23/05/2023) and the dotted line highlights the EU Commission formation (01/12/2024), which delimit the period in which we perform the analysis.

For example, in the Polish Wikipedia, beyond one peak in correspondence with the European Parliament Election on June 6th, 2024, and in line with the other language editions shown, we observe other significant peaks on October 15th and December 12th, 2023. The first date corresponds to the date of Polish national parliamentary elections, and the second to the day after a vote of confidence, held on December 11th in the Sejm (lower house of the bicameral parliament of Poland) and lost by the Prime Minister in charge. Analogously, in the Dutch Wikipedia we find a peak on November 22nd, 2023, in correspondence with the Dutch general election. In both cases, even though the selection of pages in our dataset is based on the European Parliament Elections, and focused on the parties and candidates running for those elections, the effect of national elections and politics seems to be stronger even on this selection of articles, as we observe the highest peaks for these language editions in correspondence of national events. This finding suggests that Wikipedia readers' attention may be triggered more heavily by national than European level events.

Looking at cross-country patterns in Figure 16, we notice that readers' attention seems even more focused on their own country than observed for editors (Figure 4), as the diagonal line appears even more marked. We observe more page views for male than female politicians in most language editions, with few exceptions like Finnish and Maltese, and comparable values in Romanian and Czech.

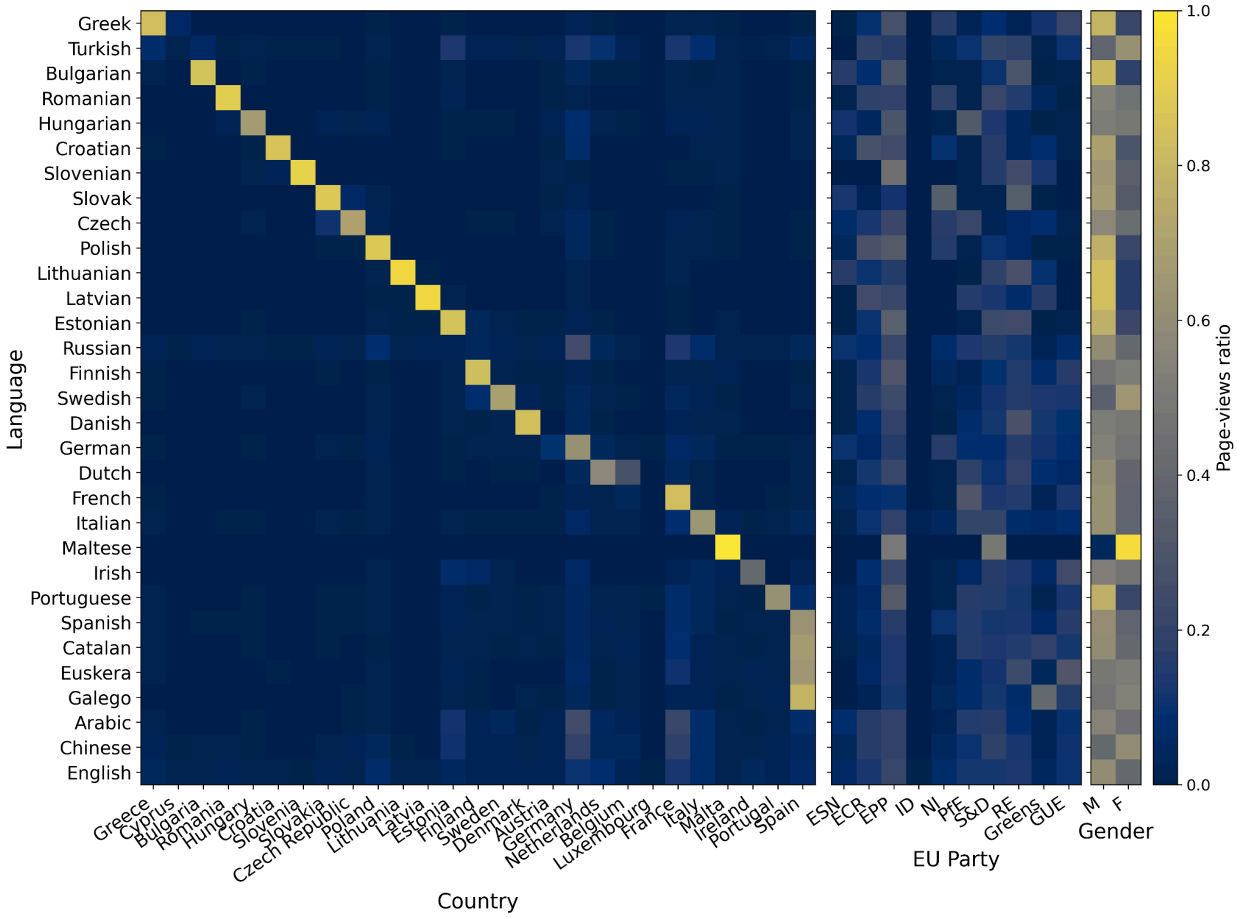

Figure 16: Page views share across different countries, European parties and genders. Proportion of page views in each language edition on all content about the different countries (left), EU parties (centre) and gender (right) and the total number of edits in the language edition.

To understand the relation between readers' attention and editor activity, in Figure 17 we look at how it relates to different metrics described above. Both in English (left) and Italian (right) we see a general alignment between page views and editorial activity, with some interesting outliers like Sahra Wagenknecht Alliance, emerging German political party who received significantly more attention from editors than from readers on the English Wikipedia, getting to attract on her page the highest number of edits when including vandalism, and the highest number of distinct editors, while receiving a much lower number of page views than other more popular pages, an order of magnitude lower than the most read article on the English Wikipedia, Alternative for Germany. Similarly, the page receiving the most vandalism edits on the English Wikipedia is Lega Nord, which is not among the pages receiving the highest numbers of page views. In the Italian Wikipedia, instead, the most read article, Roberto Vannacci corresponds to the one receiving the highest levels of editorial activity according to all the variables considered.

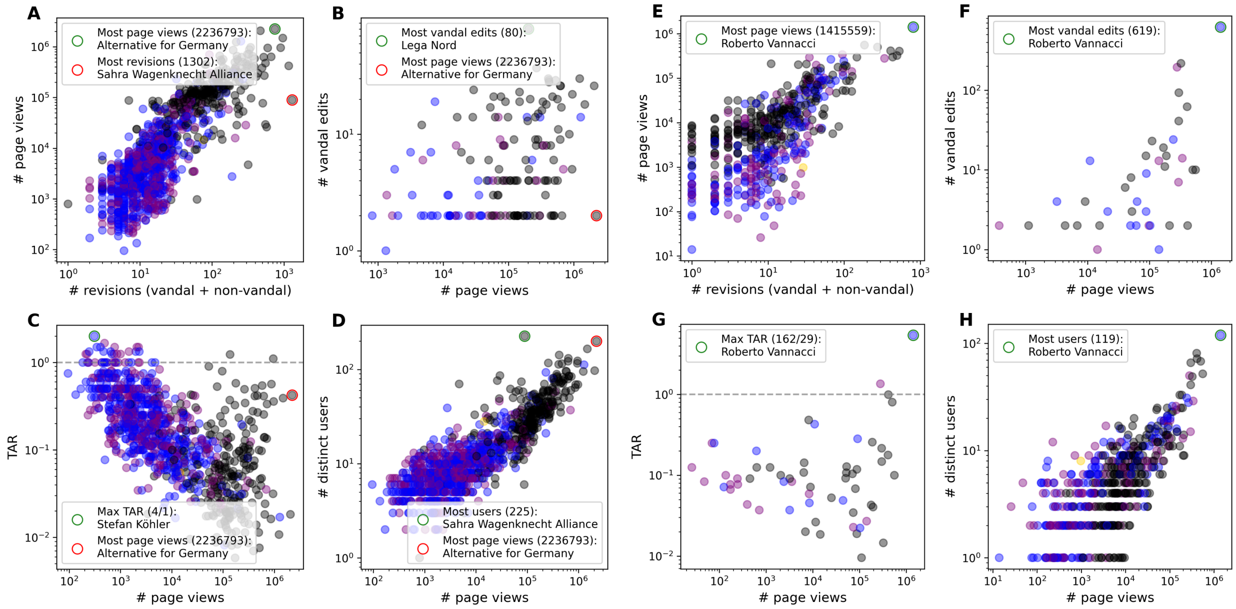

Figure 17: Attention patterns across different variables. Scatter plot for the pages in English (A-D) and Italian (E-H) Wikipedia of the number of page views with the previous defined metrics: total revisions (A, E), vandal edits (B, F), TAR (C, G) and distinct users (D, H). In each plot, each point is a Wikipedia page, coloured according to gender: male in blue, female in purple, non-binary in yellow and agender in black. The pages with highest Y-variable and X-variable are highlighted in red and green, respectively.

Multilingual analysis

In this section, we describe and analyse the lexicon used in the edits and comments through the electoral period in the Wikipedia pages related to the 2024 European Parliament election. This lexicon analysis is basically language-agnostic, but with the particularity that the filtration of stop words requires for a multilingual framework, since the necessary treatment cannot be the same for all languages. The filtered stop words are extracted from a popular GitHub repository[6].

Methodology

Added and deleted words extraction and processing

The extraction of added and deleted words from Wikipedia revisions was performed using the Wikipedia API. For each article, we retrieved all historical revisions in chronological order, including revision metadata (timestamp, author, user ID, comment, revision ID) and the complete content of each revision. Bot edits were systematically excluded by checking for the "bot" role in user metadata or the presence of "(bot)" in usernames and edit comments. To identify textual changes between consecutive revisions, we applied a word-level diff algorithm using Python's difflib[7], which compares the content of revision *n* with revision *n-1* after tokenizing both texts by whitespace. This process yielded two lists per revision: words added (tokens present in revision *n* but absent in *n-1*) and words deleted (tokens present in *n-1* but absent in *n*). The extraction was performed separately for both article pages and their corresponding talk pages.

The preprocessing pipeline applied multiple filtering stages to remove noise and non-substantive tokens from the extracted word lists. First, HTML tags, wiki markup elements (e.g., `<ref>`, `{{cite}}`), and common CSS/formatting artifacts were removed using regular expressions and predefined pattern lists. Second, unwanted punctuation characters were stripped, and accent marks were normalized using Unicode decomposition (NFKD normalization). Third, as mentioned before, language-specific and English stop words were filtered out, and tokens containing digits or matching noise patterns (e.g., CSS prefixes like "colorspanish") were excluded. Additionally, tokens shorter than 3 characters were removed to eliminate residual markup fragments. Finally, revisions containing any single word repeated more than 50 times were excluded entirely, as such patterns typically indicate automated spam rather than legitimate content contributions. This preprocessing ensured that subsequent analyses focused on substantive lexical changes reflecting actual editorial content rather than technical artifacts. At the end of this preprocessing pipeline, we have for each revision in our dataset, a cleaned list of added and deleted words by that edit.

[6] Stop words multilingual repository: https://github.com/stopwords-iso/stopwords-iso

[7] **difflib** is part of Python's standard library and provides utilities to compare sequences and identify changes (in terms of tokens) between consecutive revisions; see https://docs.python.org/3/library/difflib.html.

Deleted-to-Added words ratio (DAWR)

To quantify the balance between content addition and removal at the page level during the electoral period, we introduced the Deleted-to-Added Words Ratio (DAWR), defined as the ratio of total deleted words to total added words across all non-vandal revisions for each Wikipedia page. Formally, DAWR is computed as:

$$\mathrm{DAWR} = \frac{\text{deleted words} + 1}{\text{added words} + 1},$$

where the summation is taken over all legitimate (non-vandal) revisions of a page, and the constant +1 is added to both numerator and denominator to avoid division by zero and to prevent extreme values for pages with very few edits. A DAWR value close to 1 indicates balanced editorial activity with roughly equal amounts of content addition and deletion, suggesting either stable content undergoing refinement or active debate with back-and-forth changes. Values significantly below 1 indicate net content growth, typical of pages receiving substantial new information, while values above 1 suggest net content reduction, which may reflect quality control efforts, vandalism removal, or editorial disputes leading to content removal. This metric complements traditional edit count measures by capturing the directional nature of editorial changes and providing insight into the developmental trajectory of individual Wikipedia articles during the electoral period. This metric is introduced for the first time in literature via this report.

Results

Words added over time by user type and level

Figure 18 shows the temporal evolution of the number of words added to articles (upper panel) and talk pages (lower panel) throughout the electoral period, disaggregated by user type. We observe that registered users contribute the vast majority of content across both spaces, with their contributions showing marked peaks around the election date and in the subsequent months. Anonymous users (shown in red) contribute a small but consistent proportion of words, particularly to article pages, where their activity shows more pronounced spikes during the electoral campaign period. Bot activity (grey) is relatively small, and it is primarily concentrated in articles rather than talk pages, which aligns with the typical use of bots for maintenance tasks such as updating templates, fixing formatting, and reverting vandalism.

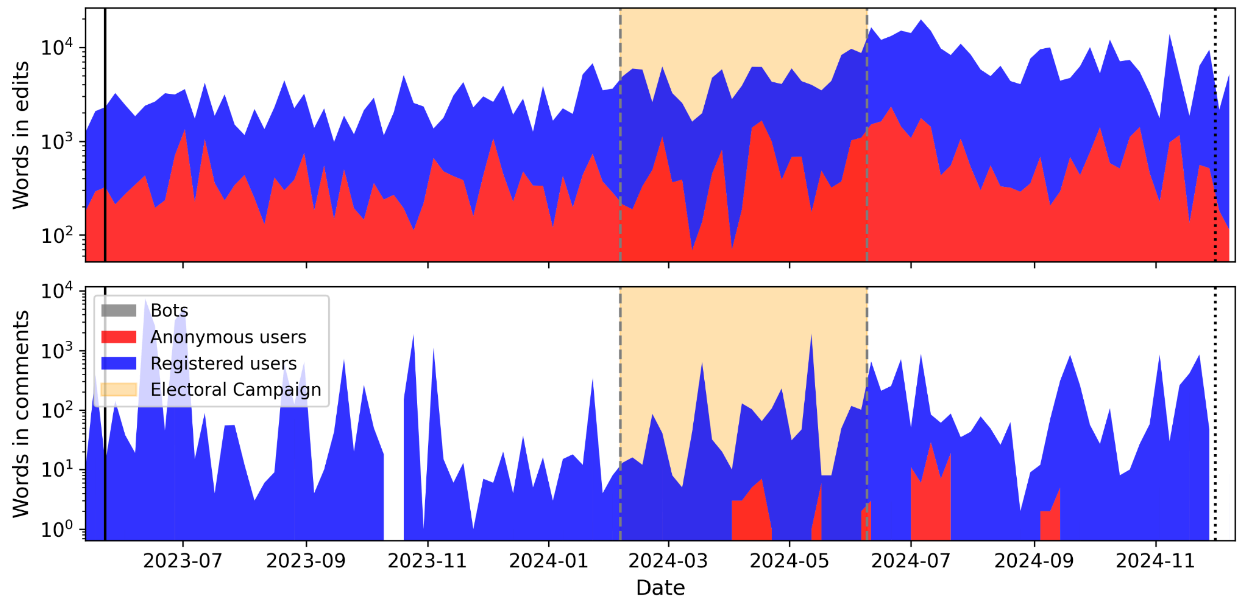

Figure 18: Number of words added through the electoral period by user type. Number of words added in edits (top) and comments (down) in the pages during the electoral period for all language editions. The share of the number of words is shown in colours: anonymous users in red, registered users in blue and bots in grey**.** The orange shaded region delimited by grey dashed lines is the electoral campaign, the solid line is the Election announcement (23/05/2023) and the dotted line highlights the EU Commission formation (01/12/2024), which delimit the period in which we perform the analysis.

The temporal patterns reveal that while all user types increased their activity during the electoral campaign (highlighted by the grey-shaded region), the most substantial peaks occur after the election date, particularly for registered users. This suggests that much of the content creation and discussion happened not only during the campaign itself but continued intensively in the period following the elections, likely reflecting the need to update articles with election results, document the formation of the new parliament, and discuss the implications of the electoral outcomes. The lower panel shows that talk page activity, while lower in absolute volume, follows similar temporal patterns, indicating coordinated efforts between content creation and community discussion.

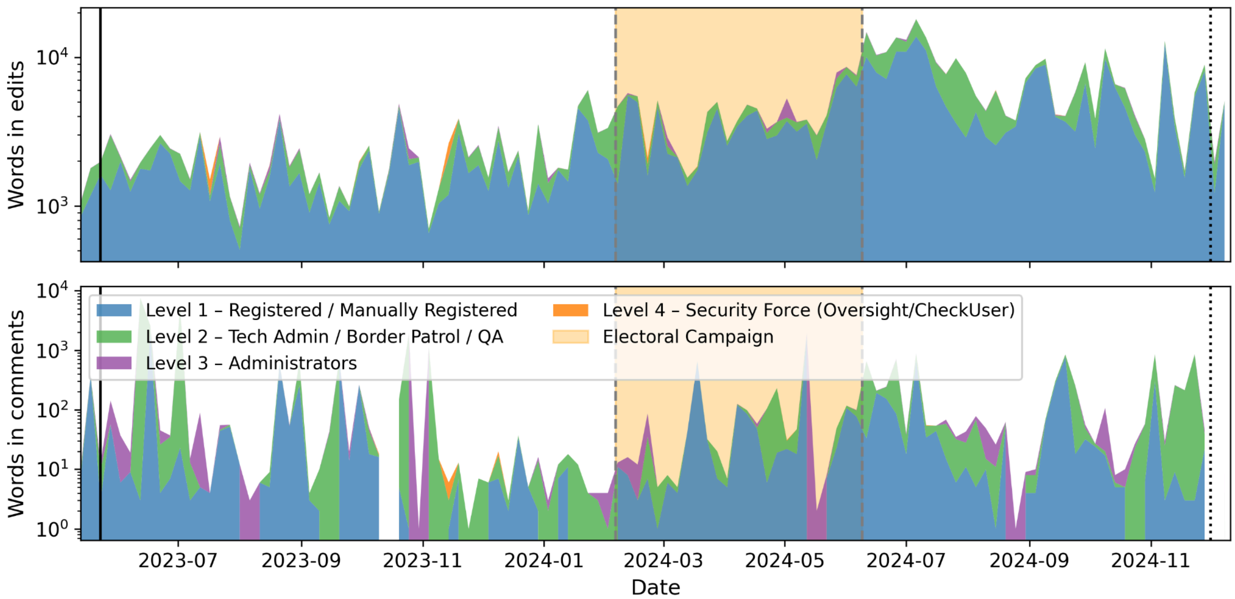

Figure 19: Number of words added through the electoral period by user level. Number of words added in edits (upper) and comments (lower) in the pages during the electoral period for all language editions. The share of the number of words through the user levels is shown in colours: Level 1 in blue, Level 2 in green, Level 3 in purple and Level 4 in orange**.** The orange shaded region delimited by grey dashed lines is the electoral campaign, the solid line is the Election announcement (23/05/2023) and the dotted line highlights the EU Commission formation (01/12/2024), which delimit the period in which we perform the analysis.

When examining word contributions by user level rather than type, Figure 19 reveals the hierarchical structure of editorial activity within Wikipedia communities. Level 1 users (registered users with basic privileges, shown in blue) dominate content creation across both articles and talk pages, contributing the majority of words throughout the entire period. This pattern reflects the fundamental principle of Wikipedia as a platform where most content comes from the broader community rather than from administrators or specialized roles. This hierarchical pattern of contribution aligns with previous findings on Wikipedia's organisational structure (Arazy et al., 2015), where a broad base of regular users produces most content while experienced users focus more on coordination and quality control.

Level 2 users (technical administrators, border patrol, and quality assurance roles, shown in green) maintain a consistent presence, particularly visible in talk pages where their moderation and quality control functions are more actively exercised. The contribution of Level 3 users (administrators, in purple) is proportionally small but steady, which is expected given their reduced numbers and their focus on oversight rather than direct content creation. Level 4 users (oversight and checkuser roles, in orange) show minimal contribution volumes, consistent with their even smaller number, and specialized security-focused functions.

The temporal distribution shows similar peaks across all user levels around the election period and afterwards, suggesting coordinated community response to electoral events. Notably, the proportion of activity from higher-level users increases slightly in talk pages compared to articles, reflecting their enhanced role in dispute resolution and content quality discussions.

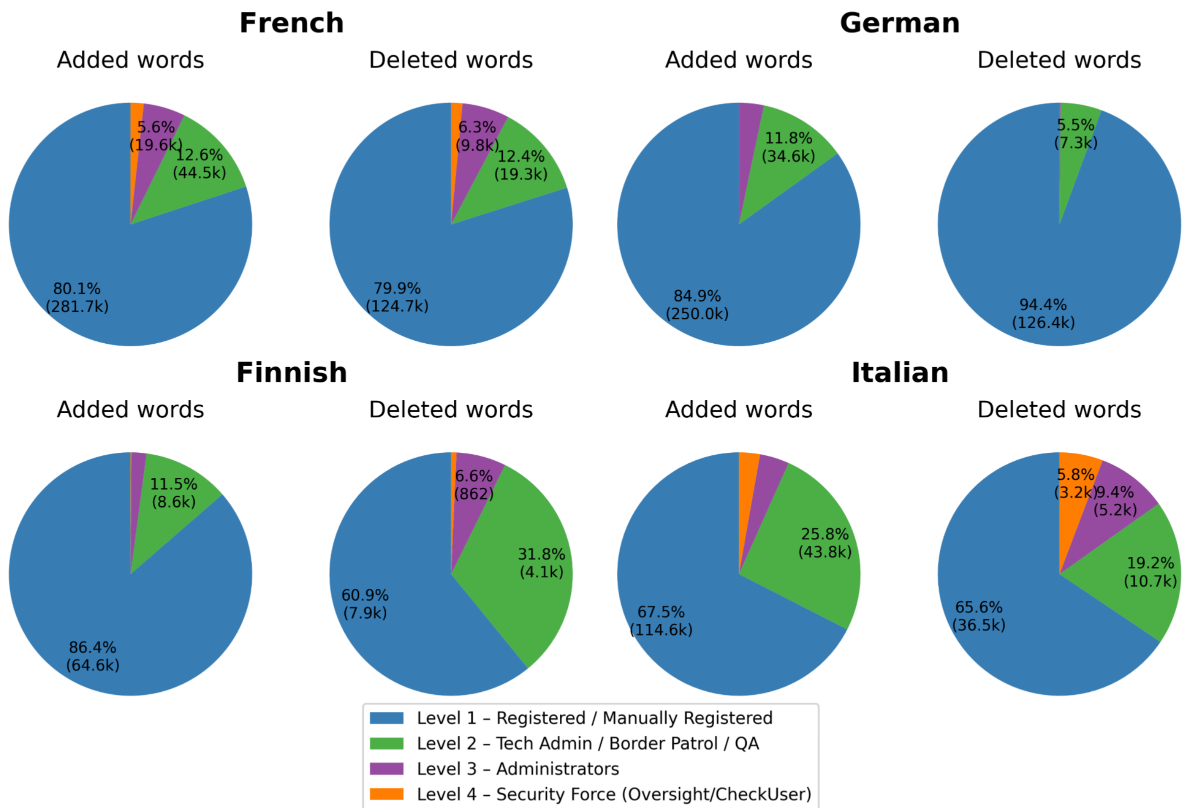

Figure 20: Share of added and deleted words through the edits by the different user tiers in different languages. Proportion of added words (left) and deleted words (right) by different user tiers for the 4 language editions: French, German, Finnish and Italian. The share of the number of words through the user levels is shown in colours: Level 1 in blue, Level 2 in green, Level 3 in purple and Level 4 in orange**.**

Figure 20 provides a comparative view of the distribution of editorial work across user levels for four major European language editions: French, German, Finnish, and Italian. For each language edition, the left panel shows the proportion of words added, while the right panel displays the proportion of words deleted by editors belonging to each user tier. Level 1 users (registered users) who generally constitute the vast majority of the user base, contribute a large majority of both added and deleted content. However, differences emerge among communities. The prominence of Level 1 users is marked especially in the German Wikipedia, while Level 2 users have a significant presence especially in the Italian and in the Finnish Wikipedia; interestingly, in the Italian community they are proportionally more active adding than deleting content, like in German, while the opposite happens in Finnish with Level 2 users accounting for over 30% of deleted words vs only 11.5% of added words. Third Level users, the Administrators, are proportionally much more active deleting than adding words both in the Finnish and in the Italian Wikipedia. Users from the highest tier have a noticeable impact especially in the Italian Wikipedia, and again more with deleted than added words.

Added and deleted words in articles vs talk pages

To understand how editorial discourse differs between encyclopedic content creation and community discussion, we examined the lexical patterns in articles versus talk pages separately. Figure 21 presents word clouds showing the most frequent added and deleted words in both spaces for the English Wikipedia during the electoral period.

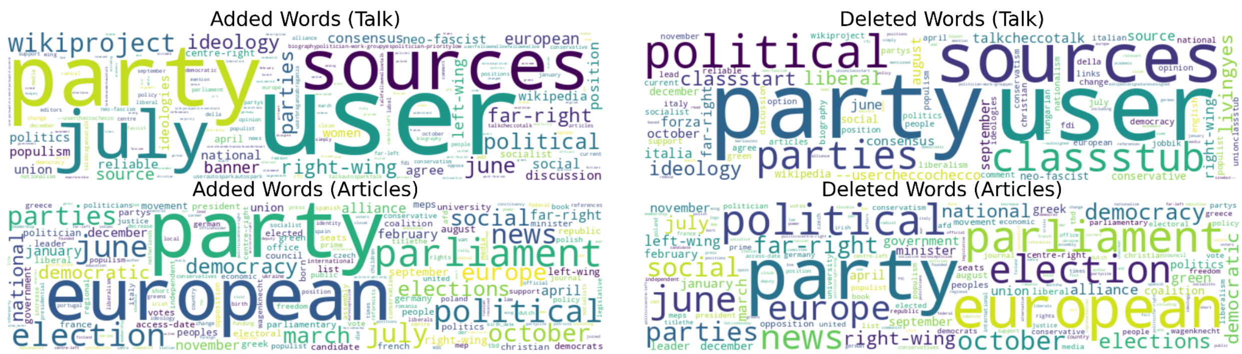

Figure 21. Added and deleted words cloud in discussions and articles. Added (left) and deleted (right) words in the talk pages (top) and articles (down) at the different revisions during the electoral period in the English Wikipedia. Word size is proportional to the sum of the frequency of each word across each collection of pages.

In article pages (lower panels), the added words (left) reflect encyclopedic content focused on political terminology: words like "party", "election", "parliament", "political", "member", and "group" dominate, alongside specific references to political actors and organisations. Geographic terms like "Germany", "France", and "European" are prominent, reflecting the multinational nature of the European elections. The deleted words (right) show similar political vocabulary, suggesting continuous refinement and updating of political information as the electoral situation evolved, with editors removing outdated or disputed content while adding new information.

Talk pages (upper panels) reveal a markedly different lexical landscape. The added words include more meta-discussion terms such as "article", "Wikipedia", "edit", "source", and "information", reflecting the collaborative and procedural nature of talk page discussions where editors coordinate their work and discuss article improvements. Policy-related terms appear more frequently here than in articles, as these spaces are where editors explicitly negotiate content decisions and invoke community norms. The deleted words in talk pages show similar meta-editorial vocabulary, though with less frequency overall, as talk page content is typically less subject to removal than article content, serving as a permanent record of community deliberation.

Added and deleted words (vandal edits)

Vandalism represents a particular challenge for Wikipedia communities, especially during politically sensitive periods such as elections. To characterize vandalistic behaviour, we analysed the lexical content of edits that were identified as vandalism through our filtering methodology (described in the Methods section). Figure 22 shows word clouds of added and deleted words specifically for vandal revisions in both talk pages and articles.

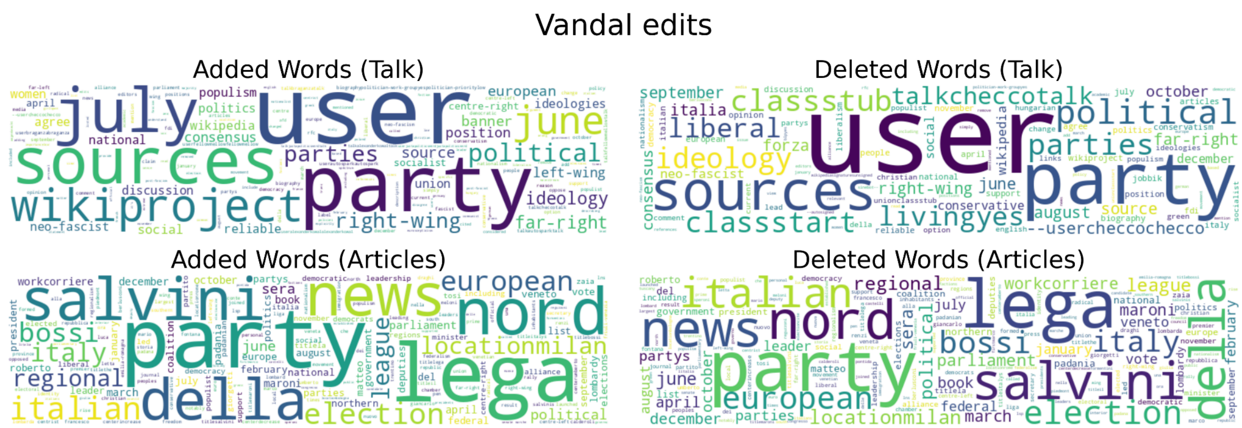

Figure 22. Added and deleted words cloud in Vandal revisions in discussions and articles. Added (left) and deleted (right) words in the talk pages (top) and articles (down) at the different vandal revisions during the electoral period in the English Wikipedia. Word size is proportional to the sum of the frequency of each word across each collection of pages.

The patterns in vandal edits differ substantially from legitimate editorial activity. In articles (lower panels), vandal additions include very specific words as “Salvini”, “news”, “Lega Nord”, “liberal”, “ideologies”, reflecting that in this context targeted vandalism is a reality. We also observe terms related to extreme political positions and personal attacks that editors attempted to insert into biographical articles of politicians. Notice that vandalism may imply deleting all content in a certain page or section, which makes this lexicon more difficult to understand. The relatively lower frequency of vandal activity in talk pages compared to articles reflects that vandals primarily target the publicly visible article content rather than community discussion spaces.

Added and deleted words by gender

The gender-disaggregated analysis of lexical patterns reveals interesting differences in how male and female politicians are discussed on Wikipedia during the electoral period. Figures 23 and 24 show added and deleted words for articles about male and female politicians in French and Spanish Wikipedia respectively.

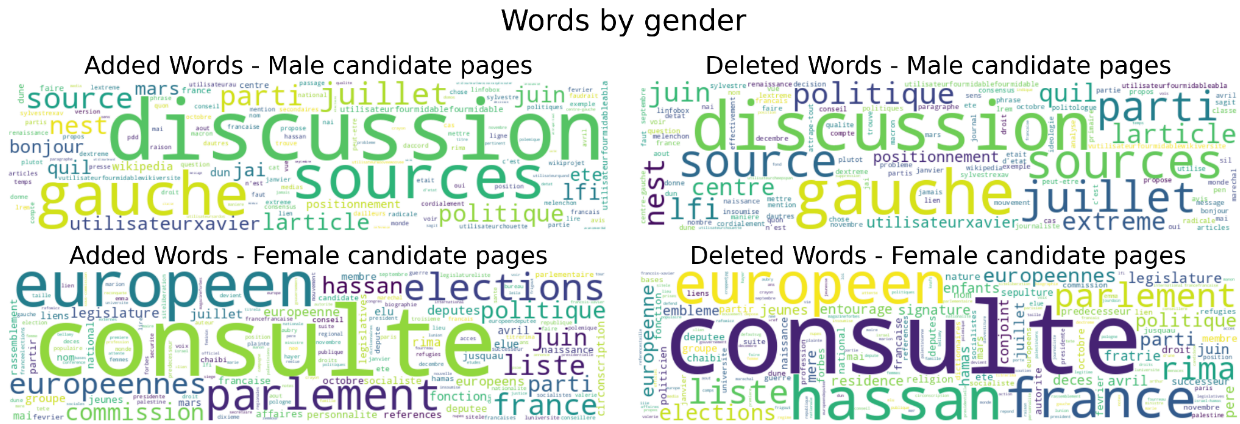

Figure 23. Added and deleted words cloud in discussions and articles, separated by gender in French Wikipedia. Added (left) and deleted (right) words in pages corresponding to present and past MEPs, aggregated by gender: male (top) and female (down), at the different revisions during the electoral period in the **French** Wikipedia. Word size is proportional to the sum of the frequency of each word across each collection of pages.

In the French Wikipedia (Figure 23), articles about male politicians show a broader range of political terminology and organisational references, with frequent mentions of specific parties, parliamentary groups, and political positions. The deleted words for male politicians include similar political vocabulary, suggesting continuous updating and refinement of political information. For female politicians, while the core political terminology remains prominent, we observe the appearance of specific controversial figures in the deleted words, notably "Rima Hassan", a left-wing politician who sparked significant debate.

In Spanish we observe among deleted words a prominence of the name of Irene Montero, a woman politician elected for Podemos (GUE EU party), and words "condena" and "supremo", which refer to the politician being convicted by the Supreme Court of Spain to pay a compensation of 18,000 euros to a man she accused of abuse. The word cloud reflects this heated debate about whether and how this information should be included in the article.

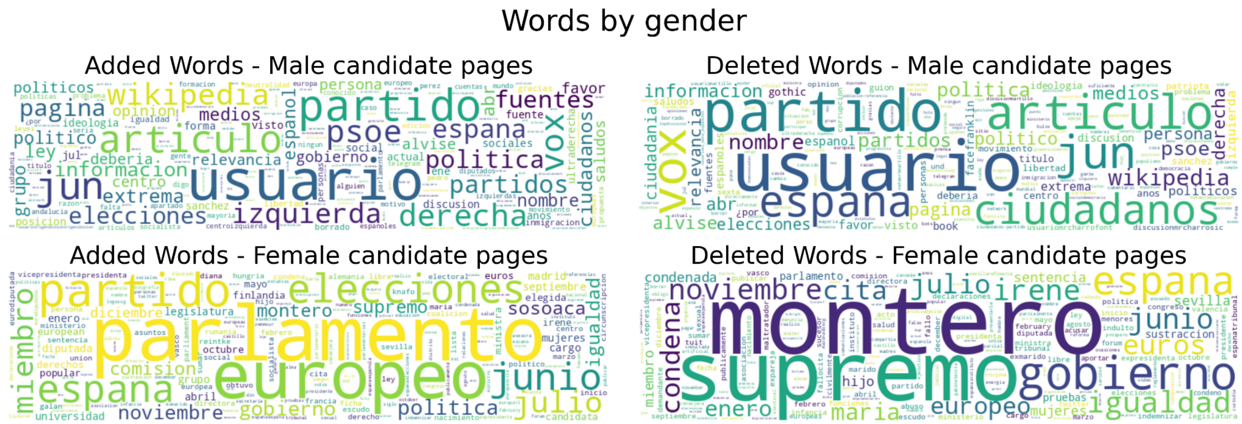

Figure 24. Added and deleted words cloud in discussions and articles, separated by gender in Spanish Wikipedia. Added (left) and deleted (right) in pagers corresponding to present and past MEPs, separated by the two common genders, male (top) and female (down), at the different revisions during the electoral period in the **Spanish** Wikipedia. Word size is proportional to the sum of the frequency of each word across each collection of pages..

Recent research on gendered online discourse during elections (Sala et al., 2025) found that female candidates in the 2024 European elections received toxic messages with higher intensity on social media platforms, including more sexist slurs and sexually explicit content. While here we do not focus on toxic messages, these results suggest that articles about female politicians involved in controversies attract particularly intense editorial attention and may require more active content moderation. However, it is important to note that these patterns may also reflect the specific controversies surrounding individual politicians rather than systematic gender bias, and further analysis would be needed to disentangle these factors.

Word frequency in article and talk revisions

Figure 25 examines the relationship between word frequencies in article pages versus talk pages for the Italian Wikipedia, providing insight into whether the same vocabulary is used in both collaborative spaces. Each point represents a word that appears in both contexts, positioned according to its frequency in articles (x-axis) and talk pages (y-axis), with color intensity indicating total occurrences.

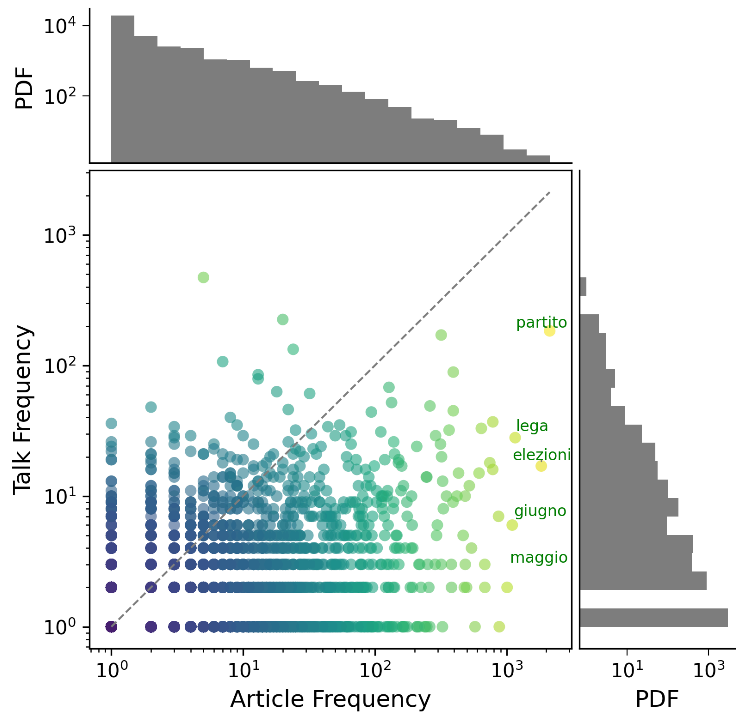

Figure 25. Words frequency distribution across articles and pages. Scatter plot of the words' frequency in the talk page and in the article page for the Italian Wikipedia. Each point is a word that appears in both article and talk pages, coloured according to the sum of occurrences, showing that there is a particular vocabulary for each of the pages. In grey, we also show the distribution of word frequency in log-log scale, showing a fat tail distribution.

The scatter plot reveals several distinct patterns. Most words cluster near the axes, indicating vocabulary that appears predominantly in only one space, either heavily in articles or heavily in talk pages, but not both equally. Words lying along the diagonal represent terms used with similar frequency in both contexts. The grey distribution shown in log-log scale demonstrates a characteristic fat-tailed distribution, where a few words account for the majority of occurrences while most vocabulary items appear infrequently. This power-law pattern is typical of natural language and has been consistently observed in Wikipedia content.

Words appearing frequently in articles but rarely in talk pages likely represent encyclopedic content terms—names of politicians, parties, geographic locations, and substantive political concepts that form the article content itself. Conversely, words appearing frequently in talk pages but rarely in articles include meta-editorial vocabulary about Wikipedia processes: "articolo" (article), "fonte" (source), "modifica" (edit), and references to Wikipedia policies and procedures.

This separation of vocabularies confirms that articles and talk pages serve genuinely different functions in the collaborative process: articles contain the encyclopedic knowledge product, while talk pages host the procedural and deliberative discourse through which that knowledge is negotiated and maintained.

Deleted vs Added Words (DAWR)

We examine the ratio between deleted and added words per article, captured by the DAWR, to compare editing patterns among pages. In Figure 26, we observe the distribution of this variable as a function of other article metrics.

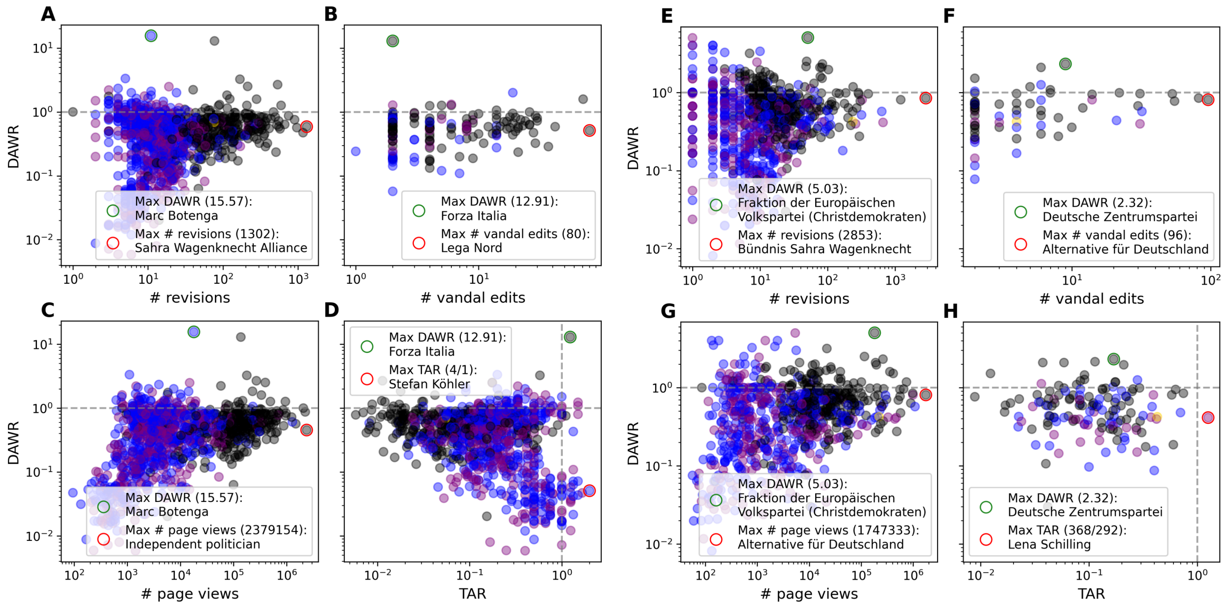

Figure 26. **DAWR patterns across different variables.** Scatter plot for the pages in **English** (A-D) and **German** (E-H) Wikipedia of the DAWR (Deleted to Added Words Ratio) with the previous defined metrics: total revisions (A, E), vandal edits (B, F), page views (C, G) and TAR (D, H). In each plot, each point is a Wikipedia page, coloured according to gender: male in blue, female in purple, non-binary in yellow and agender in black. The pages with highest Y-variable and X-variable are highlighted in red and green, respectively.

The Deleted-to-Added Words Ratio (DAWR) provides insight into the editorial stability and developmental trajectory of Wikipedia articles during the electoral period. Figure 26 shows scatter plots examining how DAWR relates to various page metrics for English and German Wikipedias.

Panels A and E reveal a generally positive relationship between total revisions and DAWR, indicating that more heavily edited pages tend to have higher ratios of deleted to added words. This pattern suggests that mature articles with extensive editing histories undergo more refinement and content removal as the community iteratively improves quality and removes outdated, disputed, or unnecessary information. However, the relationship is not strictly linear, with considerable variance, especially among highly edited pages, reflecting diverse editorial dynamics across different topics.

The relationship with vandal edits (panels B and F) shows that pages experiencing more vandalism tend to have higher DAWR values, which is intuitive since vandalism removal directly contributes to word deletion. Notably, some pages with moderate vandalism levels display very high DAWR, possibly indicating articles where vandalism removal is compounded by substantive content disputes requiring significant deletion of legitimate but disputed content.

Page views (panels C and G) show a weak positive correlation with DAWR, suggesting that high-attention articles tend to undergo more content refinement. This may reflect both higher scrutiny from readers who identify problems, and more active editor engagement on popular topics. The Talk-to-Article Ratio (panels D and H) displays interesting patterns, with some high-TAR pages showing low DAWR, indicating that intensive discussion often is not related to content deletion, especially for biographies (no “agender” pages).

Comparing English and German Wikipedias, we observe similar overall patterns but with the German Wikipedia showing somewhat tighter clustering in several relationships, possibly reflecting more consistent editorial practices or different community norms around content stability and deletion.

Policies analysis

In this section, we present the analysis of policy citations by different Wikipedia language communities. Complementing the legal analysis in deliverable D2.3, we inspect the edit history of different linguistic communities to shed light on where, how, why and by whom policies are cited within editorial activity, both in article revisions and in talk pages.

Methodology

Policy mention extraction

The most typical way to cite a Wikipedia policy is through "WP:[policy name or acronym]", where "WP" stands for "Wikipedia" and indicates that the page belongs to the namespace dedicated to community norms and issues, and is followed by the name of the policy or its acronym or abbreviation[8]. For example: "WP:Verifiability", "WP:NPOV" as an acronym for "Neutral point of view", "WP:OR" or "WP:NOR" for "No original research". However, this citation methodology is not only used for policy citations, but it is used to mention the action of certain bots, use of templates, etc. This citation takes place in the edit summary or at the comments in the talk page.

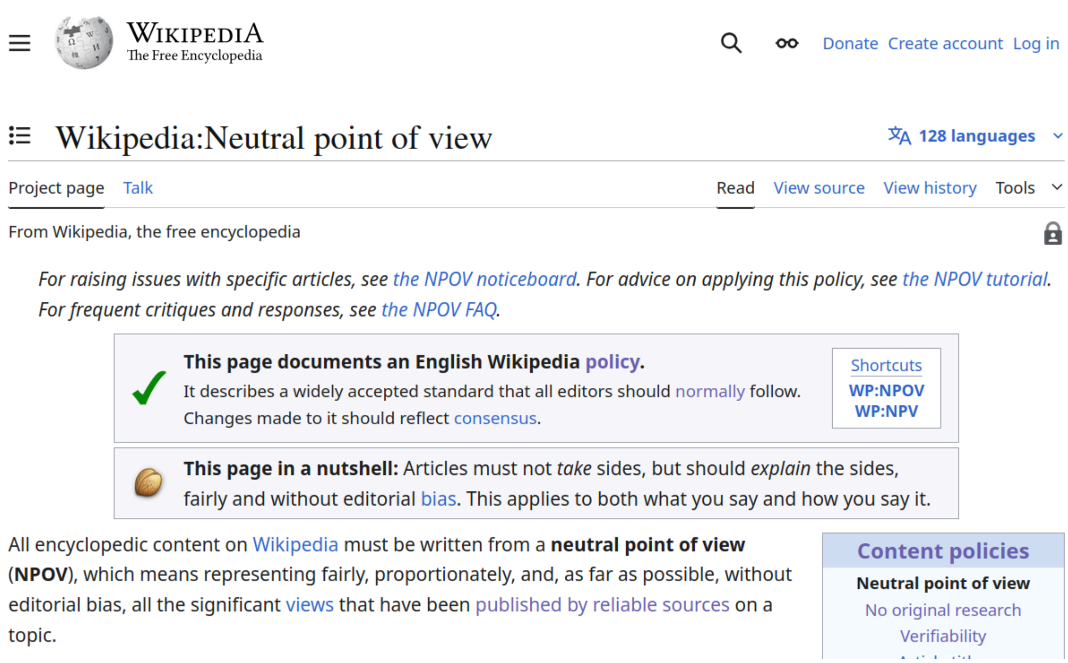

Figure 27. Wikipedia Neutral point of view policy page. Screenshot at date December 15th, 2025, from the page "Wikipedia: Neutral point of view" on the English Wikipedia.

---

[8] See the designated page on Wikipedia about policies and guidelines: https://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines.

In all the major language editions, each policy or guideline has a dedicated page, where its shortcuts are indicated in a structured way in a template, as shown in Figure 27 below for the case of Neutral Point of View, where two shortcuts are indicated: "WP:NPOV" and "WP:NPV".

Furthermore, each policy page may have subsections with their own shortcuts; for example, the NPOV policy implies "fairly representing all significant viewpoints that have been published by reliable sources, in proportion to the prominence of each viewpoint in those sources" and "giving due weight and avoiding giving undue weight meaning articles should not give minority views or aspects as much of or as detailed a description as more widely held views or widely supported aspects". Therefore, "Due and undue weight" is a subsection of NPOV, with shortcuts "WP:DUE" and "WP:UNDUE", as shown in the Figure below; similar other subsections exist such as Balance ("WP:BALANCE") or Balancing aspects ("WP:BALASP") or Making necessary assumptions ("WP:MNA"). In the case of "Neutral point of view", we find 16 subsections having some shortcut.

Due and undue weight

"Wikipedia:UNDUE" redirects here; not to be confused with Wikipedia:UNDO.

Neutrality requires that mainspace articles and pages fairly represent *all* significant viewpoints that have been published by reliable sources, in proportion to the prominence of each viewpoint in those sources.[c] Giving **due weight** and avoiding giving **undue weight** means articles should not give minority views or aspects as much of or as detailed a description as more widely held views or widely supported aspects.

Shortcuts
WP:DUE
WP:UNDUE

Figure 28. Wikipedia Neutral point of view subsections. Screenshot at date December 15th, 2025, from the page "Wikipedia: Neutral point of view" on the English Wikipedia.

Extraction process for the English Wikipedia

In order to identify references to community policies, we followed an iterative process combining a bottom up and a top-down approach. We started from the English Wikipedia, where we found the most complete categorization of policies, and extended it to the other language editions.

On the one hand, we followed a top-down approach to extract a list of policies that we considered relevant to our study, i.e. policies related to content and conduct, and guidelines related to content and behaviour. We did this by considering all policies and guidelines belonging to these categories or to their relevant subcategories:

- https://en.wikipedia.org/wiki/Category:Wikipedia_policies
- https://en.wikipedia.org/wiki/Category:Wikipedia_content_policies
- https://en.wikipedia.org/wiki/Category:Wikipedia_content_guidelines
- https://en.wikipedia.org/wiki/Category:Wikipedia_behavioral_guidelines
- https://en.wikipedia.org/wiki/Category:Wikipedia_conduct_policies
    - https://en.wikipedia.org/wiki/Category:Wikipedia_edit_warring
    - https://en.wikipedia.org/wiki/Category:Wikipedia_personal_attacks

For each of the policies identified, we extracted all of its shortcuts listed in the corresponding page, as well as all the shortcuts of all its subsections.

On the other hand, as a first approximation with a bottom up approach, we detected all references to pages in the Wikipedia namespaces, including abbreviations, by identifying any word preceded by the prefix “Wikipedia:” or “WP:”, or its variants “WP_” or “WP ”. In this way we extracted a first list of all the words, phrases and acronyms used in our dataset with some of these prefixes, as a proxy for the mentions of policies and community norms and pages.

Afterwards, we matched the resulting list of mentions to the Wikipedia namespace with the list of policy names and shortcuts retrieved from the policies and guidelines categories, to have a list of all the occurrences of these policies within our dataset and discard mentions to other pages from the Wikipedia namespace. This matching allows for small variants in the policy names and shortcuts to match the citation techniques from Wikipedia editors. For example, the policy Neutral Point of View is often cited as Neutral_point_of_view.

For the English Wikipedia, we went one step further and also considered the cases in which a policy name or some of its shortcuts appear even without the “WP” prefix, which was spotted several times through the dataset. In this case, we only considered cases when the policy name or shortcut was included in capital letter, and excluded acronyms of less than two characters as they would introduce noise (the presence of shortcuts “OR” and “AND” are trivial examples of noise that is filtered with this filter). We manually checked the occurrences obtained in this way and found that with these constraints we were not including noise.

Finally, as a given policy or some of its subsections could be mentioned through a variety of shortcuts, we merged all the occurrences related to the same policy and labelled with one of the principals shortcuts.

Extraction process for other language editions

For the other language editions, we followed an analogous approach: on the one hand, we followed interlanguage wiki links from the policies identified in the English Wikipedia, and then extracted all shortcuts from the corresponding page. In some language editions, like German wiki, the shortcuts in the wikitext do not follow the same structure as in English wiki, and retrieving this information is not as simple. To solve this, we also extracted all shortcuts that, when searched on Wikipedia, redirect to the policy page (which can be extracted directly from the API). We considered the possibility of extracting the policies for each language edition from its category structure or policy lists, but we did not find in any language a categorization of policies comparable in size and quality to the one from the English Wikipedia; in most cases, if categories or lists of policies exist, they do not separate policies related to content and behaviour from policies and guidelines related to other aspects, such as style, or even from other community pages. Therefore, we decided to stick to the categorization from the English Wikipedia, which also has the advantage of selecting the same set of policies consistently across the different language editions considered. The disadvantage is that we might miss some policy existing in a specific language and not in English in case they exist; however, in our observation from the major language editions we considered this was not the case, at least among the most frequently mentioned policies. The complete list of policy shortcuts and policies analysed for each language edition are available at a Zenodo folder (Abella and Laniado, 2025b).

For each language edition we detected all the occurrences of strings of characters preceded by the "Wikipedia" or "WP" prefix in its different variants, in English and in the considered language version when different. For example, in the Catalan Wikipedia, the Wikipedia namespace is named "Viquipèdia" and its abbreviation is "VP"; however editors do often use the English name or abbreviation, so we also included the English version of the prefixes.

We then matched all the resulting occurrences to the list of policies and shortcuts obtained for the given language as described above. Again, as we observed that editors often tend to use the English conventions even when not officially listed in the corresponding page, we also matched policy mentions with their English name or shortcut. In cases when a shortcut overlaps between English and the local language, and they correspond to different policies, we considered it as a citation of the local language policy.

Given the limited amount of data available in our dataset, the results for most language editions considered are not meaningful. Therefore, we performed the extraction for a selected set of language editions: English, German, Spanish, French, Italian, Portuguese, Polish, Dutch and Russian, in which the results where we found a significant number of occurrences of policy mentions. In this report, we show representative examples of the results computed for particular languages.

Policy mention textual context

In order to characterise the context in which different policies are mentioned, we collected the words co-occurring with each policy citation

We retrieved policy mentions from edit summaries and from talk pages comments; for each policy mention occurrence, we collected all the words co-occurring within the same edit summary, or talk page comment, removing the cite itself. We kept these two spaces separated to be able to compare activity in the article and in the talk page, and we merged them for some general analyses. This methodology creates a collection of context words associated with each policy of our dataset.

TF-IDF normalization

To identify distinctive vocabulary associated with each Wikipedia policy, we applied Term Frequency-Inverse Document Frequency (TF-IDF) normalization to the context text surrounding policy citations. TF-IDF (Salton and Buckley, 1988) emphasizes terms that are distinctively associated with a specific document (in our case, a specific policy), while down-weighting terms that appear frequently across many documents. For each policy, we aggregated all context text (extracted during policy detection from the edit summaries and talk page comments) into a single document, creating a corpus where each document represents the complete contextual usage of one policy across all revisions. Prior to vectorization, we applied a policy-specific preprocessing pipeline that removed: (1) Wikipedia-specific markup and HTML artifacts already filtered during word extraction, (2) common stopwords, (3) policy-related terms to avoid circular emphasis (e.g., variants of the policy's own shortcut code, such as "NPOV", "npov", "neutrality" for the Neutral Point of View policy), and (4) generic Wikipedia terminology (e.g., "wiki", "page", "edit", "user"). We then constructed TF-IDF vectors with unigram tokens, a minimum document frequency of 2, and a token pattern requiring at least 3 alphabetic characters. This normalization scheme down-weights terms that appear frequently across many policies while emphasizing terms that are distinctively associated with specific policies, enabling us to characterize the lexical signature of each policy through its top-weighted terms.

Policy-related deleted words

To assess which content tends to be deleted when invoking a given policy, for each occurrence of a policy citation within an edit summary, we associated with it all the words deleted in the corresponding revision (if there were any). We then aggregated all the deleted words associated with different citation occurrences of the same policy, creating a collection of deleted words associated with each policy.

Policy network and clustering

To understand the relationships between different Wikipedia policies and how they form thematic groups by its citation context, we constructed a similarity network based on the lexical context in which policies are mentioned by editors. This approach allows us to identify which policies tend to be invoked in similar editorial situations and discourse patterns during the 2024 European Parliament elections.

Policy context similarity measure

We quantified the similarity between policies using their associated textual contexts. For each policy, we aggregated all context text (from edit summaries and talk page comments where the policy was cited) into a single document, creating a corpus where each document represents the complete contextual usage of one policy across all revisions during the electoral period. To extract meaningful lexical features, we applied Term Frequency-Inverse Document Frequency (TF-IDF) vectorization to this corpus, as described above.

We then computed pairwise cosine similarity between the TF-IDF vectors of all policies, yielding a similarity matrix *S,* where each entry represents the lexical similarity between the two policies. Cosine similarity is particularly appropriate for text analysis as it measures the angle between document vectors, effectively capturing semantic proximity while being robust to document length variations (Manning et al., 2008).

Similarity network and clustering

From the similarity matrix, we constructed an undirected weighted network $G = (V, E)$ where nodes V represent policies and edges E connect policy pairs according to the similarity. To balance network connectivity and interpretability, we filtered edges with similarity below a threshold of 0.2, ensuring that the resulting network captured meaningful relationships while avoiding excessive density.

To identify thematic clusters of policies, i.e. groups that tend to be invoked in similar editorial contexts, we applied the Louvain method for community detection (Blondel et al., 2008). The Louvain algorithm is a hierarchical modularity-optimization method that efficiently partitions networks by iteratively maximizing modularity, a measure of the density of edges within communities compared to edges between communities. The Louvain method is particularly well-suited for our analysis as it handles weighted networks naturally, scales efficiently to networks with hundreds of nodes, and does not require specifying the number of clusters *a priori*.

The resulting partition assigns each policy to a community (cluster), which enables us to characterize distinct moderation and content governance practices, as policies within the same cluster tend to be invoked in similar editorial situations.

For visualization and interpretation, we represent the network using force-directed layout algorithms (Kamada and Kawai, 1989; Fruchterman and Reingold, 1991), colouring nodes according to their cluster membership and sizing them proportionally to their node strength (sum of edge weights), which indicates how central a policy is within the similarity network. We also visualize the cluster structure through hierarchically-ordered heatmaps of the similarity matrix, where rows and columns are reordered to group policies by cluster, making the block structure corresponding to communities visually apparent.

Policy Citation ratio

In order to be able to quantify the relative weight of policy mentions within editorial activity in a given context (e.g., in a Wikipedia edition, or on a specific page or group of pages) we defined an indicator for measuring policy citation activity in proportion to overall editing activity on a page:

the PCR (Policy Citation Ratio), defined as the ratio between the number of edits including mentions to some of the policies included in our dataset (content and behavioural policies), and the overall number of edits.

Moderation policies

The way in which we selected policies ensures they are all related to content and behaviour, thus avoiding other kinds of policies and guidelines, e.g. related to style or other aspects of editing; still, not all policies are strictly related to content moderation. Therefore, to have a list of the policies more strictly related to content moderation and in particular in the context of elections, we decided to rely on manual coding.

Coding was made by the UvA team, taking advantage of the team's expertise and work examining Wikipedia community policies in relation to elections, as described in particular in deliverable D2.3. Out of the list of policies included in our dataset with at least one occurrence in the English Wikipedia, 16 policies were selected as "moderation policies", i.e. as strictly relevant to content moderation:

- Arbitration Committee/CheckUser and Oversight
- Arbitration/Policy
- Banning policy
- Biographies of living persons
- Citing sources
- Conflict of interest
- Disruptive editing
- Harassment
- Neutral point of view
- No original research
- Protection policy
- Reliable sources
- Sockpuppetry
- Spam
- Verifiability
- What Wikipedia is not

Based on this list we organized the 16 identified moderation policies into 5 functional categories based on their primary governance objectives within Wikipedia's content moderation framework. The categorization scheme was developed through analysis of each policy's scope, purpose, and role within Wikipedia's governance structure. The 5 categories are:

- **Source Quality** encompasses policies that establish evidentiary standards for encyclopedic content, requiring that information be attributable to reliable published sources and verifiable by readers.
- **Content Neutrality** groups policies that mandate impartial representation of viewpoints and define the encyclopedic scope of Wikipedia.
- **User Conduct** includes policies addressing interpersonal behaviour and ethical standards for editors, encompassing harassment, disruptive editing, conflicts of interest, paid editing disclosure requirements, and dispute resolution mechanisms for serious conduct violations through arbitration.
- **Account Control** consolidates policies and procedures governing account-level restrictions and their enforcement, including banning and blocking mechanisms, detection of usage fraud and multiple account abuse, among others.
- **Content Protection** comprises policies specifically designed to safeguard sensitive content and prevent harmful material, particularly regarding biographies of living persons.

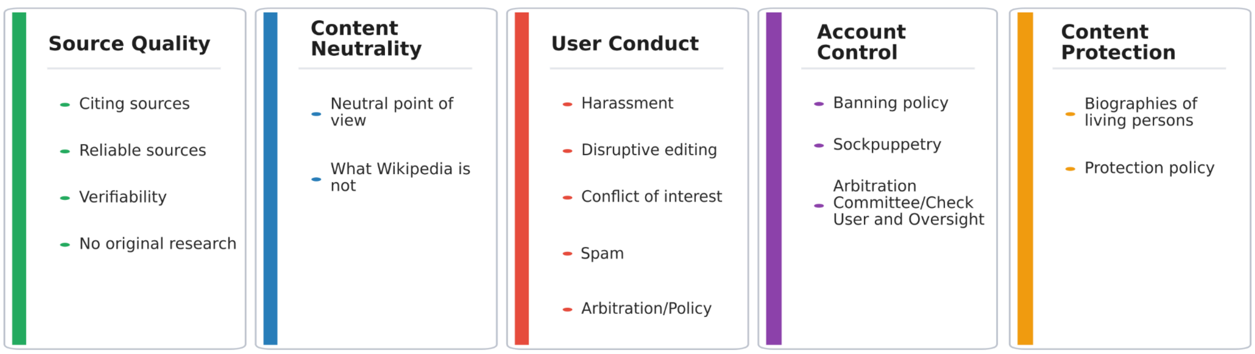

Figure 29. Moderation policies categorization. Categorization of the 16 moderation policies into the 5 functional categories according to their policy scope and usage.

The corresponding policy titles that belong to each category are shown in Figure 29. This functional categorization enables analysis of how different aspects of Wikipedia's governance system are deployed during politically sensitive periods and allows us to reveal specialization patterns across user types, experience levels, and temporal dynamics throughout the electoral cycle.

Cross-language results

We first have a look at results across languages and countries, EU Parties and genders, shown in Figure 30. Only 9 languages are included, the ones for which we found a significant number of policy citations. As seen with other measures of activity and attention, policy citations are mostly concentrated on the countries related to each language community; in English, we find a specially high proportion of policy citations on content related to Italy, and in Russian and French on content related to Germany.

We find more policy citations on pages related to men, especially in Italian and Polish, while the opposite happens in French and Dutch.

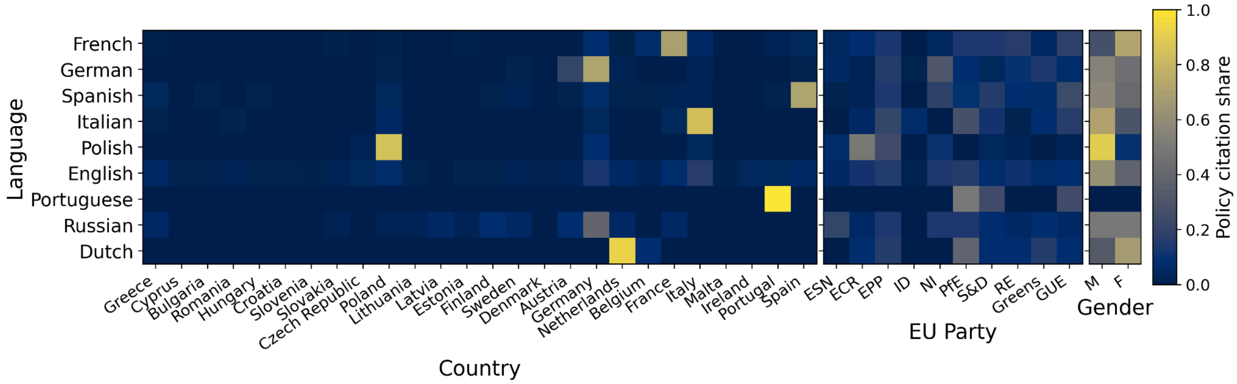

Figure 30. Policies citation share across different countries, European parties and genders. Proportion of revision citing policies in each language edition on all content about the different countries (left), EU parties (centre) and gender (right) and the total number of edits in the language edition.

Results for the English Wikipedia

We now present the results obtained for the English Wikipedia. We treat the English separately not only because of the special attention it deserves as the largest and lingua franca, taken as reference for deliverable D2.3, but also because for English we followed a more sophisticated approach for the extraction process in order to be able to identify a higher number of policy citations, and this would make the results not comparable in a straightforward way to the other language editions.

Most mentioned policies

First, we look at the policies that are more frequently cited by editors. The most cited is Consensus, followed by Vandalism and Reliable sources; and then with lower amounts of citations: Disrupting editing, Neutral point of view, No original research, Edit warring, rollback, Biographies of living people, Sockpuppetry, Spam, Verifiability. We refer to deliverable D2.3 and to the English Wikipedia for a description of each of these policies; the rest of the analysis below will help to shed further light on these policies.

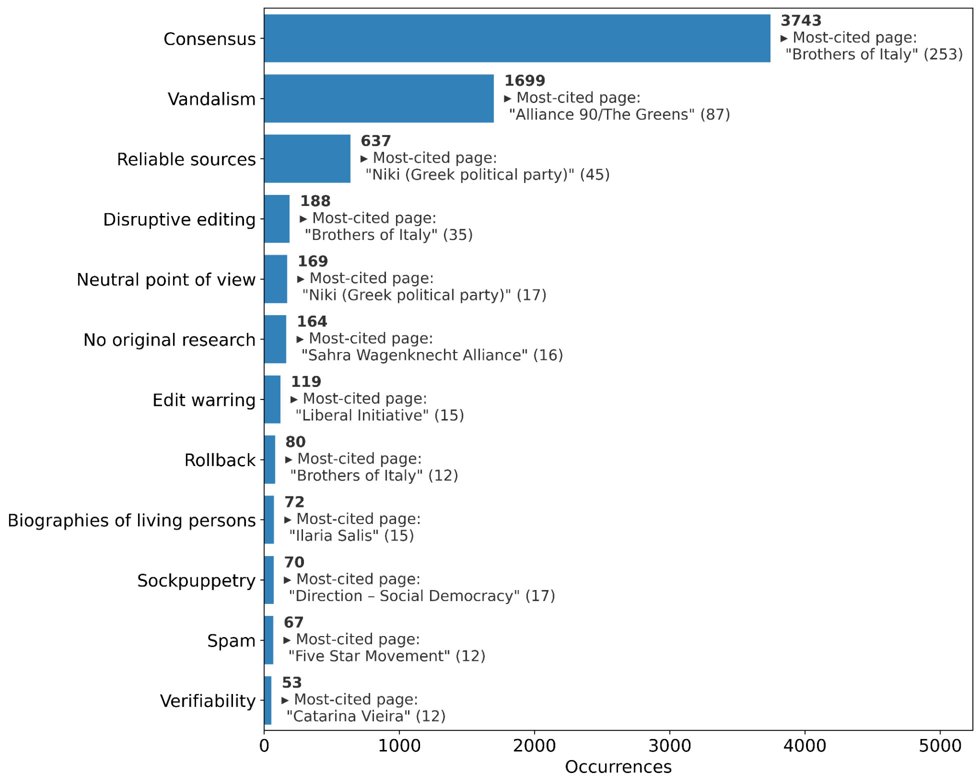

Figure 31: Top 10 mentioned policies in the English Wikipedia. Histogram of citations of each policy during the electoral period. In bold is shown the number of total occurrences, the number of occurrences for the page with the highest amount of policy citations of that particular policy.

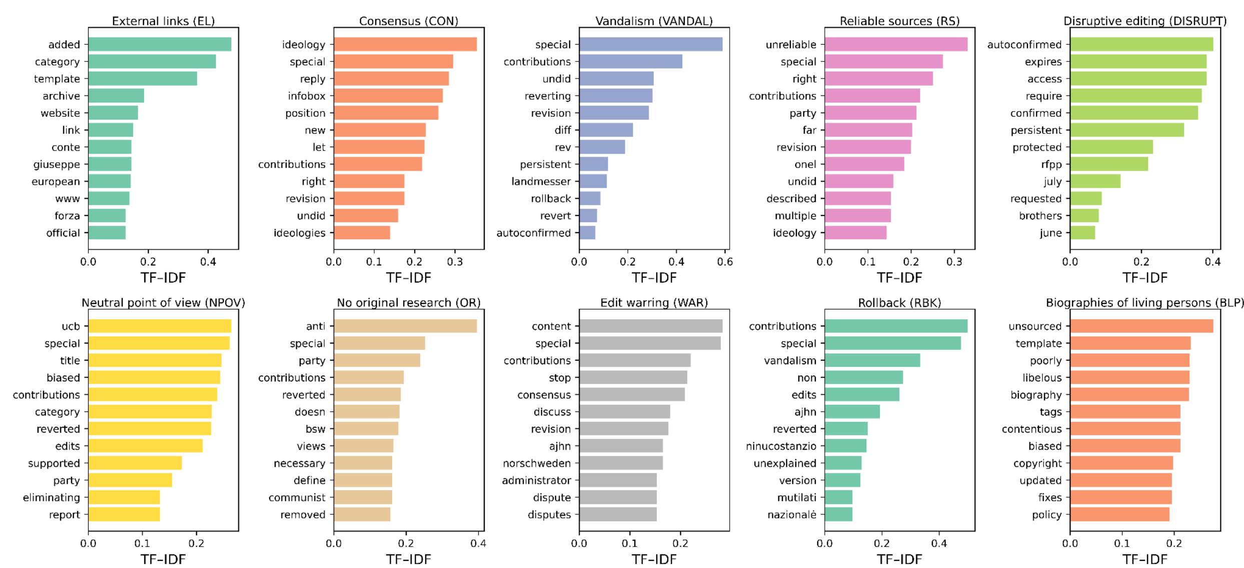

Figure 32: Policy citation context for the English Wikipedia. The 12 most frequent words co-occurring with each policy in the edit summaries or comments where they are mentioned, for the 10 most cited policies, with TF-IDF normalization.

Policy deleted words

We now perform a similar analysis where, instead of looking at the text accompanying the citation of a policy in the edit summary of comment, we consider the words deleted in a revision where the policy is mentioned in the edit summary. In this way, we can see which policies are invoked the most to delete content (Figure 33) and which are the most frequent words deleted invoking a certain policy (Figure 34).

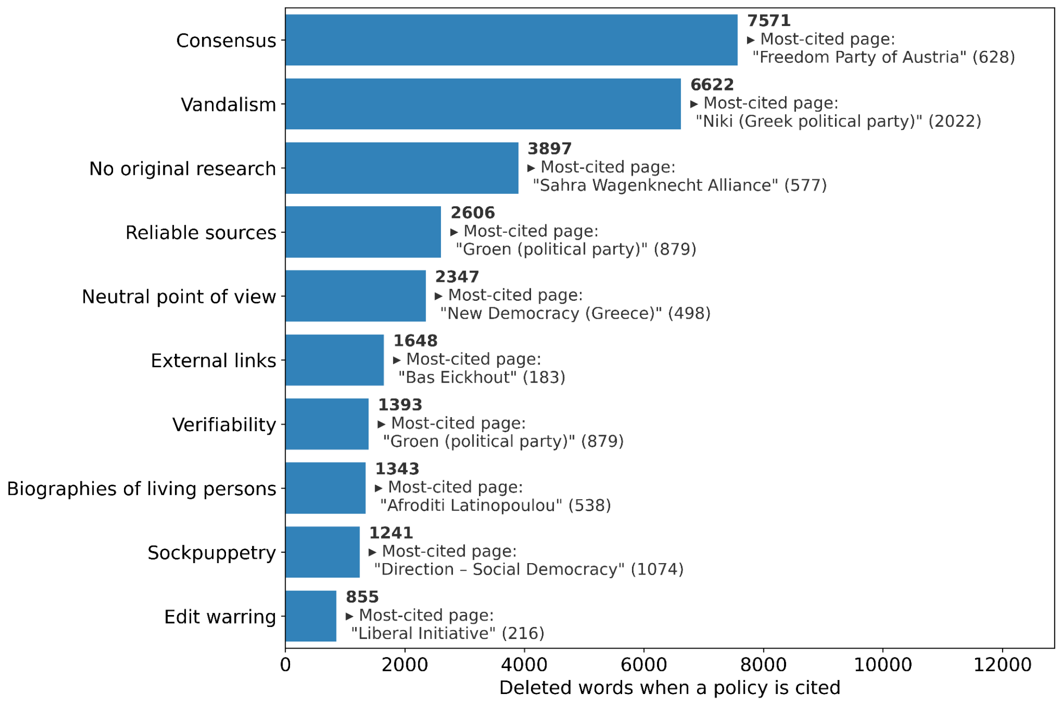

Figure 33: **Top 10 policies with the highest amount of deleted words in policy-citation edits in the English Wikipedia.** Histogram of deleted words by policy-citation edits corresponding to each policy during the electoral period. In bold is shown the total number of deleted words and the page with the highest amount of deleted words of that particular policy.

At the top in Figure 33 we find again Consensus and Vandalism, followed by No original research. This is in line with the nature of this policy, leading to delete all the content that may be considered original instead of encyclopedic, and is not supported by reliable sources.

Looking at the words more frequently deleted (Figure 34) we find politically loaded words like “right”, “left” and “wing” especially for Consensus, Vandalism, Edit warring, policies that have to do with conflicts among users over the content of the articles. We find words like “conspiracy” and “propaganda” related to the invocation of the policy about Sockpuppetry.

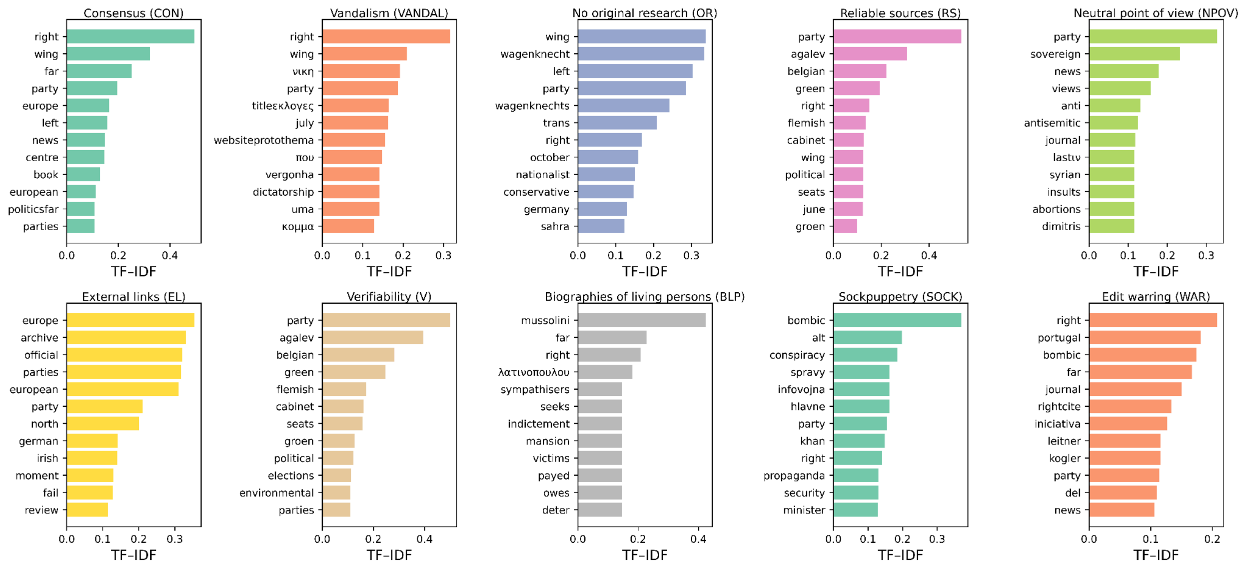

Figure 34: Deleted words in policy-citation edits for the English Wikipedia. The 12 most frequent words deleted when a certain policy is invoked in the edit summaries or comments, for the 10 top policies, with TF-IDF normalization.

Policy mentions patterns

We now take a deep look at policy citation patterns. We start by investigating which kinds of users invoke policies in our dataset, and to what extent. In Figure 35, on the top, we observe the quite low presence of anonymous users, which could be expected as these kinds of users tend to have less engagement in the community and less knowledge of its norms; instead we see a certain presence of bots, which could also be expected as the automatic messages they generate to accompany their edits may typically refer to some policy, for example to Vandalism when reverting vandalism edits. In the same figure on the bottom, we can see the share of policy mentions by users of different levels: we observe a considerable importance of higher level users, decreasing as the level increases, as it could be expected, given the lower number of higher level users.

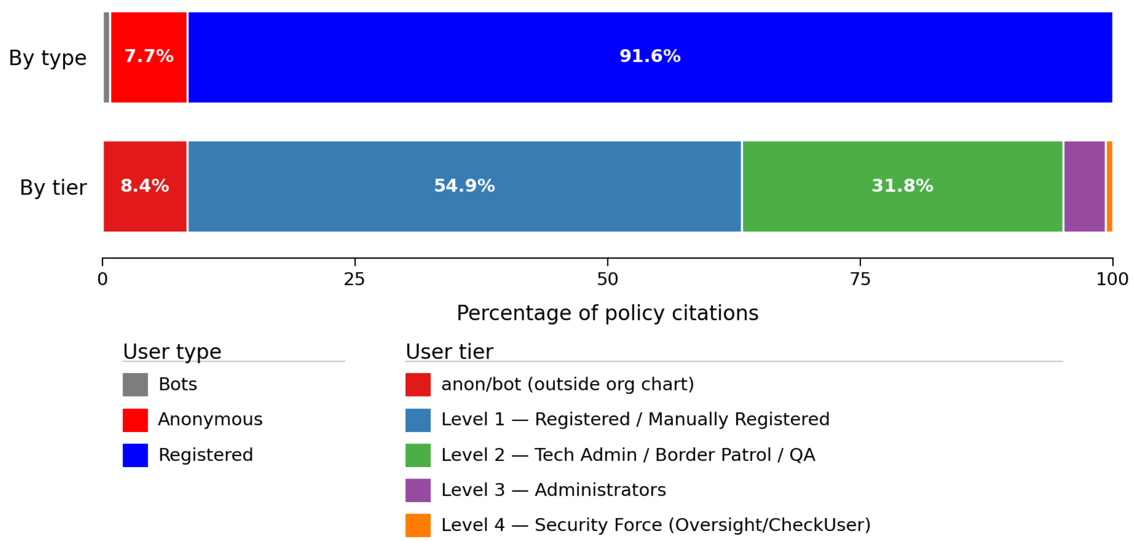

Figure 35. Proportion of policy citations by user type and tier. Stacked bar charts showing the proportion of edits invoking policies by user type (top) and by user tier (bottom) in the English Wikipedia.

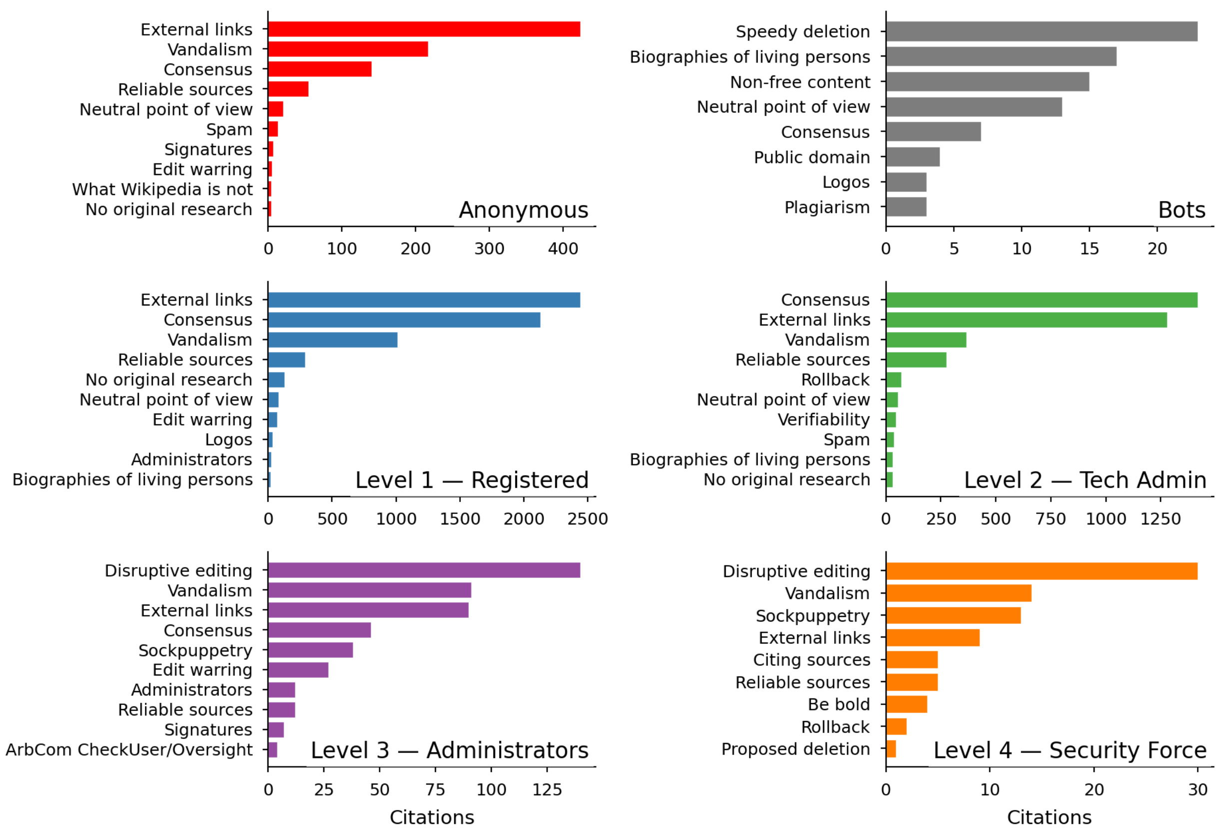

Figure 36. Citations ranking of the top policies cited by each user type/tier. Histogram of citations of the top policies cited by each type of user (Anonymous and Bots) and, in the case of Registered users, by the different user tiers during the electoral period.

Not only is the proportion of policy citations different across different kinds of users, also the policies they invoke more often are different. In Figure 36 we show the ranking of top policies invoked for the different kinds and levels of users. As expected, bots exhibit different patterns from all other types of users, since there are policies cited only by bots in this electoral context (Non-free content, Public domain and Plagiarism). Regarding anonymous users, we could say that they also participate in the Wikipedia bureaucracy, as they also invoke policies, although with lower proportion than registered users, as it could be expected. The policies they cite more often tend to coincide with those for level 1 and level 2 users; in fact, the histogram is quite similar for these three categories of users. Instead, we observe considerable differences for levels 3 and 4 users, who often invoke other policies related to user misconduct and illegitimate behaviour, such as Disruptive editing or Sockpuppetry. This is in line with the special rights and duties of these higher level editors, who are the only ones having the permissions to perform actions on other users' accounts in order to detect and prevent abusive behaviour.

In Figure 37 we inspect which kind of pages attract more policy citations in the English Wikipedia, according to country and EU Party, each combined with gender. We observe a higher proportion of policy citations for Italy, and for Germany, France, Greece.

EPP, NI, PfE are the parties that have the highest proportion of policy citations; for NI, PfE, RE we observe higher values for men, while only for GUE we observe a marked prevalence of the use of policy citations on women's pages.

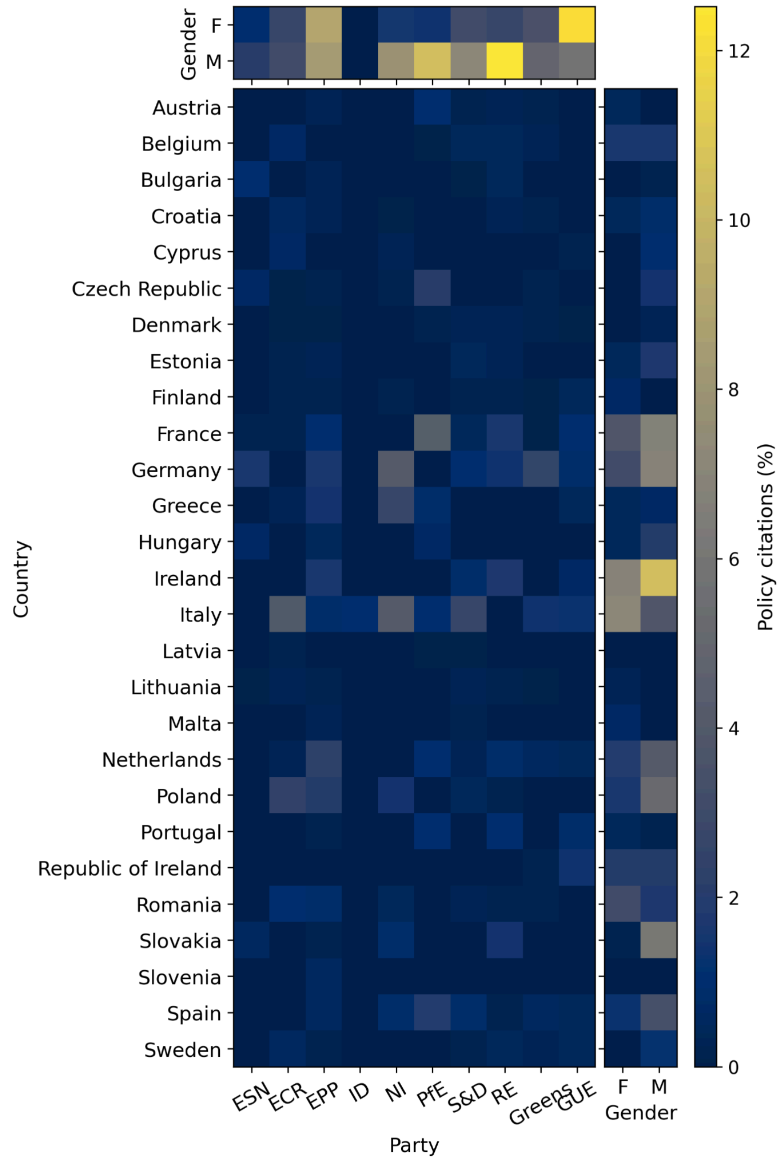

Figure 37. **Proportion of policy-citation edits across the different content.** Percentage of policy invoking edits per country, European party and gender for the **English** Wikipedia during the electoral period.

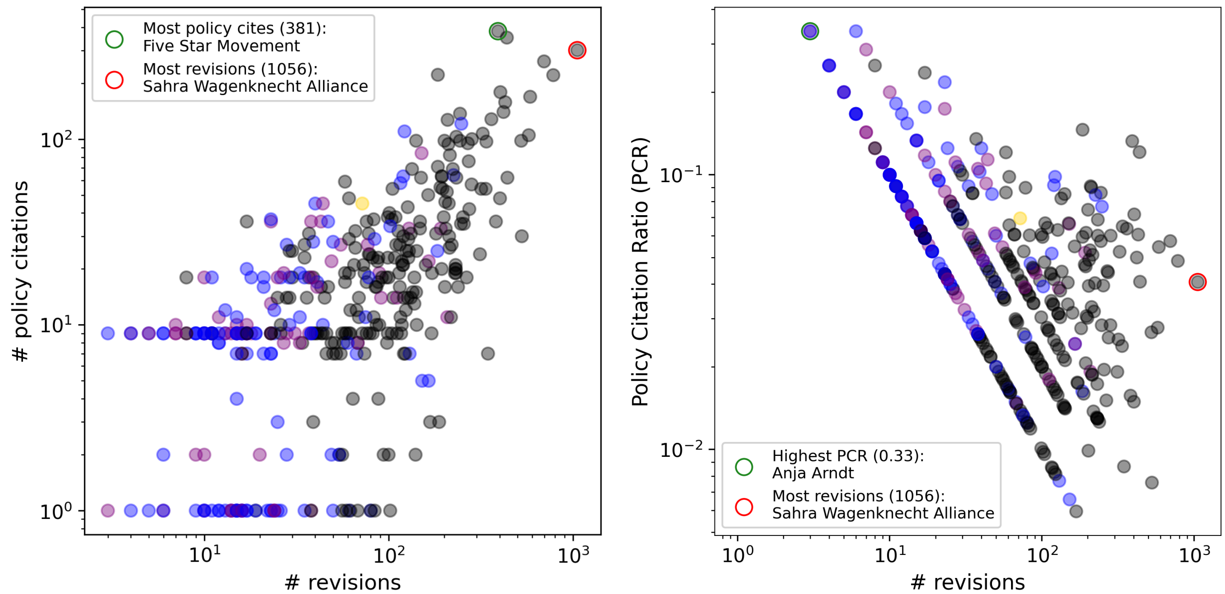

Figure 38. Policy citations and ratio with activity. Scatter plot for the pages in English Wikipedia of the number of policy citations (left) and the Policy citation Ratio (right) with respect to the editorial activity of the page. In each plot, each point represents a Wikipedia page, coloured according to gender: male in blue, female in purple, non-binary in yellow and agender in black. The pages with highest Y-variable and X-variable are highlighted in red and green, respectively.

We now examine deeply policy citations across pages. In Figure 38 we see on the left the policy citations in function of the total revisions of a page, which show a considerable alignment between the two variables; interestingly, no page overcomes 100 revisions without having at least some policy citation, and the pages getting to hundreds or thousands revisions have also the highest numbers of policy citations. This suggests that invoking policy is something intrinsically connected to editorial activity especially above a certain scale.

Looking at the proportion of policy citations in function of the number of revisions (on the right), we do not see a clear pattern, as the proportion is spread without a clear dependence on the volume of activity, but for a higher value of PCR (Policy citation Ratio) for pages having few revisions, and at least some policy citation.

Policy network and clusters

Now instead of considering policies individually and separately, we consider their relationship with each other, based on their lexical similarity, and the resulting network and clusters of policies more closely related to each other.

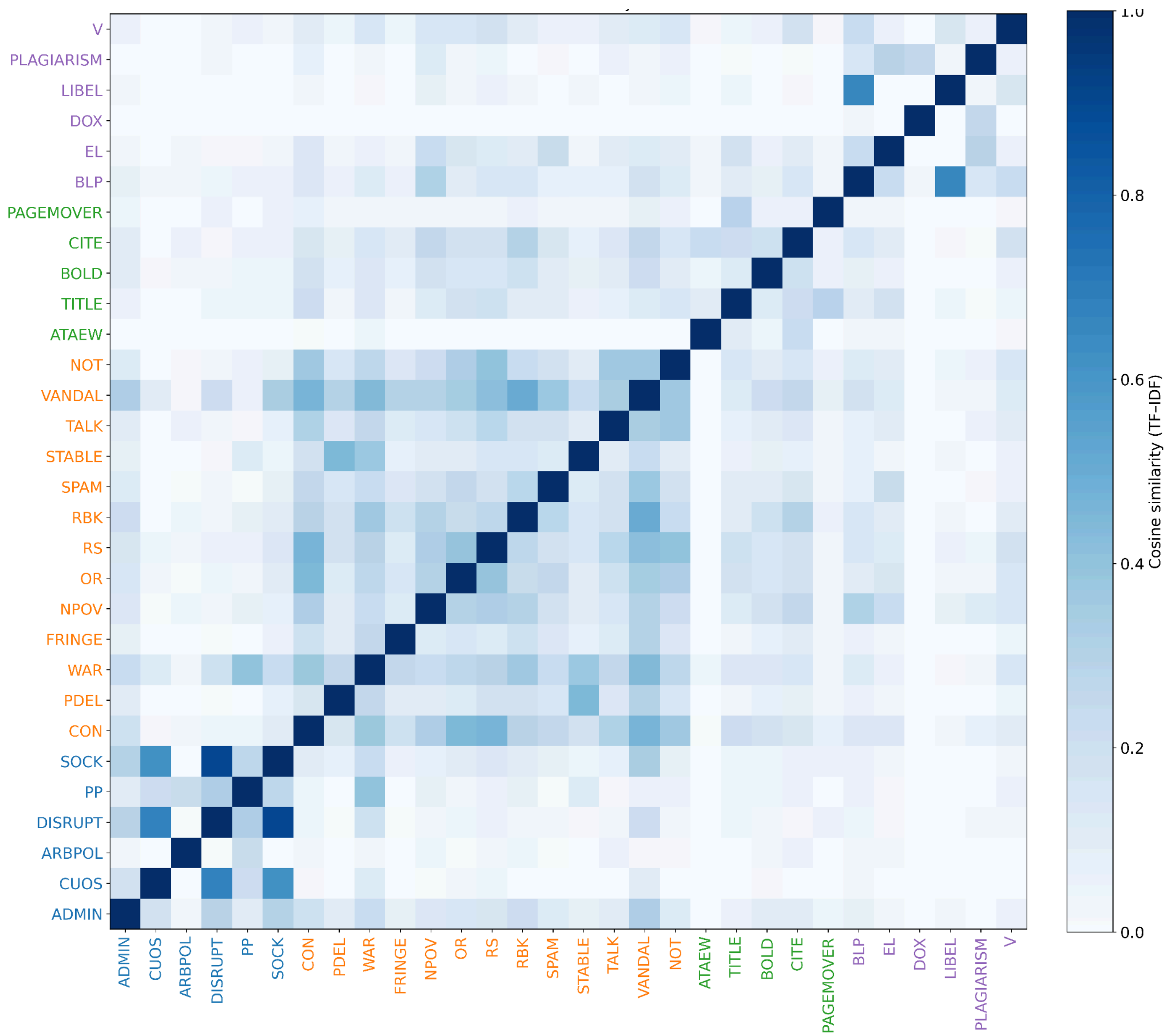

Figure 39. **Lexical similarity relation between the different policies in English Wikipedia.** Lexical similarity matrix between policies, computed as the cosine similarity between the lexical context in which they are mentioned. Colours of the policy names represent the clusters identified through the Louvain method for community detection.

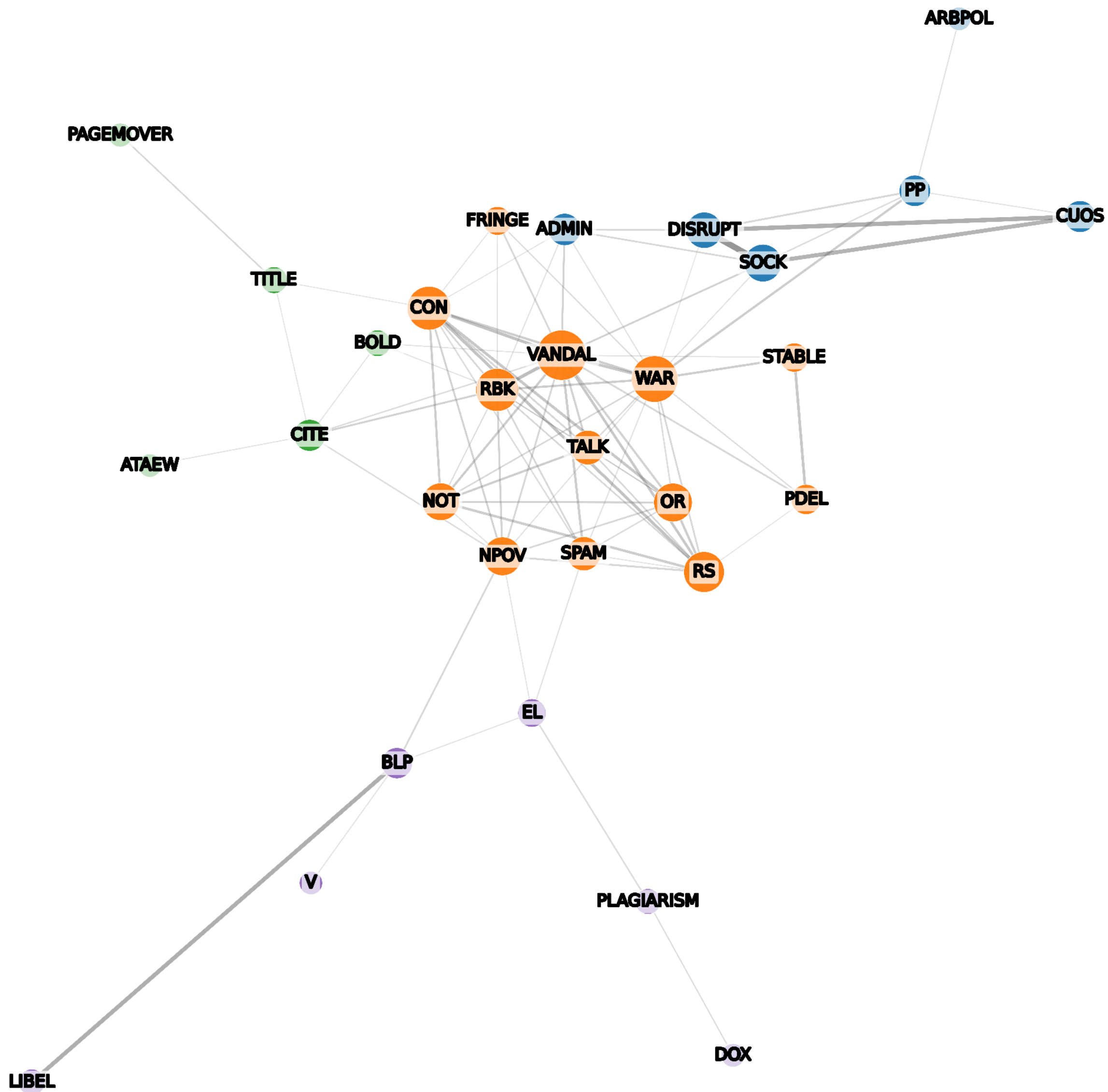

Figure 40. Network of lexical similarity of English Wikipedia policies in the electoral context. Representation of the network, where policies are nodes, connected via edges showing the lexical similarity. The width of the link between two policies is proportional to the lexical similarity between them. Node size represents the strength of the node, i.e. how similar it is to other policies in the network, and node colours represent the communities identified through the Louvain method for community detection.

Figure 39 shows a matrix representation of the similarity between each pair of policies, while Figure 40 shows the resulting network, where policies that are closer to each other are connected by an edge, and the layout presents the network topology displaying nodes so that their distance in the spatial representation tends to mirror their distance in the network.

Colours are consistent across the figures and represent the clusters identified by the Louvain algorithm for community detection. We see a major central cluster, depicted in orange (Cluster 1), that encompasses the Wikipedia pillars NOT (What Wikipedia is not) and NPOV (Neutral Point of View), and other main content policies of Wikipedia such as OR (No Original research) RS (Reliable Sources).

We then find Cluster 0, in blue, with policies related to abusive editing practices, such as disrupting editing or sockpuppetry, and admin's response actions; this cluster is connected to the previous one through nodes like WAR (Edit warring) and Vandalism.

Cluster 2, depicted in green, includes policies mostly related to different aspects of article content editing such as article titles or citations; interestingly, also BOLD ("Be bold" policy) is included, a policy that is an invitation to adopt an attitude of boldness when editing Wikipedia, and includes its opposite CAREFUL (Be careful) and other subsections related to specific aspects such as GLC (Graphical layout changes).

Finally Cluster 3, depicted in purple, includes mostly legal issues such as Plagiarism, Libel, DOX (Harassment). Interestingly, also V (Verifiability), and BLP (Biographies of living people) are included in this cluster, which points to the delicate nature of this kind of articles, that may incur in legal issues when publishing inaccurate information on a person. This cluster is connected to the main cluster through NPOV and SPAM.

In Figure 41 we show the words most frequently associated with the policies of each cluster. Cluster 0 has to do mostly with terms related to user accounts and permissions, and administrative actions. Cluster 1 with the dynamics of editing, and interesting with words such as "ideology", "left", "right", "political"; it seems that associated with these policies we find more political debate. In Cluster 2 we find words associated with editing and reverting; in Cluster 3 we find typical Wikipedia technical words related to templates and categories, but also some names of politicians and political parties.

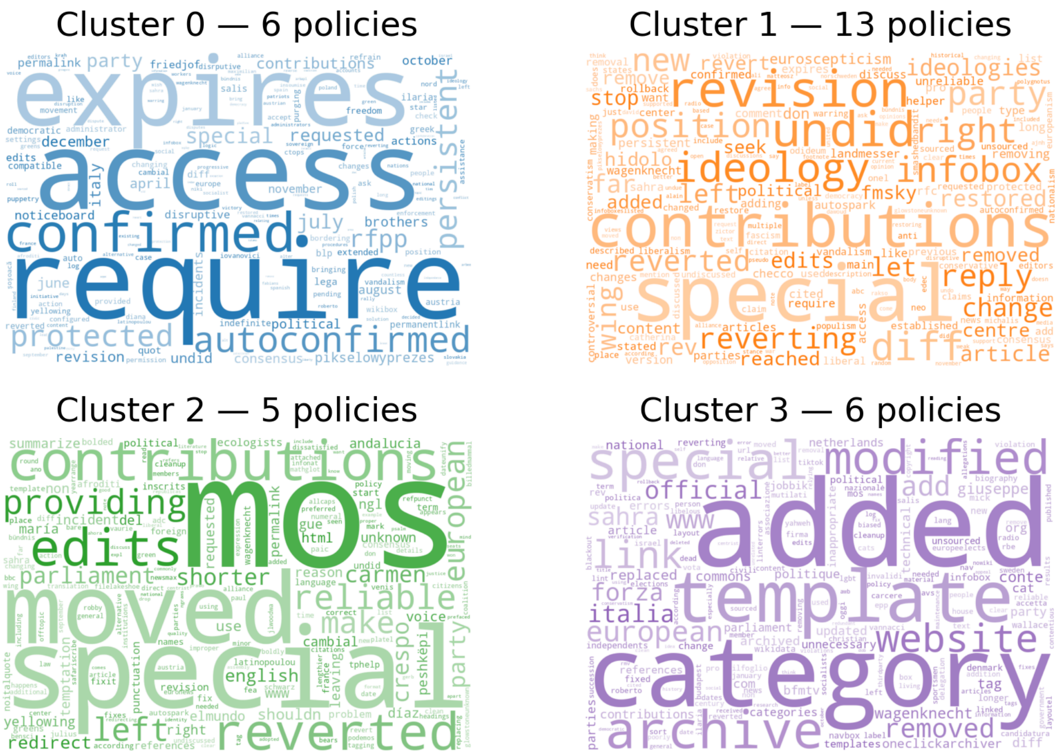

Figure 41. Clouds of words for each policy's communities in English Wikipedia. Word cloud representation of the lexical context of each community of policies. Word size is proportional to the sum of the frequency of each word co-occurring with the policies from a given cluster. The colours represent the clusters identified through the Louvain method for community detection.

Moderation activity

So far, we have focused our analysis on all policies extracted with our selection that includes behavioural and content policies and guidelines. To focus specifically on the policies more strictly related to content moderation in the context of elections and the DSA, we created a subset of 16 "moderation" policies, manually selected by UvA as explained above. We refer to citations of this group of policies as "moderation activity", that represents the 12% of the total policy citation activity; in fact, the most cited policies such as "Vandalism" or "Consensus" are not included in this restricted list of moderation policies.

As explained above, we further grouped these moderation policies into 5 categories (Figure 28). This categorization allows us to discriminate moderation practices taking place with 5 major objectives: *source quality, content neutrality, user conduct, account control* and *content protection*. Instead of analysing individual policies, in the following analysis we present results based on this categorization to characterise moderation patterns taking place during the electoral period.

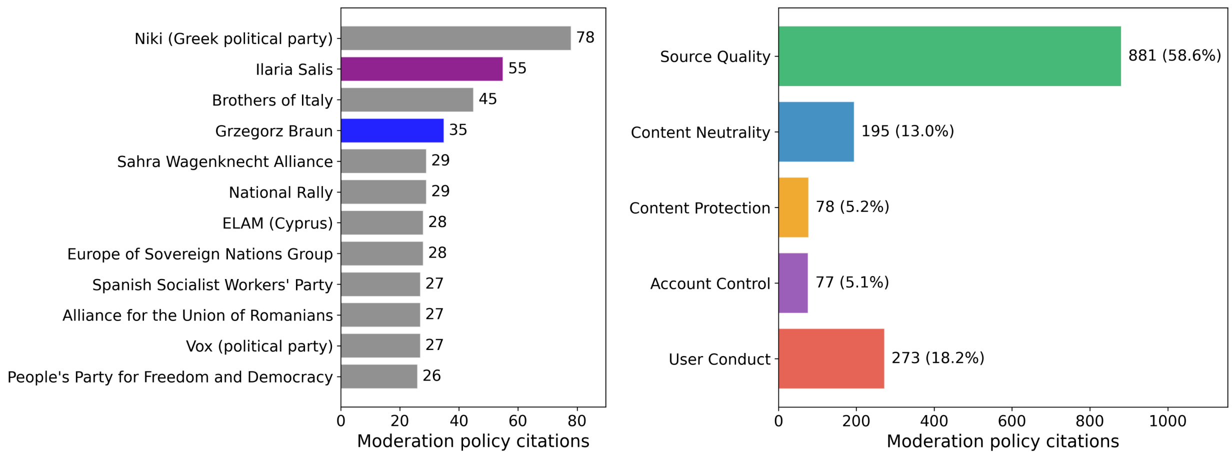

Figure 42. Moderation policy citations by top pages and by policy categories. *Left:* Top 12 articles receiving most moderation policy citations in English Wikipedia during the electoral period. Colour highlights the gender associated with the page: grey for agender pages, blue for male biography pages and purple for female biography pages. *Right:* Moderation policy citations percentage by category.

In Figure 42, on the left we can see the ranking of the articles receiving most moderation policy citations. The top 12 pages sum to 434 citations, i.e. ~29% of all moderation activity, showing a strong "hot-page" effect while still leaving a long tail across many other pages. The top pages shown are mostly entries devoted to political parties, therefore not associated with a gender and depicted in grey.

We observe a prevalence of nationalist and radical-right parties. A majority of the pages are widely described in sources as far-right or ultranationalist parties: National Rally ("Rassemblement national" French party), Brothers of Italy, Vox, ELAM, AUR, Niki, plus the Europe of Sovereign Nations (ESN) Group. The two biography pages are associated with high-profile legal controversy cases, also related, although in different ways, to the far-right ecosystem. Ilaria Salis is an Italian activist who was arrested and incarcerated after alleged assaults on neo-Nazis at a far-right event in Budapest, and then released from house arrest in 2024 thanks to the legal immunity gained as an elected MEP. Grzegorz Braun is a Polish far-right politician who was charged with seven criminal offences stemming from multiple incidents, and for whom the European parliament lifted immunity in 2025. In both cases, controversies entering the legal field seem to heavily trigger moderation policy citations, raising the level of moderation activity in a biographical article to the one of the most controversial party articles.

On the right in Figure 42 we see how moderation activity is distributed among the five categories we identified. Source Quality policies are prominent, representing alone almost 60% of moderation activity; they are followed by User Conduct and Content Neutrality policies. Moderation activity related to Account Control and Content Protection is minor, representing about 5% for each category.

Moderation policy citation activity represents 13.5% of all policy citation activity in article pages, and only 7.3% in talk pages. The distribution across the different moderation categories is not the same for article and talk pages: as we can observe in Figure 43, the prominence of Source Quality is more marked in talk pages. Policies about User Conduct and Account Control are invoked mainly in article pages and much less in talk pages; as we will observe below, these two categories are typically cited by higher level users having special administrative rights.

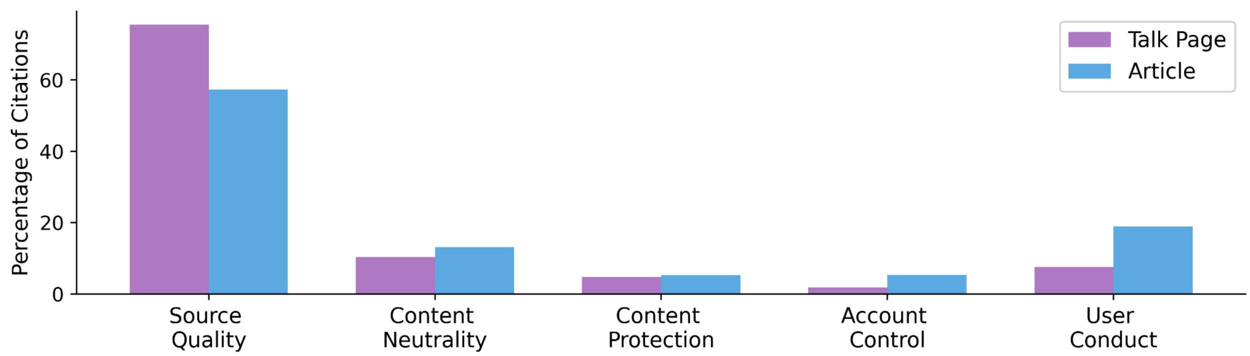

Figure 43. Moderation policy categories in talk pages vs article pages. Side-by-side bar chart comparing the percentage distribution of moderation policy citations across functional categories in talk pages versus article pages. Values are normalized separately within each page type.

Figure 44 examines how moderation policy invocations differ across pages categorized by gender, in other words, for pages about male politicians, female politicians, and agender pages (i.e. non-biographical pages devoted to entities such as political parties, events or institutions). For agender pages, the Source Quality category dominates the moderation profile, comprising the vast majority of moderation policy citations. This suggests that verification and sourcing concerns represent the primary moderation challenge across this type of page during the electoral period. Regarding biographical pages instead, the inset zoom allows for a closer inspection of the prevalent moderation categories for male and female pages specifically. Within this detailed view, we observe that Content Protection policies exhibit an elevated usage on male and female pages. This is indicative of the special treatment devoted to biographical articles of politicians, that involve legal implications and demand even stricter criteria than other articles, as described in deliverable D2.4; the policy "Biographies of living persons", belonging to this category, explains in detail the strict adherence to Wikipedia's core content policies necessary for this kind of articles that "require a high degree of sensitivity", as well as other issues such as the due regard for the subject's privacy, and the administrative actions like page protection and blocks.[9]

[9]See: https://en.wikipedia.org/wiki/Wikipedia:Biographies_of_living_persons#Role_of_administrators

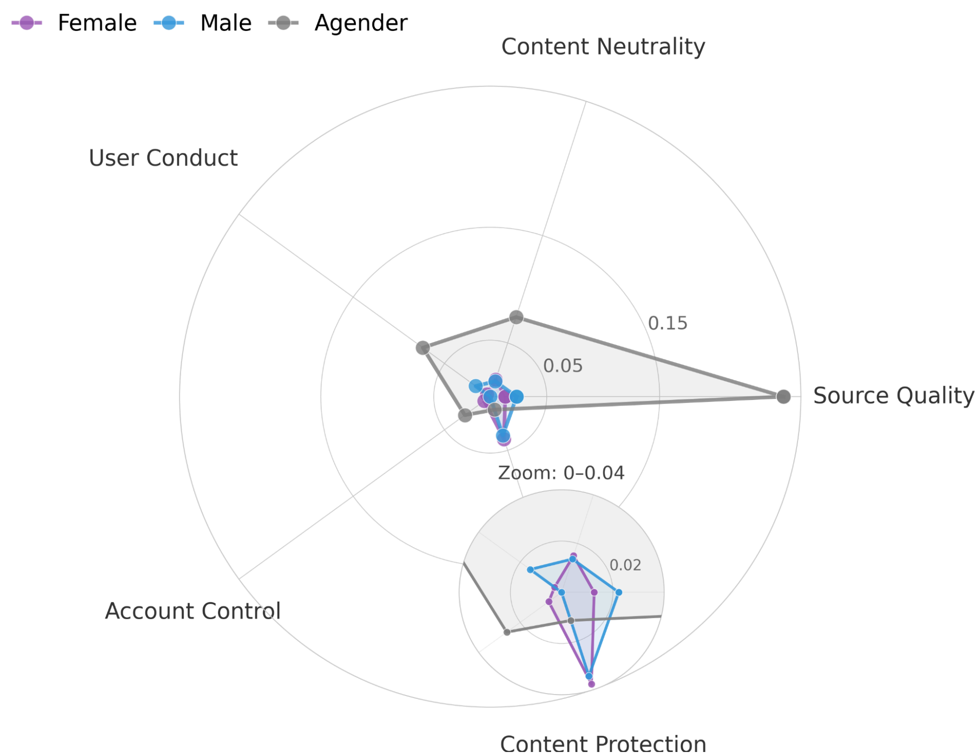

Figure 44. Moderation profiles by gender. Radar chart comparing the mean moderation profile (per-page category proportions) for pages grouped by the gender of the page subject (female, male, agender). The inset provides a zoomed view (0–0.04) to observe the male and female pages moderation profile.

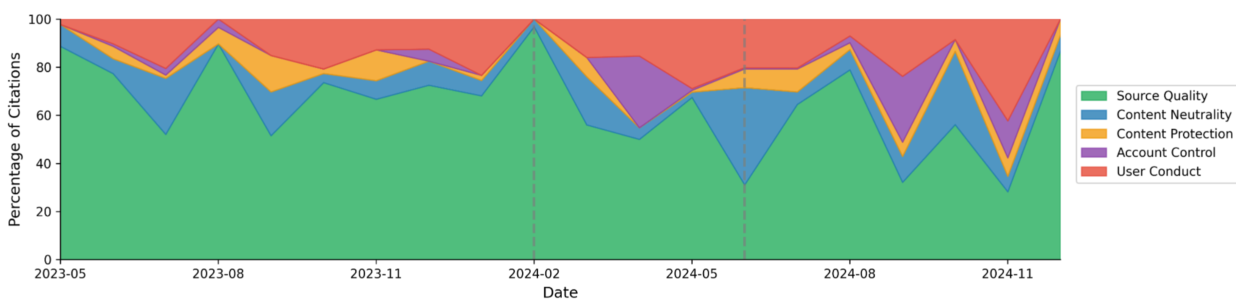

Figure 45. Temporal evolution of moderation policy usage by category. Stacked area chart showing the monthly share (percentage) of moderation policy citations attributed to each functional category over time. Dashed vertical lines indicate the start and end of the campaign period, and the last coincides with the election day.

Figure 45 presents the temporal dynamics of moderation policy usage throughout the electoral period, showing how the relative emphasis on different policy categories evolved from the election announcement through the campaign period to the formation of the European Commission. As expected, throughout the entire period, Source Quality policies (green) maintain a dominant presence, consistently accounting for 60-80% of all moderation citations in most months. This shows a general persistent emphasis of the community on sourcing and verification along all the period surrounding the elections, in line with the core principles of the project.

Content Neutrality (blue) exhibits a varying pattern, with elevated usage on specific moments, including the month of the election. These are likely corresponding to periods of intense political controversy or dispute over article framing. Account Control (purple) is characterised by low general presence with some marked peaks, a couple of months before and after the election, indicating moments of heightened vigilance against disruptive editing and potential sockpuppetry or coordinated manipulation attempts.

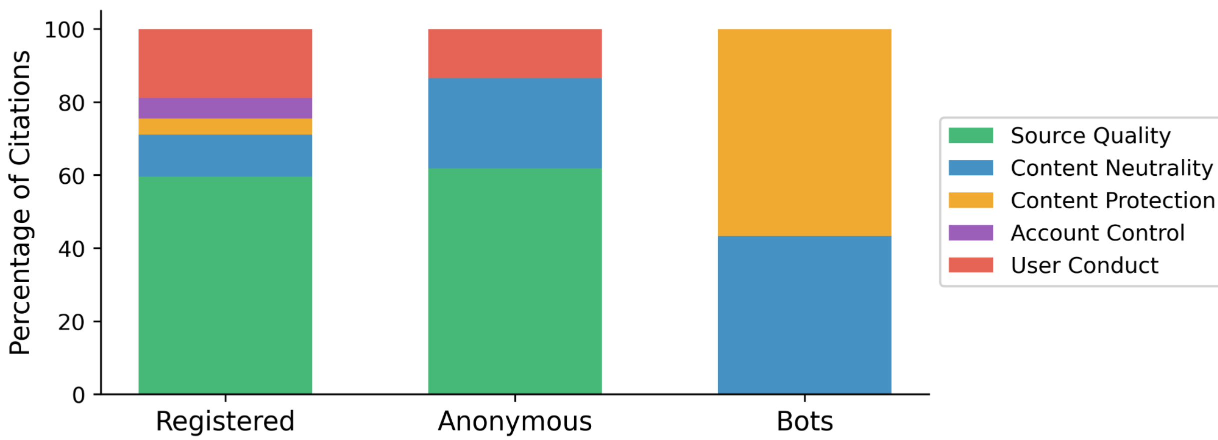

Figure 46. Moderation policy categories by user type. Stacked bar chart showing the percentage distribution of moderation policy citations across functional categories for registered users, anonymous users, and bots.

Figure 46 disaggregates moderation policy citations by user type, revealing how registered users, anonymous users and bots differ in their invocation patterns across the five moderation policy categories. Registered users (91.5% of moderation activity) demonstrate the most balanced moderation profile, invoking all five policy categories with Source Quality remaining dominant but also showing substantial use of the other policies.

Anonymous users (6.5% of moderation activity) exhibit a similar overall pattern to registered users, but without mentions of User Conduct and Account Control policies. On the other hand, Bots (2% of moderation activity) present a very different profile, with Content Protection policies dominating their citations at approximately 60% of the invocations, followed by substantial Content Neutrality activity. This pattern reflects the specialisation of bots on specific tasks, and their use of policy mentions within automated messages.

Within registered users, the moderation patterns are different depending on the user tier. Figure 47 presents a heatmap showing how different user tiers specialize in invoking particular moderation policy categories. We observe that bot users, as expected from the previous graphic, specialized in

Content neutrality and protection. Anonymous users have their main focus on Content neutrality and especially Source quality, and have a similar profile to the one exhibited by level 1 and 2 registered users, that present an even higher specialization in Source quality. Their lower engagement with Account Control and User Conduct policies reflects limited experience with administrative aspects of governance. Therefore, unregistered and registered users with a low tier that use moderation policies present a similar moderation pattern, focussing their activity into source reliability and content management.

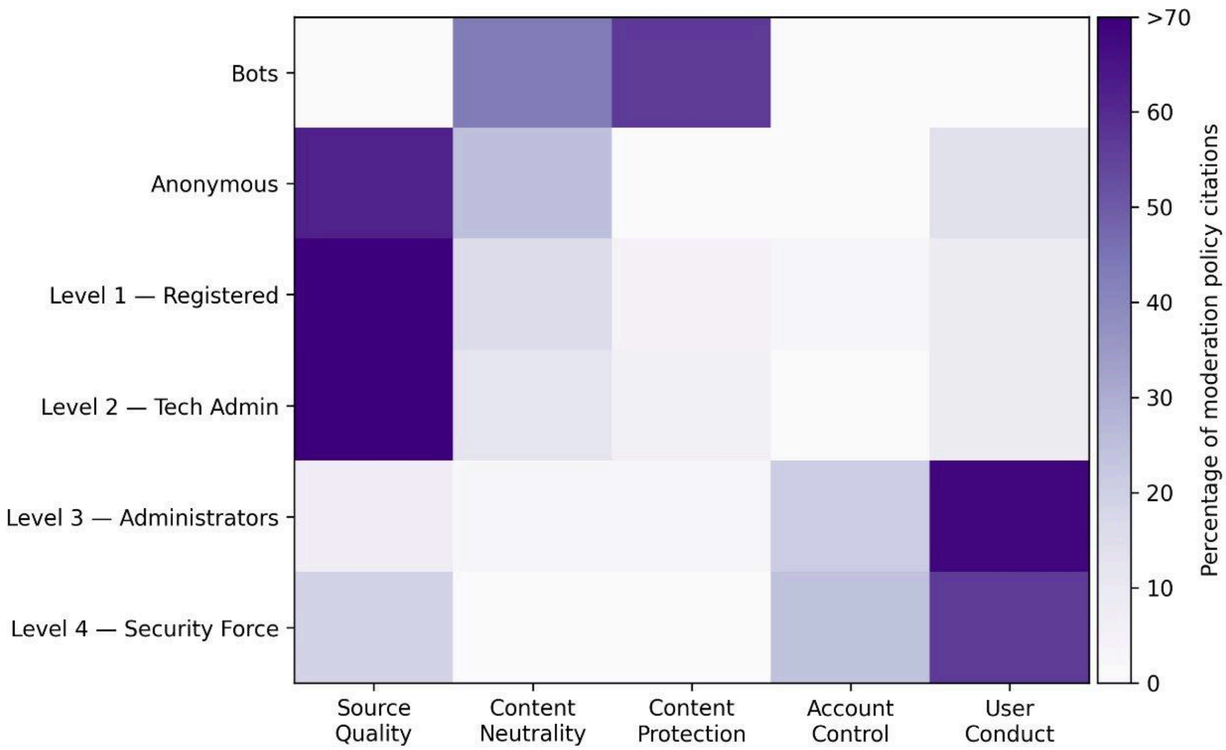

Figure 47. Moderation policy category specialization by user tier. Heatmap showing the percentage distribution of moderation policy citations across functional categories for each user tier. Values are normalized within each tier, and the colour scale is capped at 70%.

On the other hand, Level 3 and 4 users (Administrators and Security Force) show an extremely different pattern, specializing their moderation activity in User Conduct, followed by Account Control, and with barely no presence of content management (Content Neutrality and Content Protection). This specialization aligns with their mandates: investigating abusive behaviour like sockpuppetry, enforcing blocks, addressing severe behavioural violations requiring oversight or checkuser investigations.

The heatmap reveals a clear functional stratification in Wikipedia's moderation ecosystem during the 2024 European Parliament election. Lower-tier users concentrate on accessible, high-volume tasks like sourcing verification, basic conduct reporting, and addressing structural and neutrality concerns. Upper-tier users specialise in sensible administrative tasks involving account management and serious behavioural enforcement. This specialization enables efficient allocation of community attention, with different expertise levels addressing problems matching their capabilities and authority.

Results for other major languages (French, German, Spanish and Italian)

Here we report the results for the four major languages for which, beyond English, we found a significant amount of policy citation data: English, German, Spanish and Italian. The analyses are analogous to the ones performed for the English Wikipedia.

Most mentioned policies

For each of the four language editions considered, we first present two figures, where the first one indicates the list of the most mentioned policies, and the second reports, for each of the six most mentioned policies, the most frequent words co-occurring with it. This allows us to look for patterns, similarities and differences across language editions, and at the same time to inspect the textual content most frequently associated with the most frequent policies.

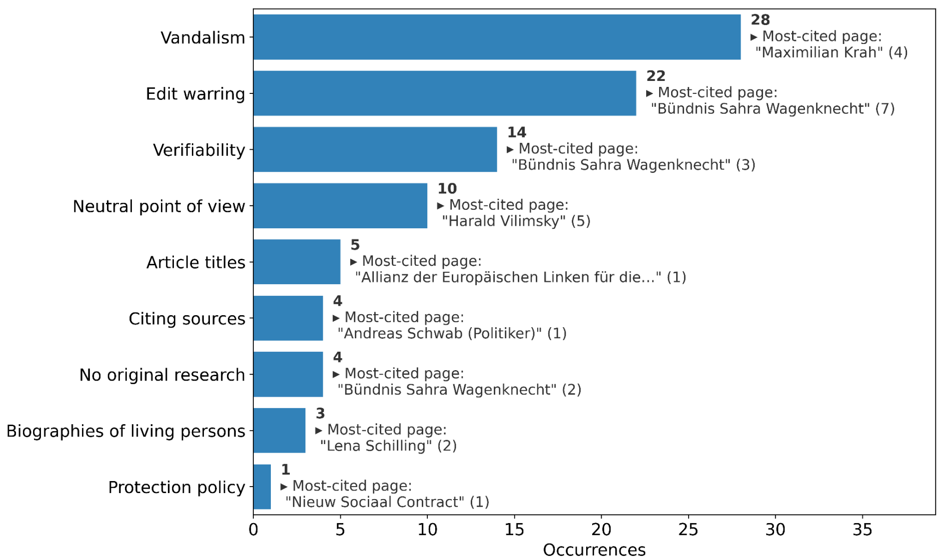

Figure 48: **Top 10 policies with the highest amount of deleted words in policy-citation edits in the German Wikipedia.** Histogram of deleted words by policy-citation edits corresponding to each policy during the electoral period. In bold is shown the total number of deleted words and the page with the highest number of deleted words of that particular policy.

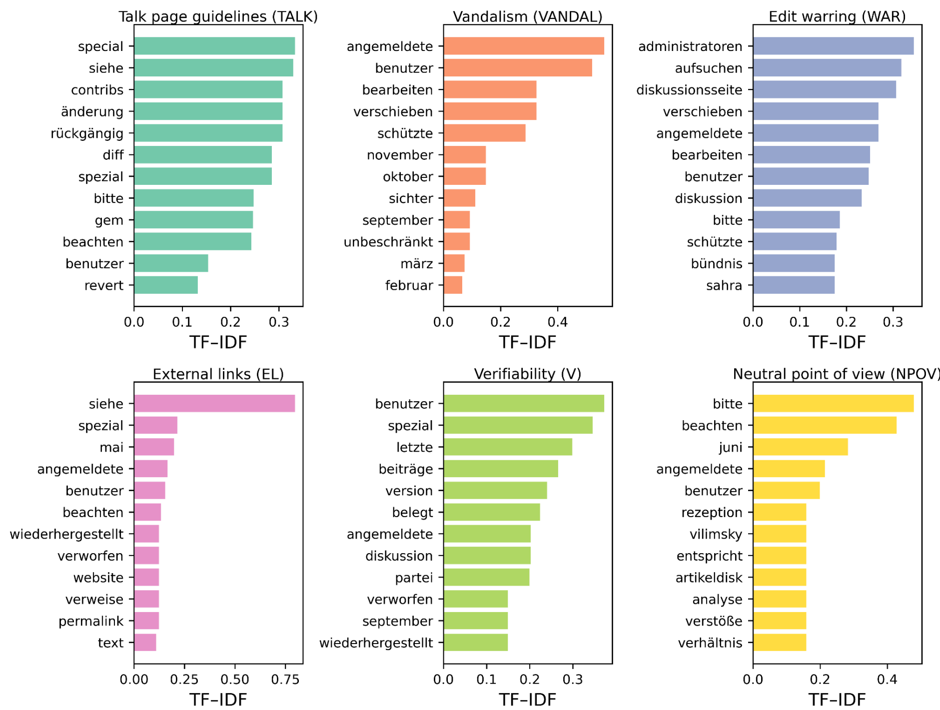

Figure 49: Policy citation context for the German Wikipedia. The 12 most frequent words co-occurring with each policy in the edit summaries or comments where they are mentioned, for the 10 most cited policies, with TF-IDF normalization.

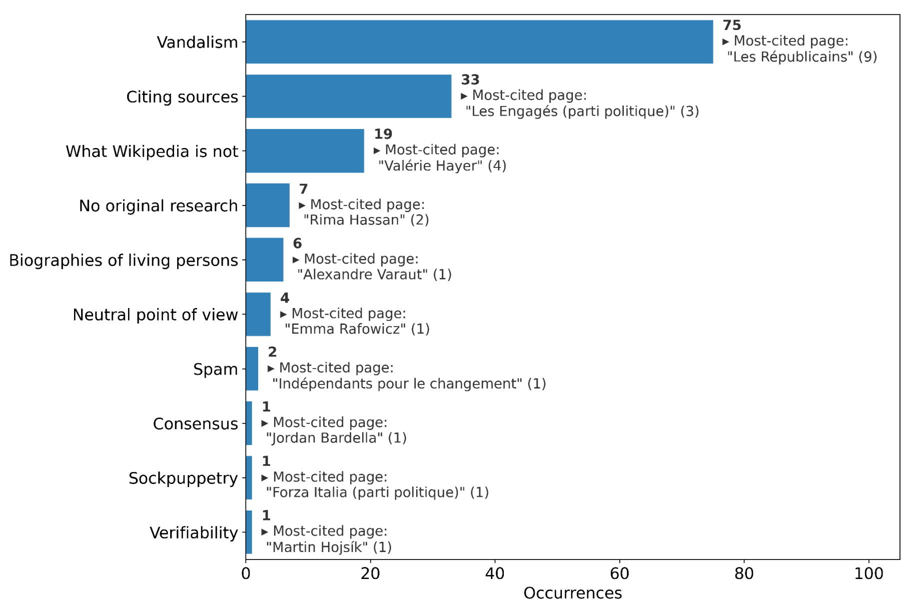

Figure 50: Top 10 mentioned policies in the French Wikipedia. Histogram of citations of each policy during the electoral period. In bold is shown the number of total occurrences, the number of occurrences for the page with the highest number of policy citations of that particular policy.

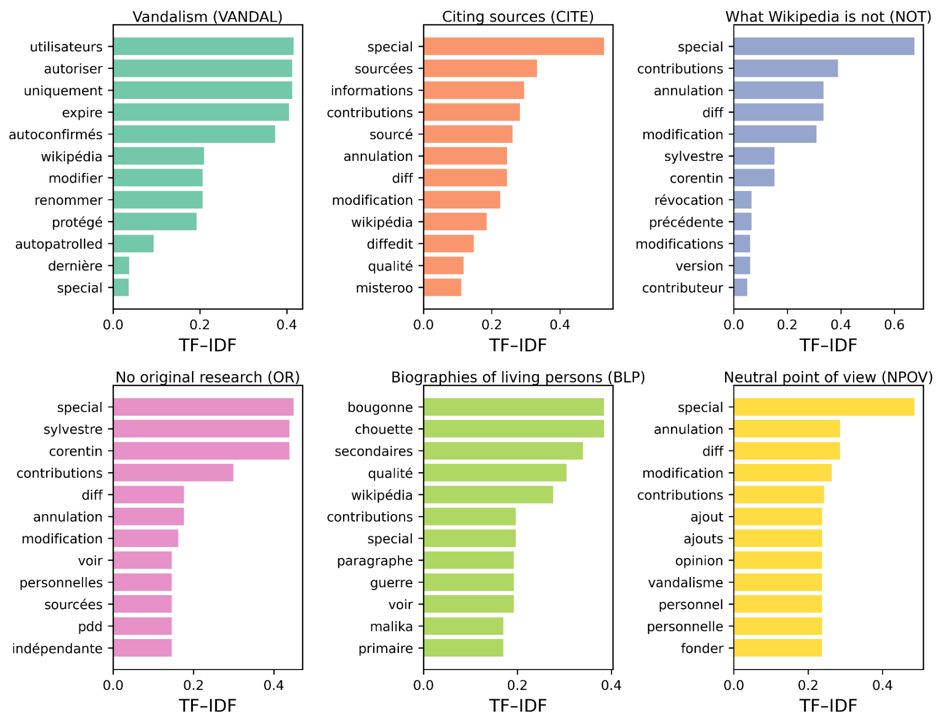

Figure 51: Policy citation context for the French Wikipedia. The 12 most frequent words co-occurring with each policy in the edit summaries or comments where they are mentioned, for the 10 most cited policies, with TF-IDF normalization.

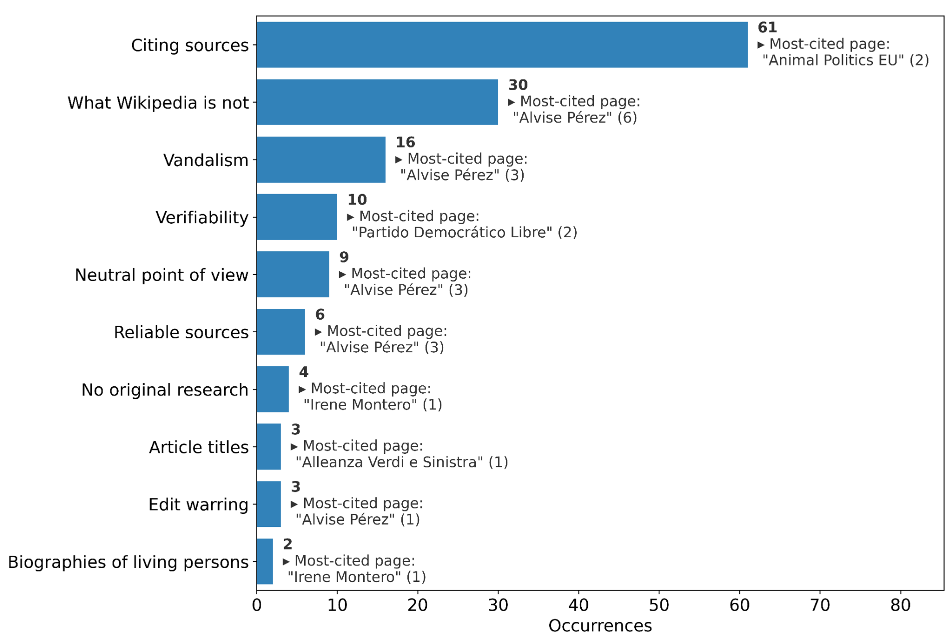

Figure 52: Top 10 mentioned policies in the Spanish Wikipedia. Histogram of citations of each policy during the electoral period. In bold is shown the number of total occurrences, the number of occurrences for the page with the highest number of policy citations of that particular policy.

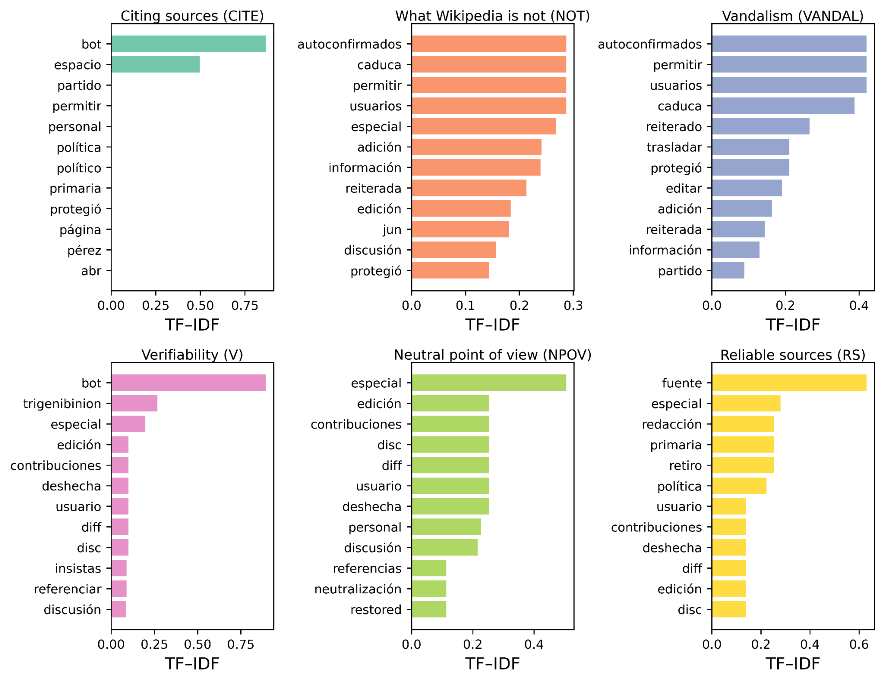

Figure 53: Policy citation context for the Spanish Wikipedia. The 12 most frequent words co-occurring with each policy in the edit summaries or comments where they are mentioned, for the 10 most cited policies, with TF-IDF normalization.

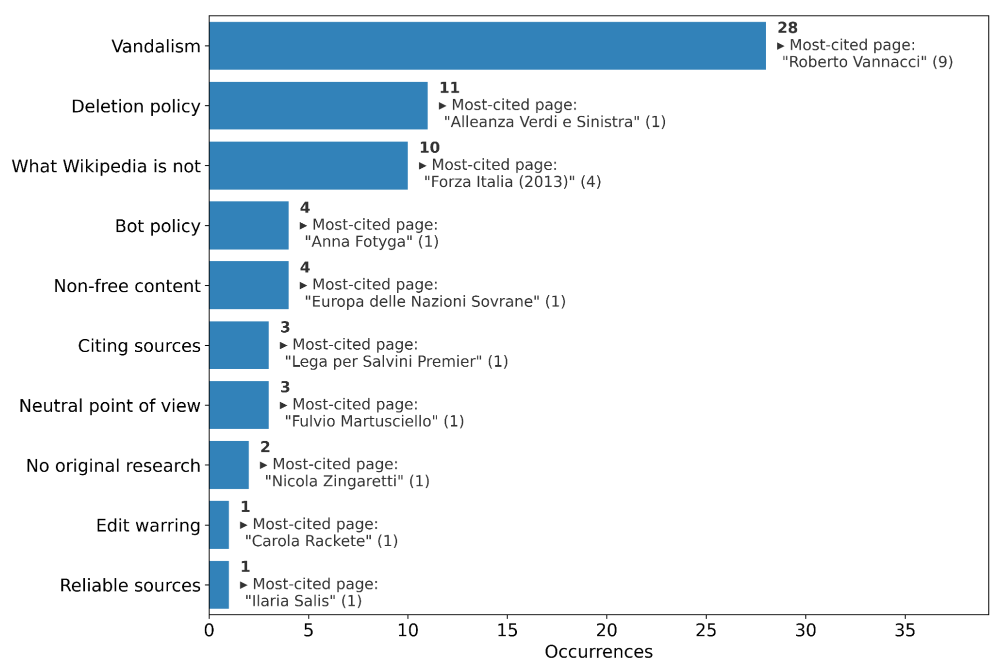

Figure 54: Top 10 mentioned policies in the Italian Wikipedia. Histogram of citations of each policy during the electoral period. In bold is shown the number of total occurrences, the number of occurrences for the page with the highest number of policy citations of that particular policy.

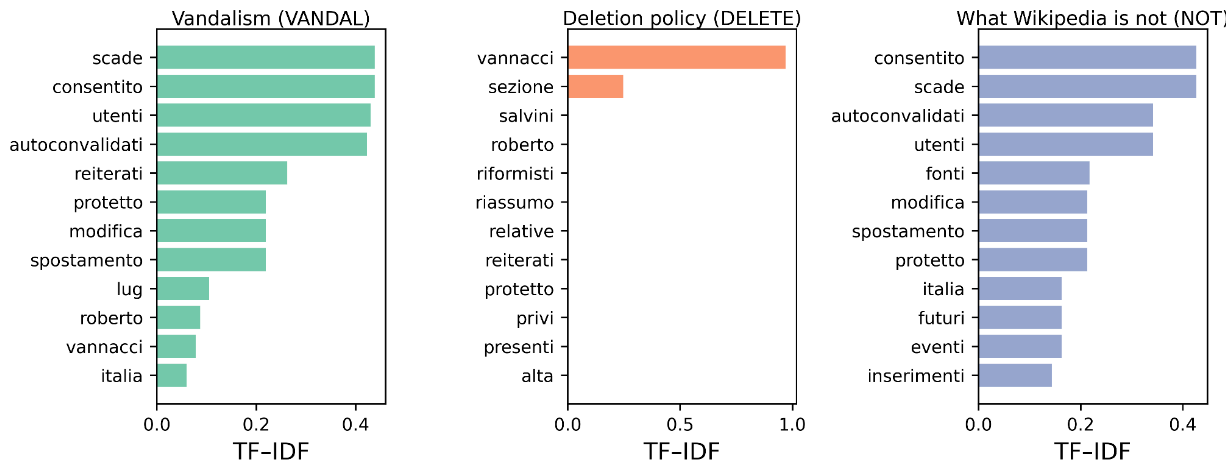

Figure 55: Policy citation context for the Italian Wikipedia. The 12 most frequent words co-occurring with each policy in the edit summaries or comments where they are mentioned, for the 10 most cited policies, with TF-IDF normalization.

We observe that Vandalism is the most mentioned policy in all language editions but for Spanish, where it is the third one; apart from this common aspect, the most mentioned policies exhibit a certain variety across languages. In German Edit warring and Verifiability; in French and Spanish Citing sources, followed by “What Wikipedia is not” which is also among the most cited in Italian Wikipedia.

The context words accompanying each policy show both universal patterns and language-specific emphases. Across all editions, policies related to sourcing and verification show similar vocabulary about citations, references, and reliability, suggesting shared editorial practices for establishing content credibility. Behavioural policies similarly show common vocabulary about user conduct, warnings, and disruption. However, the specific political terms appearing in policy contexts vary by language, reflecting each edition's focus on its associated political sphere, French terms relate to French parties and politicians, German to German political actors, and so forth.

These comparative patterns suggest Wikipedia's policy framework provides a shared foundation enabling similar governance approaches across linguistic communities, while allowing sufficient flexibility for each community to emphasize particular aspects matching its specific challenges and cultural context. The multilingual analysis reveals both the universality of certain Wikipedia governance challenges and the diversity of how different communities navigate them.

Policy deleted words

As done for the English Wikipedia, we now look for each language at an analogous measure, but focused on the words deleted in the edits that mention a given policy.

Interestingly, here we do not find at the top “Vandalism”, which suggests that vandalism edits tend to add little content, while other edits adding more consistent amounts of text are reverted invoking policies like “Citing sources” in French, “Reliable sources” in Italian; “Neutral point of view” in German and “What Wikipedia is not” in Spanish and Italian. We see words referring to political orientations deleted when mentioning the Neutral point of view policy, like “gauche” (left), “radicale” (radical) and “populisme” (populism) in French, “izquierda” (left) and “comunismo” (communism) in Spanish.

The German Wikipedia's pattern of deleted words when invoking policies reveals the specific content disputes characterizing German electoral coverage. The prominence of political and ideological terms in deletions associated with "Neutral point of view" and "Edit warring" indicates intensive negotiation over political framing and characterization. German editors apparently engaged in substantial discussion about appropriate terminology for describing political positions, movements, and controversies.

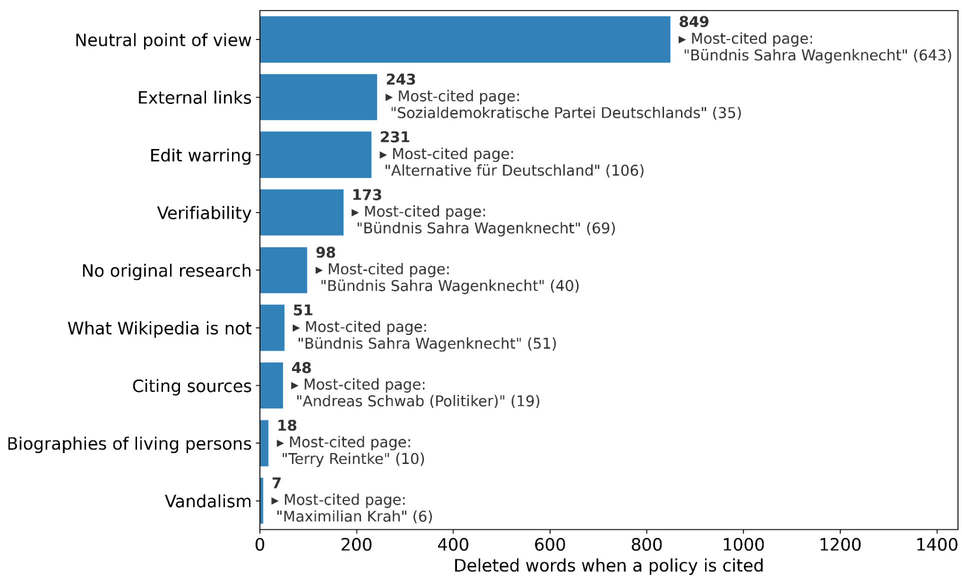

Figure 56: **Top 9 policies with the highest amount of deleted words in policy-citation edits in the German Wikipedia.** Histogram of deleted words by policy-citation edits corresponding to each policy during the electoral period. In bold is shown the total number of deleted words and the page with the highest amount of deleted words of that particular policy.

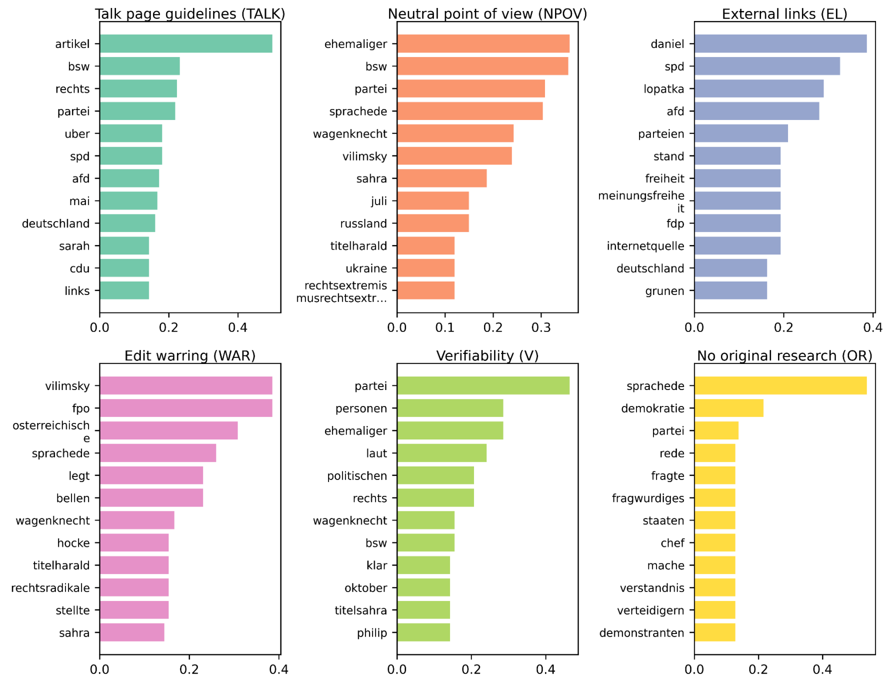

Figure 57: **Deleted words in policy-citation edits for the German Wikipedia.** The 12 most frequent words deleted when a certain policy is invoked in the edit summaries or comments, for the 6 top policies, with TF-IDF normalization.

The relatively high frequency of source-related terms in deletions associated with verification policies might suggest that the German Wikipedia actively removes poorly sourced political content, maintaining strict standards for electoral coverage. This pattern seems to align with the idea that the German Wikipedia places particular emphasis on explicit source citation and verification, creating a high-quality but potentially contentious editorial environment where sourcing disputes generate substantial discussion and content removal.

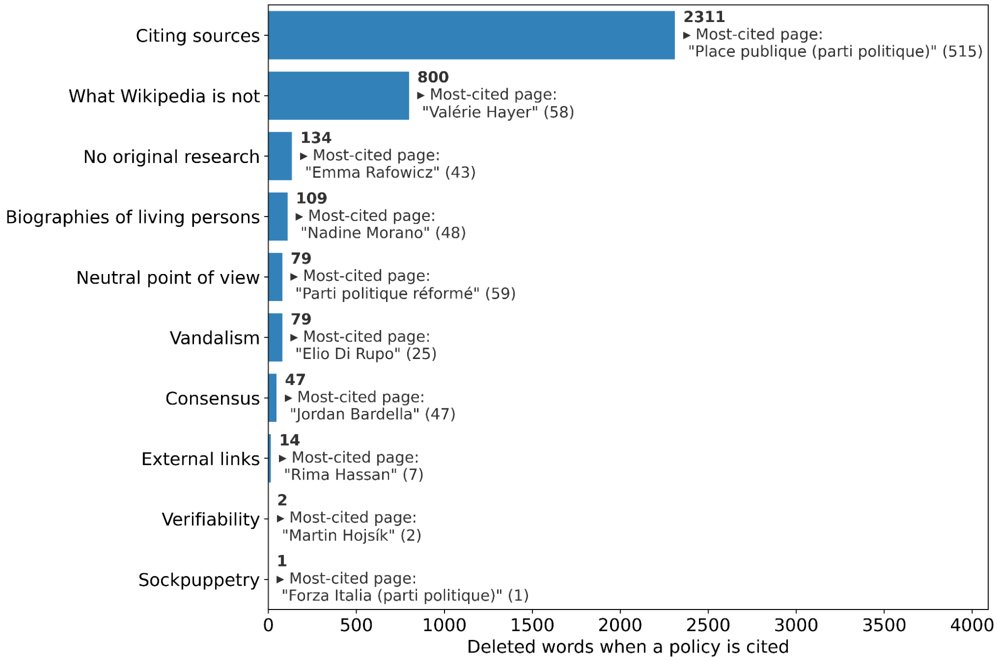

Figure 58: **Top 10 policies with the highest amount of deleted words in policy-citation edits in the French Wikipedia.** Histogram of deleted words by policy-citation edits corresponding to each policy during the electoral period. In bold is shown the total number of deleted words and the page with the highest amount of deleted words of that particular policy.

Looking at French Wikipedia's deleted word patterns, we observe the prominence of ideological vocabulary in deletions associated with the Neutral Point of View policy, which suggests that French editors frequently contested political characterizations and labels. The appearance of these terms suggests debates over whether certain descriptors constitute neutral encyclopedic language or politically charged framing requiring removal.

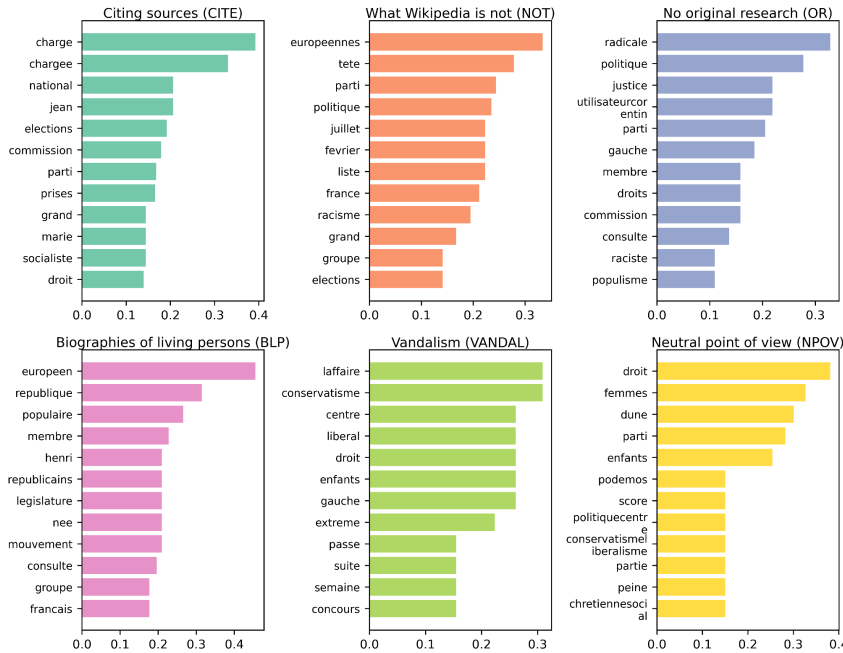

Figure 59: Deleted words in policy-citation edits for the French Wikipedia. The 12 most frequent words deleted when a certain policy is invoked in the edit summaries or comments, for the 6 top policies, with TF-IDF normalization.

The high volume of deletions associated with "Citing sources" might reflect intensive attention to source quality in French political coverage. Words deleted when invoking this policy likely represent claims that, while potentially relevant, lacked adequate sourcing according to Wikipedia's verification standards. The pattern suggests French editors strive to maintain rigorous sourcing requirements for political content, particularly during elections, when unverified claims and biased narratives circulate widely, as it occurred in German Wikipedia.

In the Spanish Wikipedia, the presence of ideological terms like "izquierda" (left) and "comunismo" (communism) in deletions associated with neutrality policies mirrors patterns observed in French and German editions, suggesting cross-linguistic consistency in how editors negotiate political characterizations while striving to maintain encyclopedic neutrality.

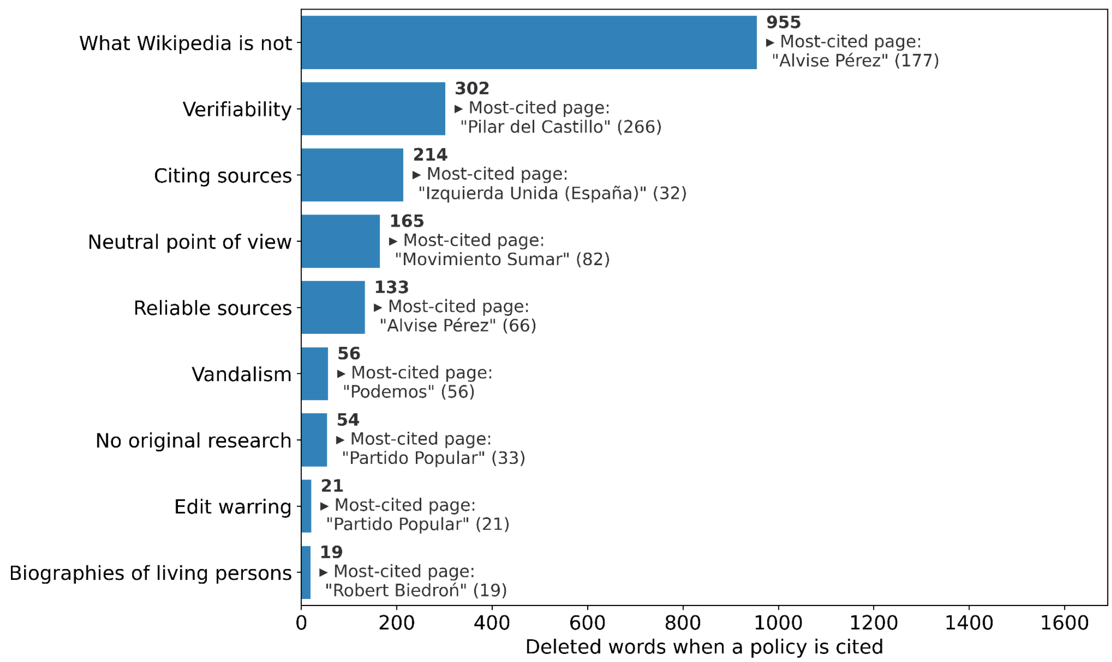

Figure 60: **Top 9 policies with the highest amount of deleted words in policy-citation edits in the Spanish Wikipedia.** Histogram of deleted words by policy-citation edits corresponding to each policy during the electoral period. In bold is shown the total number of deleted words and the page with the highest amount of deleted words of that particular policy.

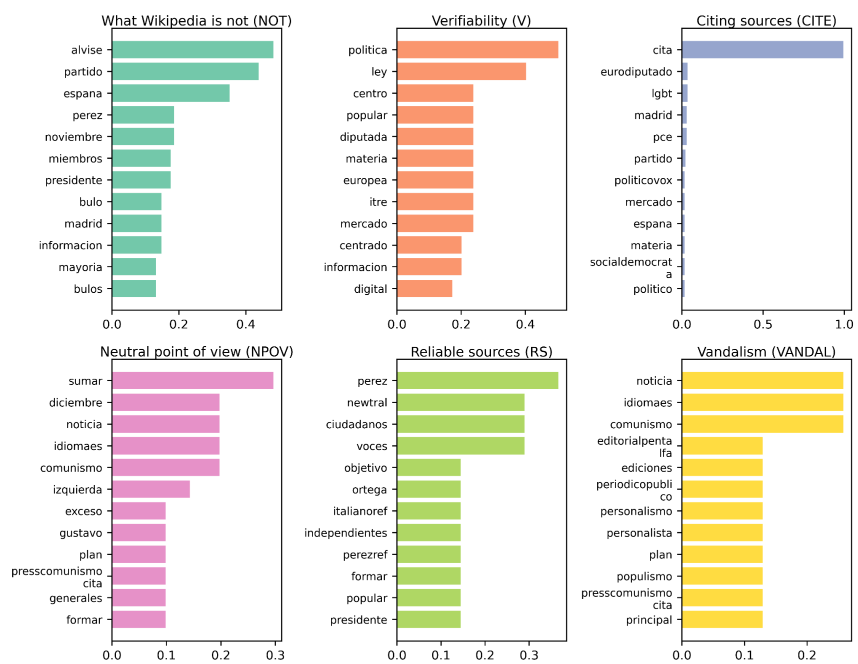

Figure 61: Deleted words in policy-citation edits for the Spanish Wikipedia. The 12 most frequent words deleted when a certain policy is invoked in the edit summaries or comments, for the 6 top policies, with TF-IDF normalization.

The frequent deletion of content invoking "What Wikipedia is not" is particularly notable in Spanish Wikipedia, appearing as the top policy for content deletion. This suggests Spanish editors frequently encountered content that, while potentially related to elections, fell outside Wikipedia's scope; perhaps including excessive opinion, advocacy, or non-encyclopedic detail. The appearance of words as "informacion" (information), “bulo”, “bulos” (fake new/s) as context words reflects the use of this policy to act against information considered not reliable, which suggests that the policy's prominence may be because of its use to distinguish between legitimate electoral coverage and inappropriate content during a politically charged period.

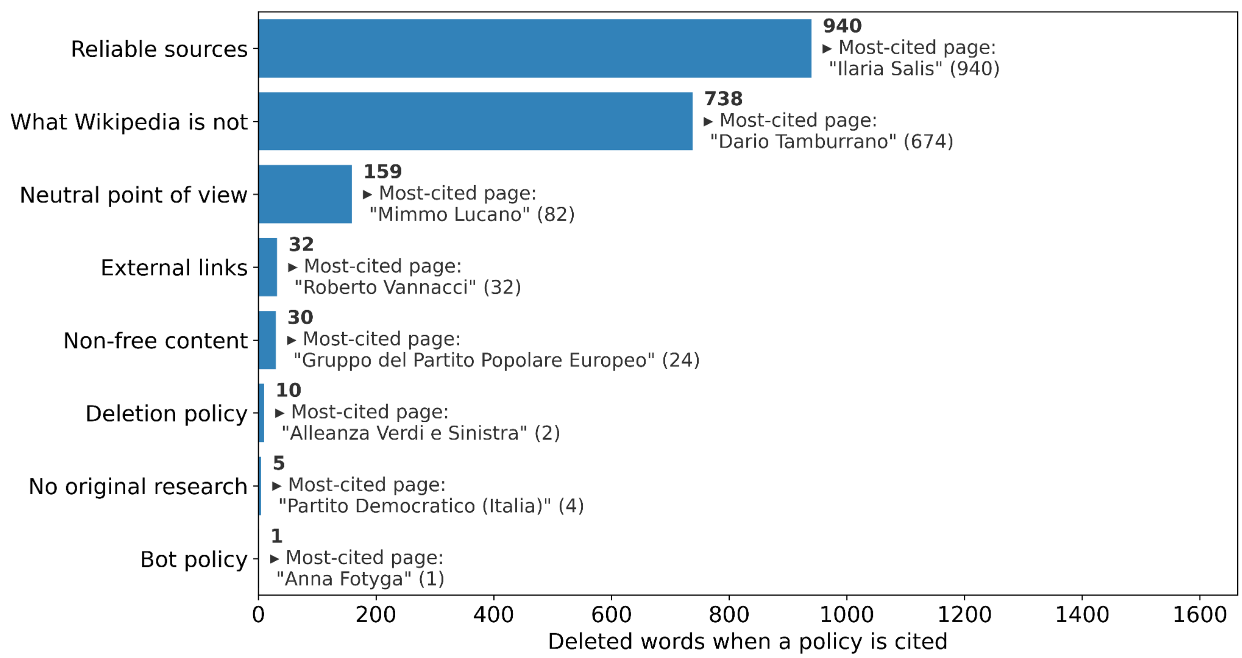

Figure 62: **Top 8 policies with the highest amount of deleted words in policy-citation edits in the Italian Wikipedia.** Histogram of deleted words by policy-citation edits corresponding to each policy during the electoral period. In bold is shown the total number of deleted words and the page with the highest amount of deleted words of that particular policy.

Italian Wikipedia's deleted word patterns exhibit the prominence of "Reliable sources" and "What Wikipedia is not" as top deletion policies, which aligns with patterns in other editions, suggesting Italian editors similarly prioritize source quality and appropriate scope.

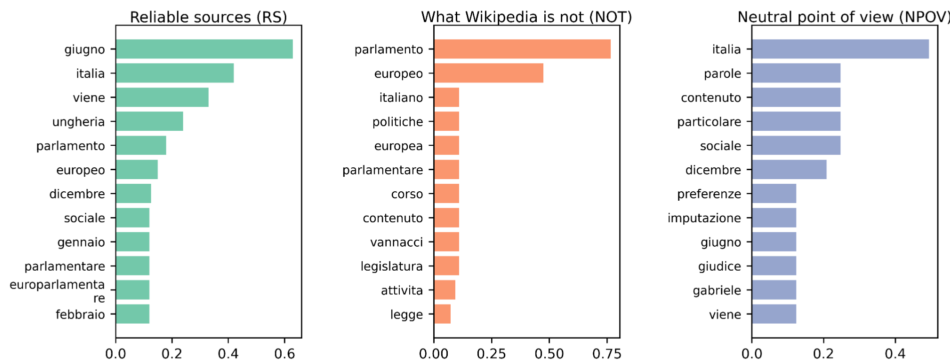

Figure 63: Deleted words in policy-citation edits for the Italian Wikipedia. The 12 most frequent words deleted when a certain policy is invoked in the edit summaries or comments, for the 3 top policies, with TF-IDF normalization.

Policy mention patterns

We now look at which kinds of users tend to invoke policies, and to what extent (Figure 55). Like for the English, we see a very low number of policy citations by anonymous users in the French Wikipedia, while this value is higher for the other language editions, especially Italian; on the right panel of each graph we can see that anonymous and bots, taken together, have a higher presence in Spanish. Looking at user tiers, we see a quite different picture across languages: in French and Italian we find mostly administrators, while in German we see a prevalence of Level 2 users (Tech admins and patrollers), and no presence of Level 4 users.

These variations suggest that, while all Wikipedia communities share the same fundamental policy framework and organisation into user roles, they develop distinctive practices with respect to policy invocation, hierarchical governance and the role of policy enforcement. Understanding these differences helps to appreciate how the same policy system generates different governance practices across linguistic and cultural contexts.

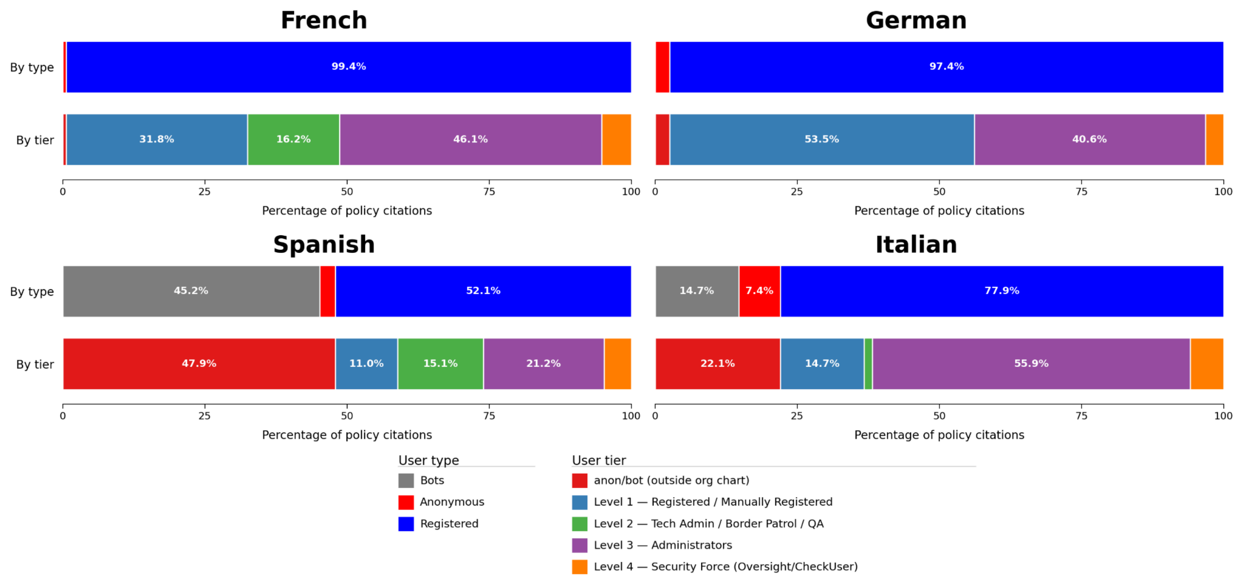

Figure 64. Proportion of policy citations by user type and tier in different languages. Bar charts showing the proportion of edits invoking policies by user type (top) and by user tier (bottom) in 4 different language editions: French (top left), German (top right), Spanish (down left) and Italian (down right) during the electoral period. The colours indicate the user type (Anonymous in red, Registered in blue and Bots in grey), and the user tier (Anon/bot in red, level 1 in blue, level 2 in green, level 3 in purple and level 4 in orange).

In the Spanish Wikipedia we see an extraordinary presence of bot activity in citing policies, that we do not observe in other language editions (Figure 64). Looking at the policies most cited by each user type in Figure 65, we find that this extraordinary activity of bots is almost entirely due to citations of a single policy, Citing sources, suggesting that this policy is invoked in automatic messages written by a source manager bot devoted to fixing sources properly. “What Wikipedia is not” is the most cited policy for most user groups, with the exception of Level 4 editors, where the vast majority of citations are related to managing Vandalism.

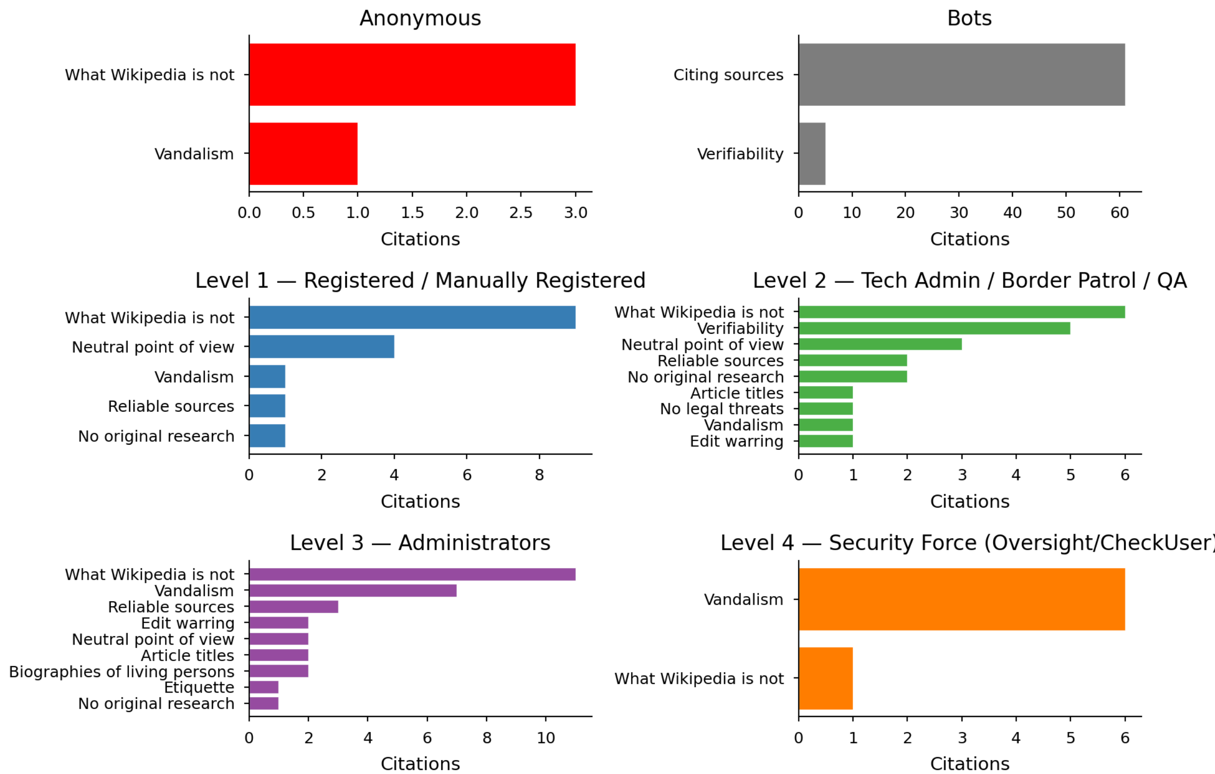

Figure 65. Citations ranking of the top policies cited by each user type/tier in Spanish Wikipedia. Histogram of citations of the top policies cited by each type of user (Anonymous and Bots) and, in the case of Registered users, by the different user tiers during the electoral period.

Policy network and clusters

In the following figures, we show the networks of policies obtained for each language edition, and the word clouds with the words most frequently co-occurring with the policies in each cluster.

The network analysis for each language edition reveals both universal patterns in how policies cluster and language-specific variations reflecting distinct editorial priorities and challenges during the electoral period. By examining the similarity matrices, network visualizations, and contextual word clouds together, we can understand both the structural organisation of policy frameworks and the specific concerns they address in different linguistic communities.

German Wikipedia

The network for the German Wikipedia (Figures 66 and 67) is similar to the English (Figures 39 and 40) in the sense that it shares the same set of central nodes, with some differences. The clusters encountered are slightly different as the policies related to legal issues and Biographies of Living People (BLP) are in a cluster together with some of the main policies like Neutral Point of View (NPOV).

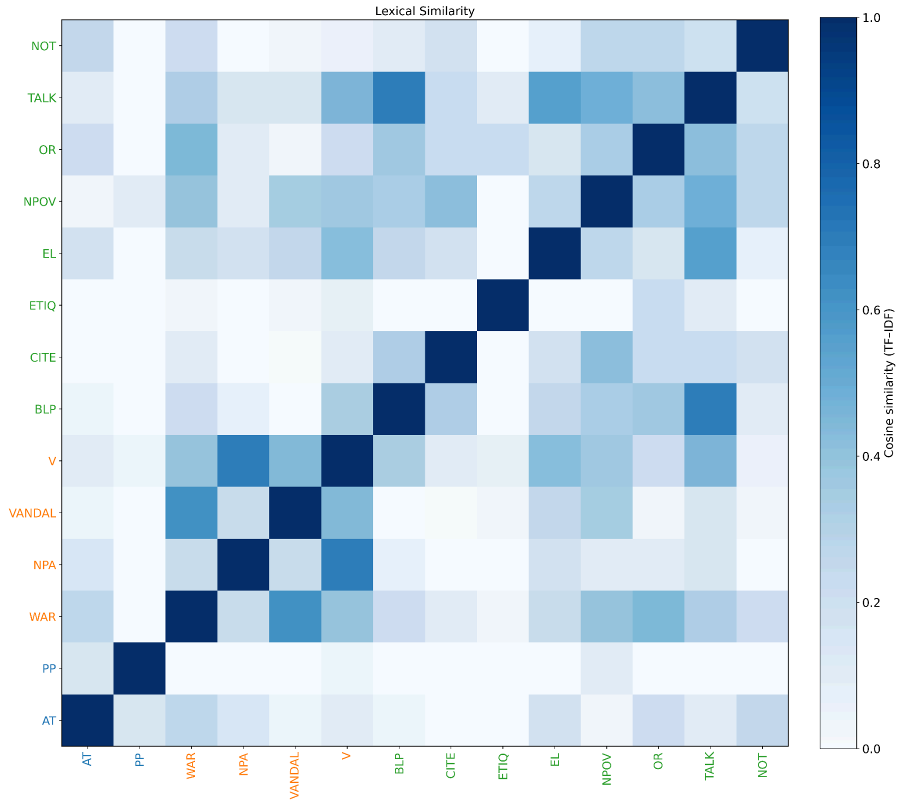

Figure 66. **Lexical similarity relation between the different policies in German Wikipedia.** Lexical similarity matrix between policies, computed as the cosine similarity between the lexical context in which they are mentioned. Colours of the policy names represent the clusters identified through the Louvain method for community detection.

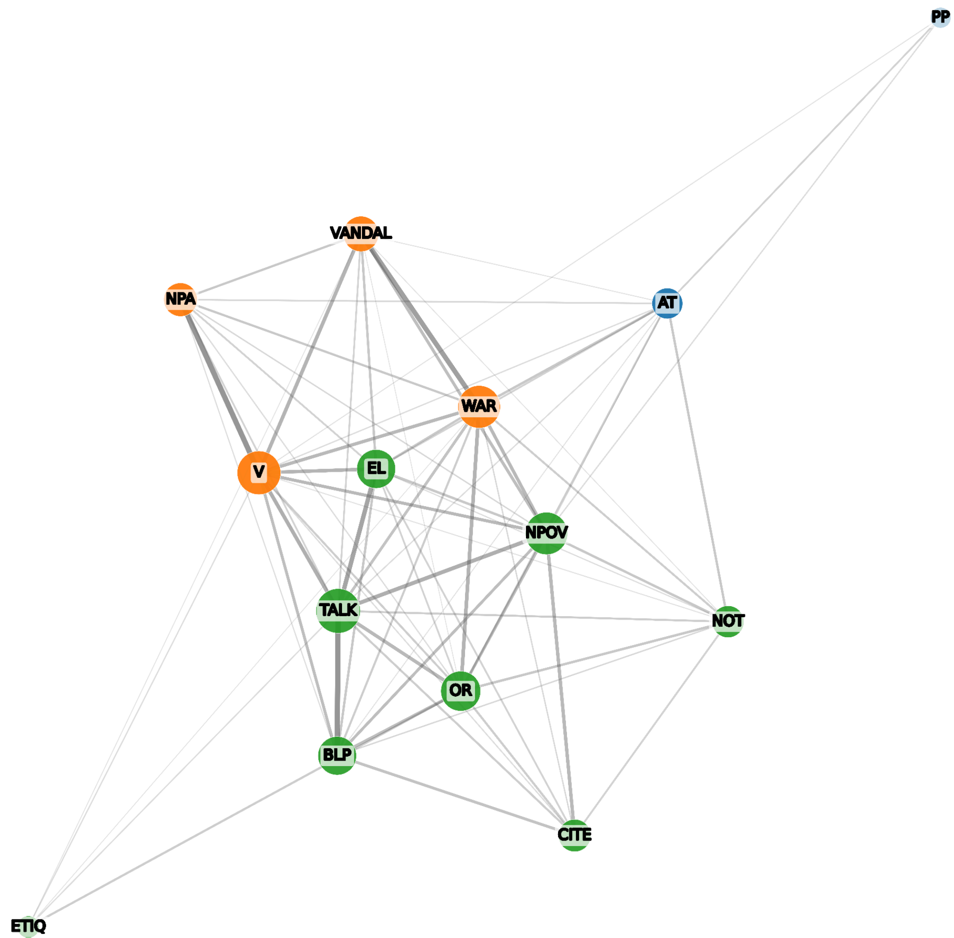

Figure 67. Network of lexical similarity of German Wikipedia policies in the electoral context. Representation of the network, where policies are nodes, connected via edges showing the lexical similarity. The width of the link between two policies is proportional to the lexical similarity between them. Node size represents the strength of the node, i.e. how similar it is to other policies in the network, and node colours represent the communities identified through the Louvain method for community detection.

Figure 68. Clouds of words for each policy's communities in German Wikipedia. Word cloud representation of the lexical context of each community of policies. Word size is proportional to the sum of the frequency of each word co-occurring with the policies from a given cluster. The colours represent the clusters identified through the Louvain method for community detection.

The word clouds for German policy clusters reveal the specific editorial discourse in each domain. Cluster vocabulary includes technical Wikipedia terminology alongside substantive political discussion, with terms related to parties, elections, and political positions. This integration of meta-editorial and substantive political vocabulary suggests that German Wikipedia's policy invocations occur in contexts where editors simultaneously negotiate both procedural questions (proper sourcing, appropriate scope) and substantive issues (political characterization, ideological framing). Particularly notable is the vocabulary around neutrality and sourcing policies, which includes both technical verification terms and substantive political language.

French Wikipedia

The similarity matrix for the French Wikipedia (Figure 69) shows strong connections within policy groups but also notable bridge policies connecting different governance domains.

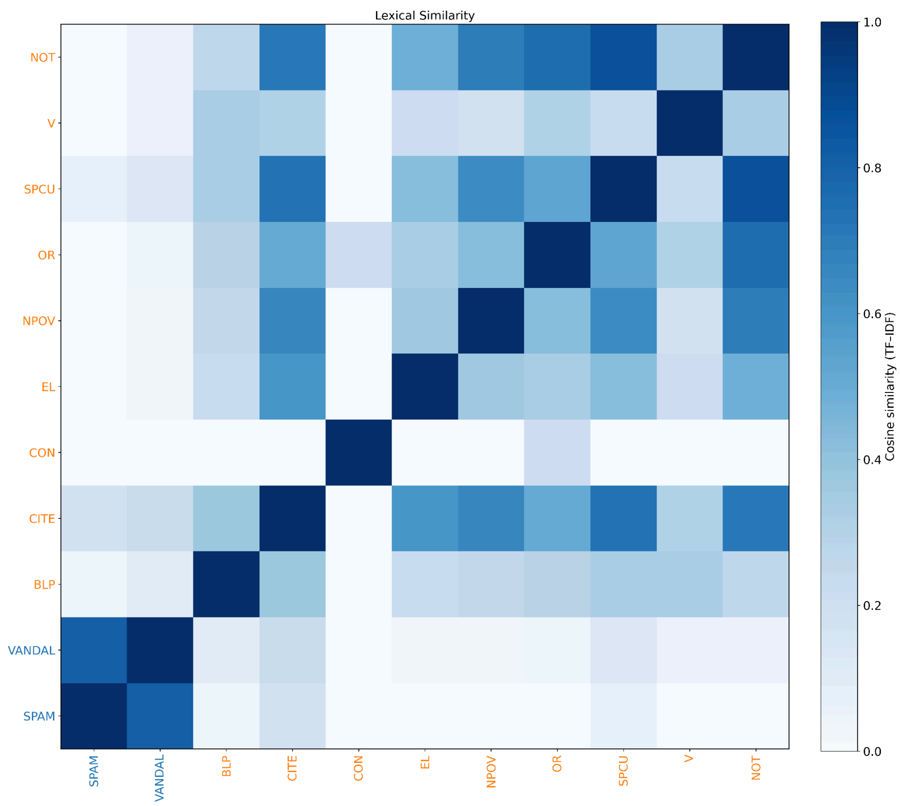

Figure 69. **Lexical similarity relation between the different policies in French Wikipedia.** Lexical similarity matrix between policies, computed as the cosine similarity between the lexical context in which they are mentioned. Colours of the policy names represent the clusters identified through the Louvain method for community detection.

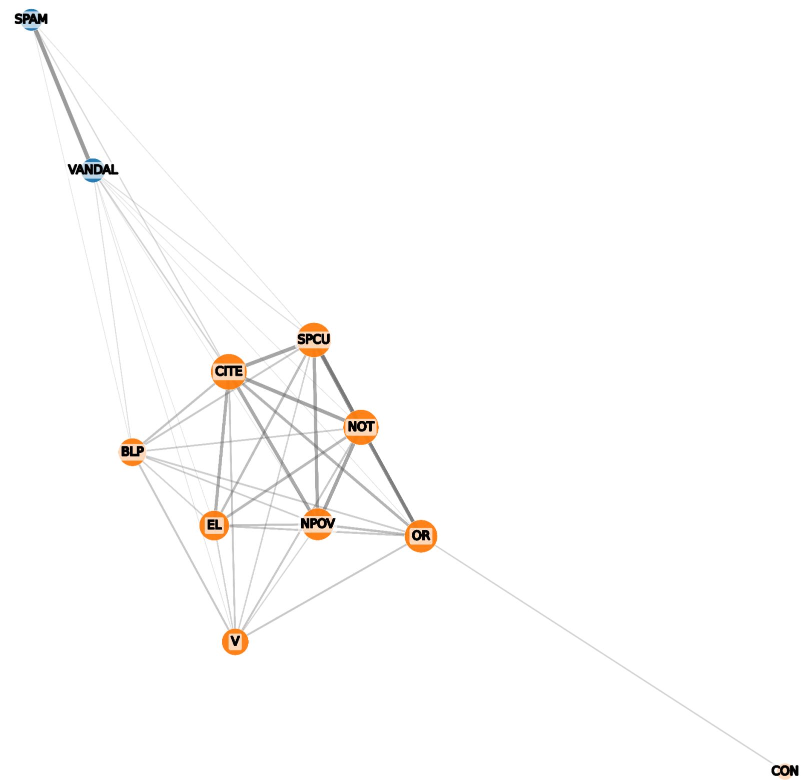

Figure 70. Network of lexical similarity of French Wikipedia policies in the electoral context. Representation of the network, where policies are nodes, connected via edges showing the lexical similarity. The width of the link between two policies is proportional to the lexical similarity between them. Node size represents the strength of the node, i.e. how similar it is to other policies in the network, and node colours represent the communities identified through the Louvain method for community detection.

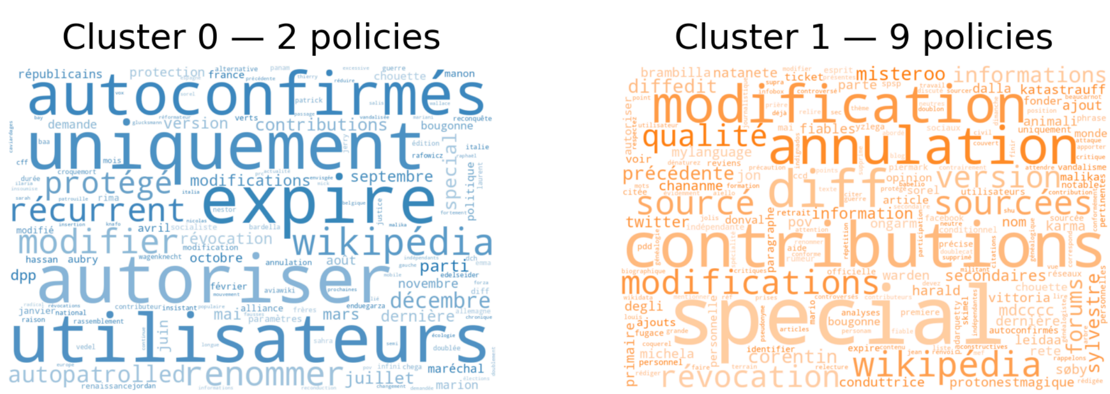

Figure 71. Clouds of words for each policy's communities in French Wikipedia. Word cloud representation of the lexical context of each community of policies. Word size is proportional to the sum of the frequency of each word co-occurring with the policies from a given cluster. The colours represent the clusters identified through the Louvain method for community detection.

The word clouds shown in Figure 71 provide insight into the specific concerns characterizing each cluster's usage in French; words related to Wikipedia practices dominate both clusters.

Spanish Wikipedia

Spanish Wikipedia's policy network displays its own distinctive features, while sharing some commonalities with other major language editions.

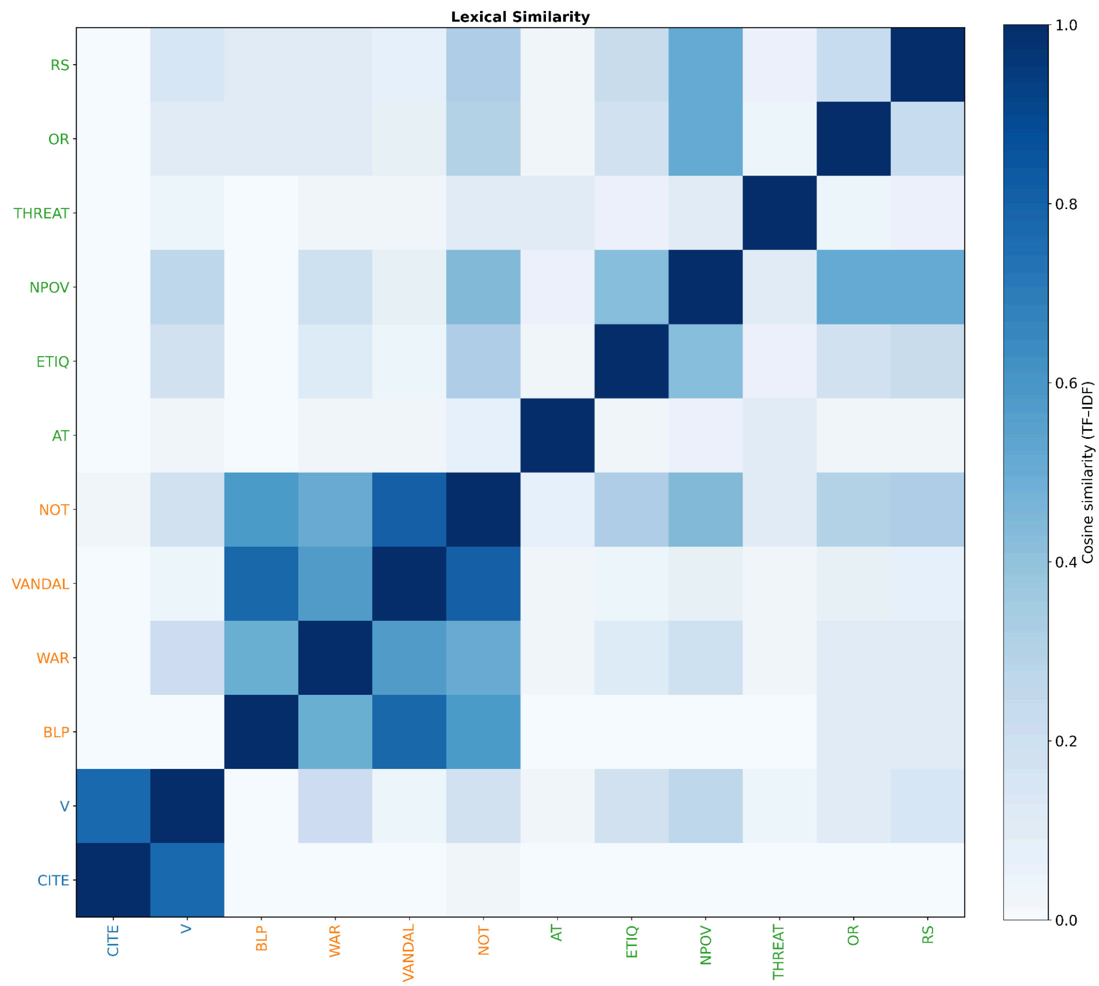

Figure 72. **Lexical similarity relation between the different policies in French Wikipedia.** Lexical similarity matrix between policies, computed as the cosine similarity between the lexical context in which they are mentioned. Colours of the policy names represent the clusters identified through the Louvain method for community detection.

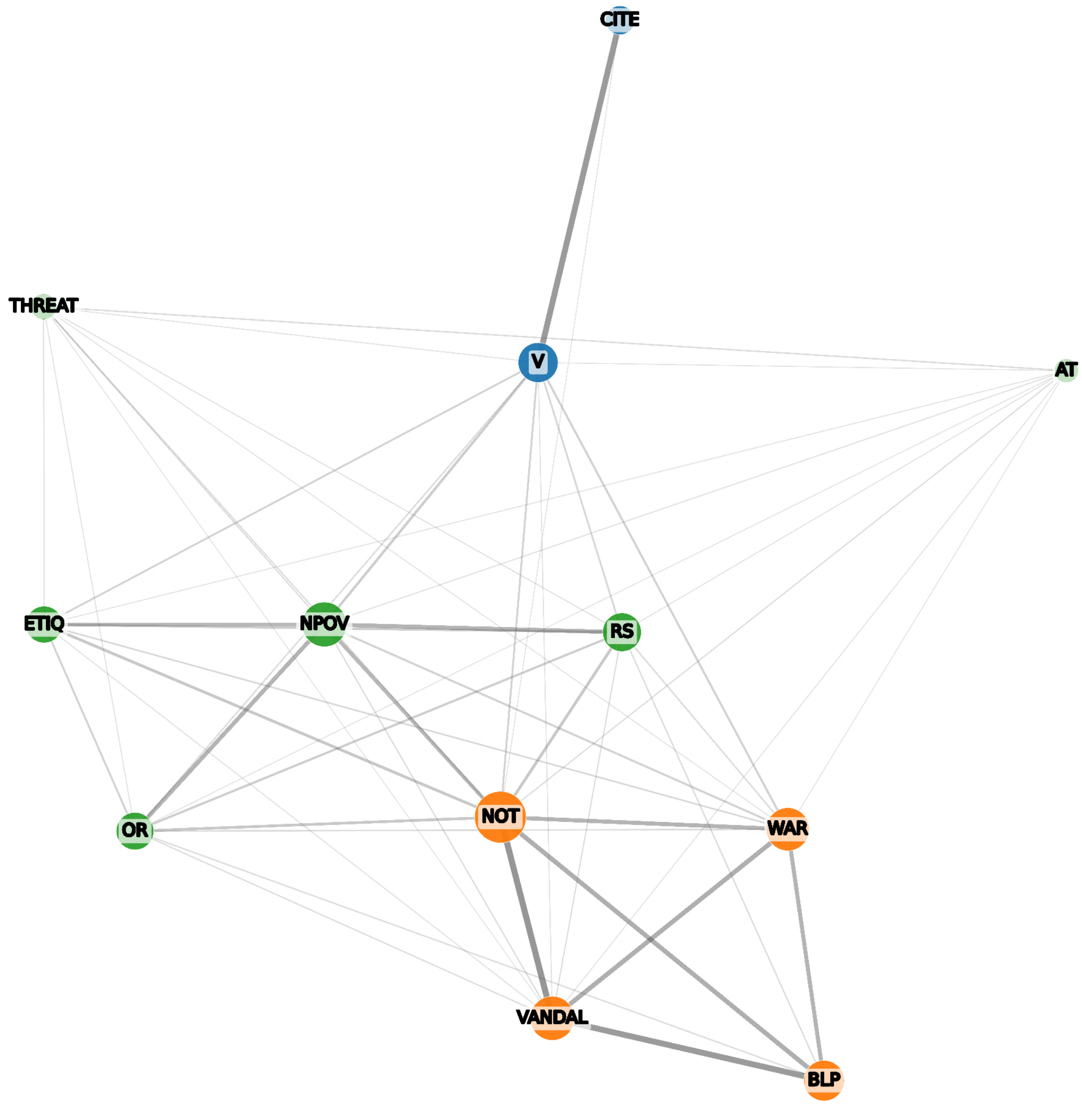

Figure 73. Network of lexical similarity of Spanish Wikipedia policies in the electoral context. Representation of the network, where policies are nodes, connected via edges showing the lexical similarity. The width of the link between two policies is proportional to the lexical similarity between them. Node size represents the strength of the node, i.e. how similar it is to other policies in the network, and node colours represent the communities identified through the Louvain method for community detection.

Particularly interesting is the prominence of "What Wikipedia is not" (NOT) in Spanish deletion patterns discussed earlier, which corresponds to its central position in the network (Figure 73). Spanish editors apparently frequently invoked this boundary-setting policy to establish what content belongs in electoral coverage versus what constitutes inappropriate opinion, advocacy, or non-reliable.

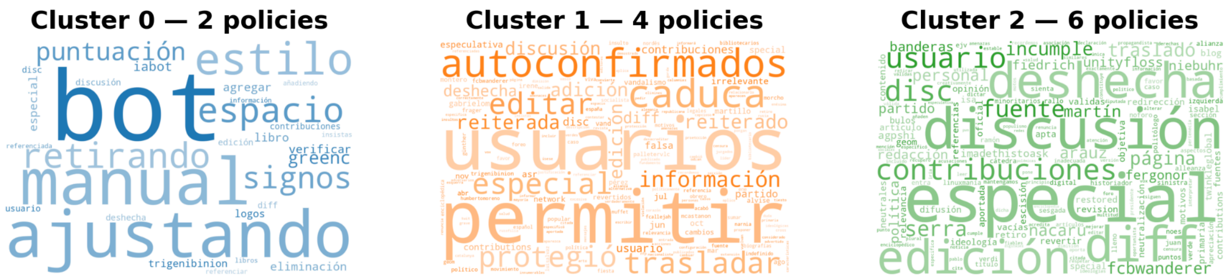

Figure 74. Clouds of words for each policy's communities in Spanish Wikipedia. Word cloud representation of the lexical context of each community of policies. Word size is proportional to the sum of the frequency of each word co-occurring with the policies from a given cluster. The colours represent the clusters identified through the Louvain method for community detection.

The word clouds shown in Figure 74 reveal the lexical contexts in which Spanish editors invoke different policy groups. Cluster 0 corresponds to the policies more cited by bots (Citing sources and Verifiability), as seen in Figure 65. Actually the word “bot” dominates this cluster, followed by a reduced number of other words mostly related to editing practices for adequately citing sources. In the other two clusters, words related to Wikipedia editing are prominent.

Italian Wikipedia

The Italian Wikipedia's policy network completes our cross-linguistic comparison, revealing how Italy's Wikipedia community structures its governance framework during electoral coverage. The network shows a relatively compact structure with fewer policies accounting for most activity, suggesting Italian editors may rely more heavily on a core set of frequently invoked policies.

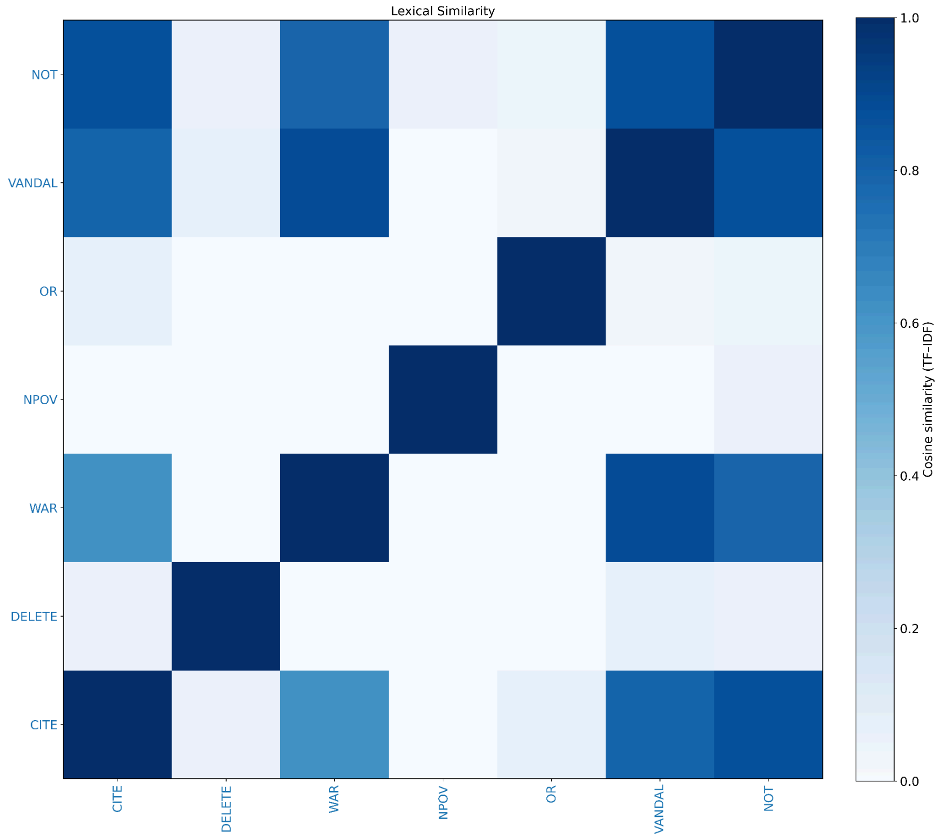

Figure 75. **Lexical similarity relation between the different policies in Italian Wikipedia.** Lexical similarity matrix between policies, computed as the cosine similarity between the lexical context in which they are mentioned. Colours of the policy names represent the clusters identified through the Louvain method for community detection.

In this case the reduced network is cohesive around a core of central policies all connected to each other (Edit warring, Vandalism, What Wikipedia is Not, Citing sources) so that we have no division into clusters.

The word cloud for the cluster representing the whole network, shown in Figure 77, includes again a prominence of words associated with WIkipedia practices and procedures, and then with lower frequency words related to controversial political figures that co-occurred with policy citations in edit summaries and talk pages, like Roberto Vannacci, Matteo Salvini and Ilaria Salis.

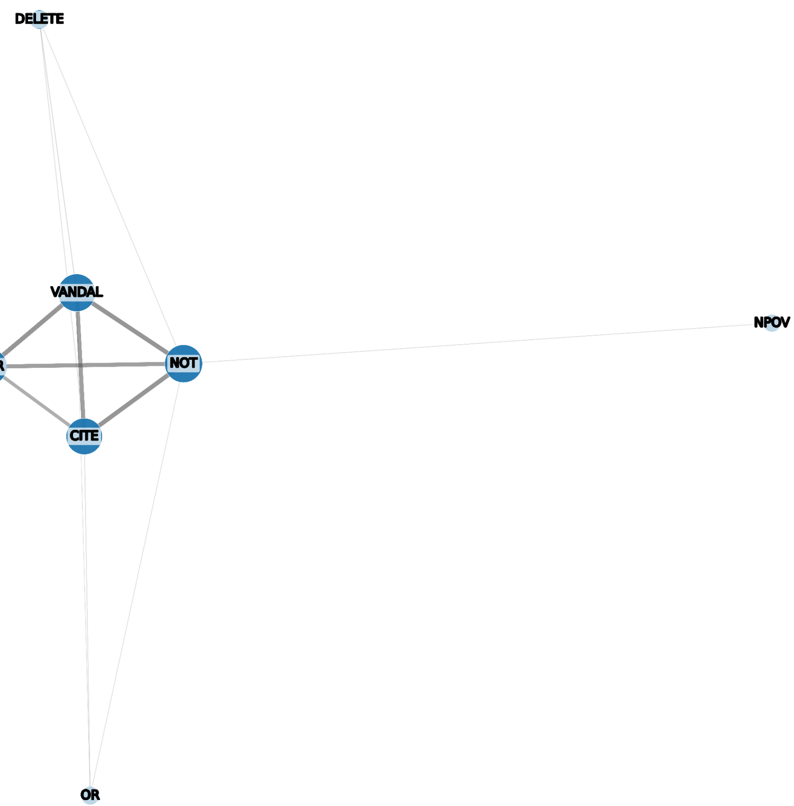

Figure 76. Network of lexical similarity of Italian Wikipedia policies in the electoral context. Representation of the network, where policies are nodes, connected via edges showing the lexical similarity. The width of the link between two policies is proportional to the lexical similarity between them. Node size represents the strength of the node, i.e. how similar it is to other policies in the network, and node colours represent the communities identified through the Louvain method for community detection.

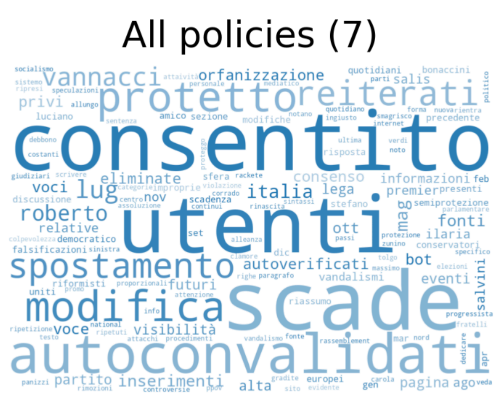

Figure 77. Clouds of words for the policy's community in the Italian Wikipedia. Word cloud representation of the lexical context of each community of policies. Word size is proportional to the sum of the frequency of each word co-occurring with the policies from a given cluster. The colours represent the clusters identified through the Louvain method for community detection.

Conclusions

The analysis presented offers a complementary view to the legal analysis in deliverables D2.3 and D2.4, based on real data from the daily work of the Wikipedia communities during the European Parliament Elections of 2024.

The deliverable provides insights on community dynamics according to different aspects, such as temporal dynamics during and after the electoral campaign, patterns of activity, deliberation, conflict, vandalism and readers' attention according to different metrics, the relationships between such metrics, the role and behaviour of editors according to their kind of profile and level, the words more frequently added and deleted in edits and discussions, in vandalism edits and in relation to gender of the politicians' pages. Results are shown only for selected language editions but are available for all language editions on the online repository of the project (Abella and Laniado, 2025b). Regarding the relationship between cultures and countries, the analyses show that, although every linguistic community is mostly concentrated on its related countries, there exist some relevant cases of strong cross-country interest by the readers and especially the editors.

The work on policies represents a first step for shedding light on how policies are used by the communities in their daily work in the context of a sensitive political event, unveiling the policies that are more frequently invoked by editors in the disputes surrounding the elections, such as the ones on Vandalism, Neutral point of view, or Reliable sources, as well as the kind of users that invoke them, the context in which each of them is cited, and the words deleted when mentioning each policy.

The findings on the relationships between policies and the resulting networks and clusters of related policies help to map governance domains, such as the ones related to the core criteria for acceptability of content according to the main pillars of Wikipedia, to legal issues and biographies of living people, or to abusive editing and administrative actions, and the relations between them. The analysis for the major European language editions highlighting common elements and differences between communities, such as the focus of German Wikipedia on Neutral point of view and Edit warring, or of the Italian and French Wikipedia on Reliable sources.

The analysis of policy invocation by different kinds and levels of users confirms the organisation of work within the community, where different editor profiles focus on different kinds of issues: anonymous and regular users focus mostly on content neutrality and sources, while administrators and higher level editors focus on user conduct and conflict management.

This is to the best of our knowledge the first effort to extensively analyse the way Wikipedians cite policies in their edit summaries and talk page comments using a computational approach to extract policy mentions; this is made possible by the open logs of Wikipedia containing the whole history of revisions, and by the availability of lists of policies and their acronyms for different language editions, although some technical issues stem from manually curated lists and language ambiguity, making not trivial the extraction process.

One limitation we found is with the size of the dataset, especially for medium and minor language editions. We believe this could also be due to the reduced interest of citizens for European elections as compared to national elections. We could expect higher interest of the public for national elections, which may be perceived by citizens as closer and with a more direct impact on their life and society, and therefore could typically attract more attention and activity on the related Wikipedia pages. Therefore, a first possibility of a line for future work is applying this analysis framework to the case of national elections.

As a further step, we are considering the possibility to extend this framework also to other kinds of events or topics, which would also allow for comparison between topics, and to identify and distinguish patterns that are specifically associated with political elections. Finally, it would be possible to think of applying this framework to a whole Wikipedia edition. This would allow one to get a comprehensive view of when, where, how, why and by whom policies are cited within a certain community and across topics.

Another direction we are considering is developing an interactive tool with dashboards to explore the data and the results of our analyses, in the same line as Miquel-Ribé et al (2022). Indeed, we have selected only a few results for showing in this deliverable, compared to the amount of data, results and plots we obtained for 31 language editions. However, one could have a specific interest in a certain language edition, page or subtopic of the elections or event or other topic under analysis. This could be the case of Wikipedia communities, policy makers, researchers, journalists or citizens. Developing a tool to make our results accessible through an interactive interface could help to increase community awareness and unveil community practices and dynamics behind a given topic.

All in all, we believe the value of this work does not only lie in the results obtained for this specific use case, but also in having developed and made available a novel framework for data extraction and analysis, in particular with respect to policy citations, whose usage can go beyond the scope of this project.

References

Abella, D. & Laniado, D. (2025a) Exploring Wikipedia community practices during the 2024 European Parliament election. *Wiki Workshop 2025*.

Abella, D., & Laniado, D. (2025b). Computational analysis of the Wikipedia editorial activity during the 2024 European Parliament election. *Zenodo. https://doi.org/10.5281/zenodo.17960820*

Arazy, O., Ortega, F., Nov, O., Yeo, L., & Balila, A. (2015). Functional roles and career paths in Wikipedia. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing* (pp. 1092-1105).

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.

Borra, E., Weltevrede, E., Ciuccarelli, P., Kaltenbrunner, A., Laniado, D., Magni, G., ... & Venturini, T. (2015). Societal controversies in Wikipedia articles. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 193-196).

Butler, B., Joyce, E., & Pike, J. (2008, April). Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1101-1110).

Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129-1164.

Jemielniak, D., & Aibar, E. (2016). Bridging the gap between wikipedia and academia. *Journal of the Association for Information Science and Technology*, *67*(7), 1773-1776.

Kaltenbrunner, A., & Laniado, D. (2012). There is no deadline: time evolution of Wikipedia discussions. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (pp. 1-10).

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31(1), 7-15.

Kharazian, Z., Starbird, K., & Hill, B. M. (2024). Governance capture in a self-governing community: A qualitative comparison of the croatian, serbian, bosnian, and serbo-croatian wikipedias. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1-26.

Lehmann, J., Müller-Birn, C., Laniado, D., Lalmas, M., & Kaltenbrunner, A. (2014, September). Reader preferences and behavior on Wikipedia. In *Proceedings of the 25th ACM conference on Hypertext and social media* (pp. 88-97).

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Menking, A., & Rosenberg, J. (2021). WP: NOT, WP: NPOV, and other stories Wikipedia tells us: A feminist critique of Wikipedia's epistemology. *Science, Technology, & Human Values*, *46*(3), 455-479.

Miquel-Ribé, M., & Laniado, D. (2018). Wikipedia culture gap: quantifying content imbalances across 40 language editions. *Frontiers in physics*, *6*, 54.

Miquel-Ribé, M., Consonni, C., & Laniado, D. (2022). Community vital signs: measuring Wikipedia communities' sustainable growth and renewal. *Sustainability*, *14*(8), 4705.

Miquel-Ribé, M., Laniado, D., & Kaltenbrunner, A. (2021). The role of local content in Wikipedia: A study on reader and editor engagement. *Área Abierta. 2021; 21 (2): 123-151*.

Sala, A., Vinagre, J., Esposito, E., Fortuna, P., Consonni, C., Porcaro, L., ... & Gomez, E. Shut up and do the dishes! Toxicity and Gender in Digital Political Spaces: Evidence from the 2024 European Elections. Retrieved from: https://osf.io/preprints/socarxiv/8xmh9_v1

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513-523.

Steinsson, S. (2024). Rule ambiguity, institutional clashes, and population loss: how Wikipedia became the last good place on the internet. *American Political Science Review*, *118*(1), 235-251.