Case study in political user behavior on Wikipedia

CHRISTOFFER NEPPARE

PONTUS BLOMBERG

Fallstudie i politiskt användarbeteende på Wikipedia

Christoffer Neppare
Pontus Blomberg

Stockholm, Sweden May 11, 2016

Abstract

Wikipedia has since its inception in 2001 become the world's largest encyclopedia and continues to enjoy a high degree of support from the general public. This report investigate differences in the user creation process of content between political and unpolitical articles. According to the results the revisions per user follow a similar pyramid pattern in both the political and unpolitical article. When investigating the number of reverts the results differ between the political and unpolitical articles. More reverts are made by semi active users in the political articles. The report also establish that some political categories exhibit a significantly longer edit history than the Wikipedia average and that these categories contain a disproportionately high number of reverts among their revisions. Conclusions are that there are differences between political and unpolitical articles in regard to who do reverts and the number of reverts per page.

Referat

Wikipedia har sedan det skapades 2001 blivit världens största encyclopedia och den fortsätter erhålla en hög grad av stöd och tillit från allmänheten. Denna rapport undersöker skillnader i användarbeteende under skapande processen av innehåll för politiska och opolitiska artiklar. Det framkom att politiska artiklar uppvisar en liknande pyramid struktur som opolitiska i hänseende till antalet revisioner per användare men att de skiljer sig i hänseende till reverts, som till en högre grad utförs av mellan-aktiva användare i politiska artiklar. Rapporten tar även och visar att vissa politiska kategorier har längre editeringshistorier än Wikipedias genomsnitt och dessa kategorier innehåller en oproportionerligt hög andel reverts bland sina revisioner. Slutsatser är att det finns skillnader mellan politiska och opolitiska artiklar i hänseende till vem som gör reverts och antalet reverts per sida.

Contents

1 Introduction

E-collaboration is a field which have emerged with the advent of ICT (Information and Communication Technology) and keeps growing at a quick pace. It refers to users creating content collaboratively with the web as a medium. E-collaboration is a subarea of Web 2.0 where users participate in the creation of content and thus simultaneously become consumer and creator (O'Reilly, 2007). A major actor in the Web 2.0 is Wikipedia which describes itself as a collaborative online encyclopedia (Wikipedia, 2016a).

Wikipedia has since its inception 15 years ago become the world's largest encyclopedia and the 7:th most visited website according to Alexa (2016). It continues to have a high degree of support and trust from the general public, often being one of the first places to visit when checking facts. It is thus of relevance to know of any cases where reliability might be impacted. Since Wikipedia is made up by user created content it is of interest to know which articles there have been conflict on when the articles were created. Seeing that political issues causes controversies in real life it might likewise affect the quality of the articles which covers political topics.

This paper will look at political and nonpolitical articles and discern whether the former exhibits different user behaviour compared to the latter. Another focus of this study is to determine whether the political articles have a more aggressive edit behaviour with more controversial edits and more reverts. This will be done through a quantitative analysis which is made by comparing category-samples of strongly political articles to a randomized control sample.

1.1 Problem statement

Do strongly political articles demonstrate systematic differences in regard to number of edits or reverts per user to less political articles?

1.2 Scope

This paper will investigate systematic differences in registered users' contributions to Wikipedia articles.This will be done by comparing articles from designated political categories and less political articles. Unregistered users will be omitted due to Wikipedia's structure not allowing us to uniquely identify them. The sample size is restricted to six categories and one randomized control sample. This report does not aim to analyze the reasons for any systematic differences and whether or not they impact the reliability of the articles, only to learn if there is a difference at all.

1.3 Purpose

The purpose with this report is to provide a quantitative study which describes the forming of a consensus within the community of Wikipedia, specifically within the political domain. This is a new research topic since there haven't been a quantitative study which compares political and non-political articles before with regard to their creation and maintenance. The report will determine if the political articles on Wikipedia is the subject to more vandalism than less political articles.

2 Background

Wikipedia has been the subject of both quantitative and qualitative studies. Scientific articles regarding the user base and usage form the majority of the corpus concerning Wikipedia. Since the user base is weighted by some user groups (e.g. male users) it might have an impact on the the way wikipedia is utilized in certain domains (Antin et al. 2011). Studies regarding the usage delve into aspects of organizational as well as discursive theory. The organizational theory is used to describe and analyze how organizations are formed to handle the task of editing a Wikipedia article. The organizational forms affect the quality and content of the created article since the process of creating and who creates the content differs (Keegan et al. 2013). The organizational theory is also used to describe the organization's maintenance and eventual decay (Mattus, 2014). The discursive theory is used to analyze how a common value system is formed within Wikipedia and within domains of Wikipedia (Dahlberg 2007, Hansen 2007).

2.1 The Wikipedia Project

2.1.1 The structure of Wikipedia

Wikipedia is an encyclopedia with quality comparable to that of traditional encyclopedias (Niederer et al. 2010). There exists articles, a general structure and users. The structure shows how the articles are related to each other. The users can add or remove relations between articles. The users can also read and alter content of the article, either through an edit or through a revert. The edits, which are also called revisions, change an article by adding and/or removing content. A revert undoes the edits made by another user (Wikipedia, 2016b). Wikipedia uses guidelines which the users are expected to follow. Repeatedly breaking the guidelines might cause a user to be blocked from further editing (Wikipedia, 2016c).

2.1.2 The lifespan of an article

Previous research has shown that the life of an article follows some patterns. An article start as a small "stub" providing an outline. Later the article will increase in size and references will be added. The last phase concerns maintenance, which means that the article has reached a stable size with a few number of minor edits mainly regarding grammar. However articles may differ in their creation. During the swelling phase the editors' opinions might conflict causing a spike in "edits" and "reverts", known as an "edit war". The duration and intensity of edit war differs. Afterwards the articles will reach a consensus, leading to the beginning of the maintenance phase (Mattus 2014, Török et al. 2013). Different articles cover certain domains of Wikipedia, e.g. the political domain. An earlier study focus on specific articles and, within these articles, performs a quantitative analysis (Brandes et al. 2008).

2.2 User created content

2.2.1 Internet quarrel

User created content describe content where the users contribute and in some cases maintain the common material. Users can contribute and thus become contributors.The contributed content can have both a positive and negative impact on the community (Chen et al. 2011). When the content added is controversial a discussion or quarrel might erupt. If the users can maintain other contributions than their own they might delete said contributions, which could lead to an increasing conflict.

2.2.2 Political topics

Sentiments within user created content such as an user written article can be based on a varying degree of political thoughts (Malouf et al. 2008). The term political articles are articles which handles subjects of a political nature. The term political can be referred to the term Politics which means "The theory or practice of government or administration."(Oxford English Dictionary, 2016).

2.2.3 Bots and scripts

Web pages which are dependent on user creator content can have robots partly control the content (Niederer et al., 2010). The robots can either be used by the administrators or in some cases the users. This means that contributors are sometimes in fact fully- or semi-automatic "bots", short for robots, or scripts. They are computer programs which can have different purposes (Wikipedia, 2016d). Two common types are grammar scripts and anti-vandalism scripts (Wikipedia, 2016e). The purpose of the former is to rewrite content so it complies with the grammar of a given language. The purpose of anti-vandalism scripts is to revert content to a prior state if the content have been the subject to vandalism.

2.3 Previous research

Previous research has shown mainly two characteristics which are commonly referred in this article. The first being that the articles' lifespan follow similar patterns (Mattus 2014, Török et al. 2013). The second being that there are systematic difference in how certain groups of articles display differences in user editing behaviour (Keegan 2013).

3 Method

The main method of analysis in this thesis is a quantitative study using metadata from Wikipedia. The choice of a quantitative study was made with two reasons in mind. The first was that controversy could be interpreted as the number of disagreements. This means that the qualitative implications need not be the focus of this study. The second reason being that a quantitative data would be easier to evaluate

By extracting the XML which describes the 'edit history' of Wikipedia, several quantitative analysis methods can be applied.

3.1 Data

We identified six broad categories of Wikipedia articles likely to include more political bias. These include political activism (Damascus Declaration, Gaza flotilla raid), political controversies (Christmas controversy, Tibetan sovereignty debate), political riots (Anti-austerity movement in Greece, 27th G8 summit), political party founders (Abraham Lincoln, Julian Assange), political theories (Agrarianism, Liberal conservatism) and political schisms (Australian Labour split, Sino-Soviet split). The reason for picking six categories was that the categories chosen should be strongly political and that the sample should not be too small. The articles were extracted with Wikipedia Category members API (Wikimedia, 2016a).

By using the English Wikipedia's comprehensive meta history archive (Wikimedia, 2016b) data was gathered with regard to the number of edits per contributor. This data was then compared between the political articles and a randomized control group. The control group was randomized with a Random API (Wikimedia, 2016c). The randomized sample size was chosen to be 200 articles since it was deemed to cancel out any anomalies. This data was then used to compare whether the user classes described earlier differed between the political sample and the reference sample.

In order to limit sources of error, contributors identified by IP address rather than an unique ID were removed as the architecture of Wikipedia can assign several persons to the same, or different, IP addresses. The probable adverse outcomes of this measure will be covered in the later discussion. Contributors with usernames containing "bot" or "script" were included as they were deemed as being of interest. They are per Wikipedia's user policy only intended to be used by non-human contributors and they make up an important part of Wikipedia's routine patrolling for malicious or faulty content.

3.2 Procedure

3.2.1 Selecting the articles

The article references were extracted using the APIs mentioned in section 3.1. The JSON objects which were returned from the APIs were read with a custom javascript script. The retrieved content was then saved to a text-file. The saved IDs would later be used for comparisons in the following section.

The meta archive mentioned in section 3.1 was downloaded with the help of Free Download Manager (FDM 3.9.7) and later opened with Liquid XML Studio 2016. Liquid XML Studio 2016 was used to inspect the structure of the downloaded XML. The XML was then read with a Java streamreader. Segments were matched with the article IDs procured earlier in 3.2.1. If the read XML-segment matched an article ID from an article in the selected categories, the segment was saved to an appropriate XML-file. As a result the six categories were combined and the reference group were saved as separate XML-files.

The Java library JAXB was then used to create Java objects from the XML. The objects could then be used for comparing the political categories to the reference files. The java objects were then inspected with the methods described in sections 3.2.2 and 3.2.3.

3.2.2 Comparing user classes

The reason for the inspection made in 3.2.2 is to discover how the users are distributed between different user groups. In each article a mean number of revisions per user were calculated. Articles with fewer than 10 revisions or more than 3000 were excluded as to reduce anomalies. The users were then sorted into different user groups depending on whether they had written more than a factor times the average, e.g. users which had done 0 to 1 times the average revisions were sorted into one group.The user groups from all articles were then compared inside each category.

For each user the number of reverts were noted as well as revisions. The reverts were then calculated for each user group. The user groups described earlier were used for the sake of continuity and since not all users also make reverts.

3.2.3 Comparing revisions to reverts

Since reverts are related with revisions the method presented in 3.2.3 lists the number of reverts by the number of revisions as the mean inside each category. When calculating the mean value articles with fewer than 10 revisions or more than 3000 were excluded as to reduce anomalies. The mean is calculated by checking which articles within a given category have at least a given number of revisions. If it has the given number of revisions the reverts within the article is summed to a category total. The total number of reverts is then divided by the number of articles

which contributed to the number of reverts within the category. The result is a normed number of reverts listed per revision within a given category.

3.3 Sifting an XML dump

The XML-document contains a degree of data which is either not needed or not suitable for the type of analysis here performed. The part which were kept for the quantitative analysis were the unique contributor userIDs. The impact of forfeiting this information will be a subject in the later discussion.

4 Results

In section 4.1 the comparisons between different contributor classes are made. The comparisons are made with regard to the number of revisions and reverts between the classes inside each sample. Section 4.2 compares in detail how the number of reverts are partly linked to the number of revisions, which is made for each category. Note that 'unpolitical' herefor refer to the randomized control group and it might therefore not be completely without articles of political nature.

4.1 Contributor classes

This section describes how different contributor classes exhibit different characteristics. The analysis is based on a grouping which sorts contributors into categories given their number of revisions. Those contributing the average or less when dividing the number of revisions with the number of contributors for each article are labelled as a lurker. The groups are listed in table 4.1. The purpose of the labels is to provide an intuitive understanding with regard to the user's contribution. They are specific to this report and hold no standardized meaning.

User group	User contribution as factor of mean contribution
Lurker	<1
Casual	1 to 2
Wikipedian	2 to 4
Expert	>4

Table 4.1 The contributors were ordered into classes.

Graph 4.1 display the distribution of users as per how many revisions they have contributed. Lurkers are the biggest user group and make up 65-75% of the total users in both categories. The following groups with a higher revision ratio is increasingly smaller, following a pyramid pattern. There is a slight skewness in the political bar chart towards the lower end, but there are no major differences. In graph 4.3 one can see the same distribution being present across all the categories.
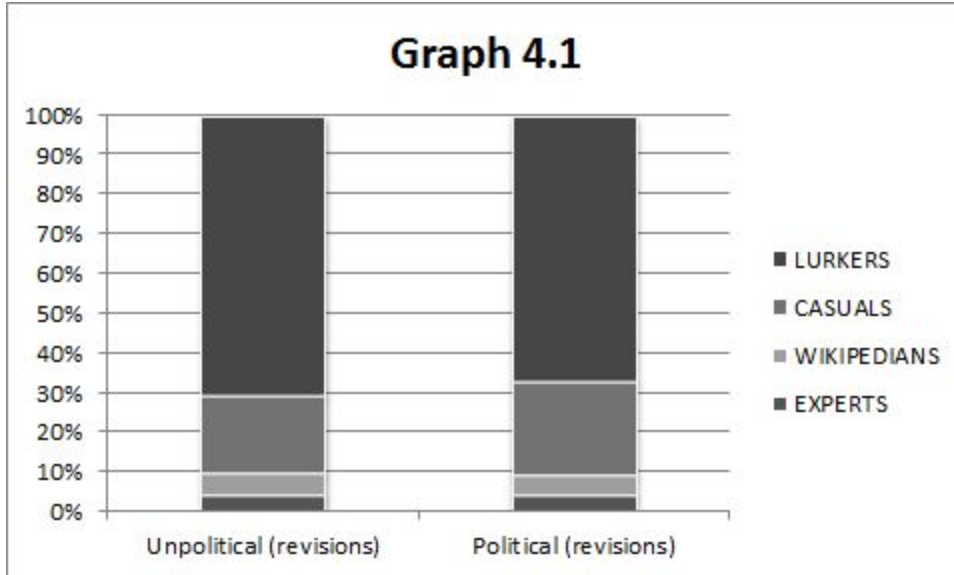
Graph 4.1 Displays the revisions made by the different user groups

Graph 4.2 displays the percentage of the reverts each contributor group contributed with. They are paired with the earlier contributor groups (in graph 4.1) to highlight disproportional representation. In both bar charts the most marked increase is among the Experts, the very frequent users, who contributes with a large part of the reverts. It is more noticeable in the unpolitical bar chart however were the reverts are almost split between the Lurkers and Experts. The political bar chart display a larger middle group of Casuals and Wikipedians leading to a smaller Expert group. This increase in the size of the middle groups are present across all the political categories as shown in graph 4.4.
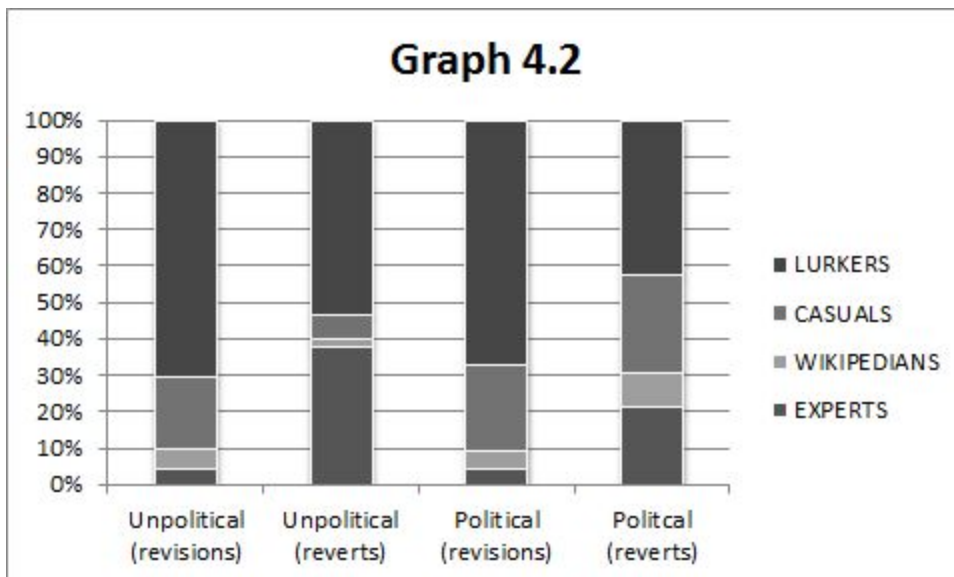
Graph 4.2 Compares the revisions and reverts by user group
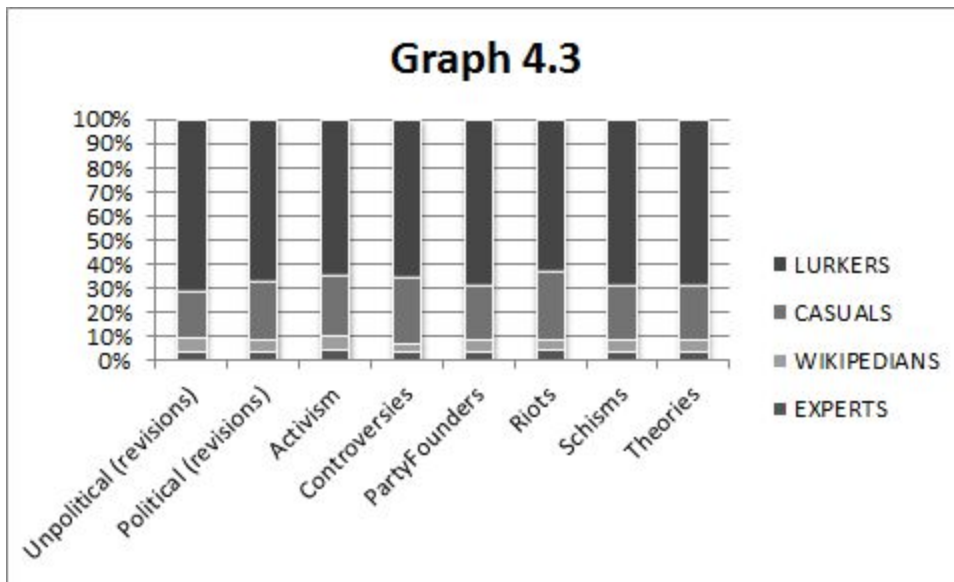
Graph 4.3 Compares the number of revisions per category by user group
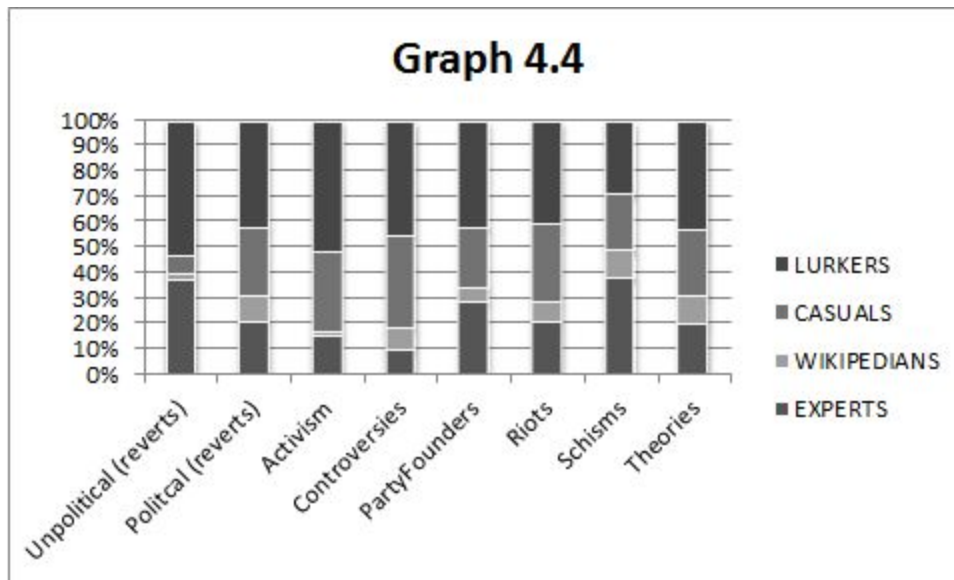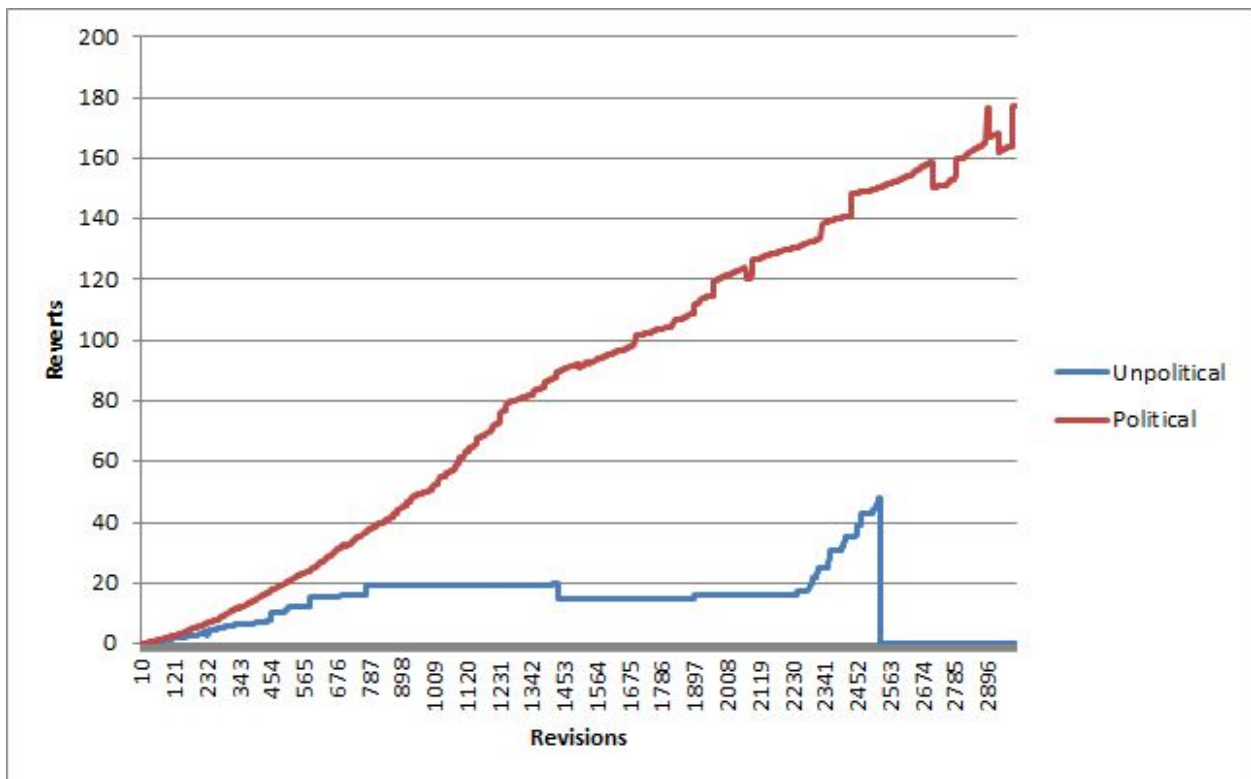

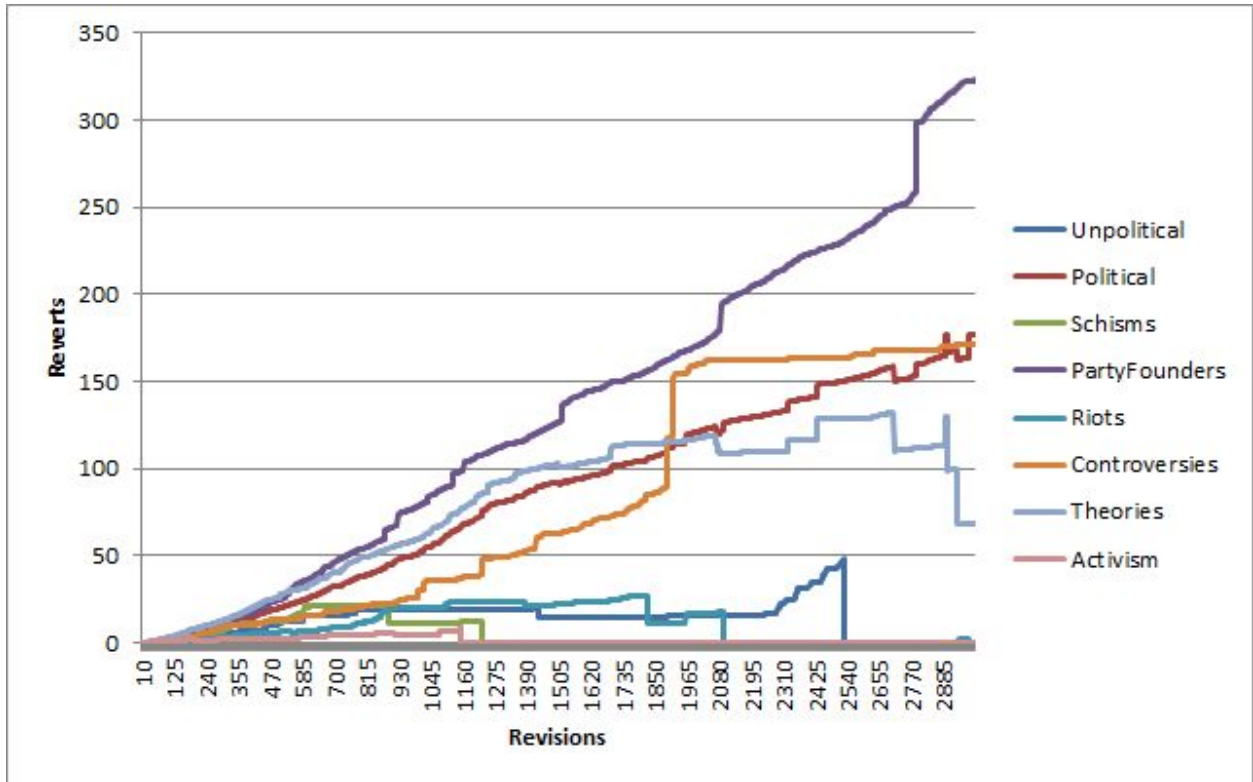Graph 4.4 Compares the number of reverts per category by user group

4.2 Revert ratio by category

The following graphs display the average number of reverts across the number of revisions for different categories. Graph 4.5 shows that the combined average of the political categories is significantly higher than the unpolitical category throughout the main part of the inspected revisions. Graph 4.6 retain both lines for comparison but also show each individual political category as its own line.

Political party founders consequently have a higher revert ratio than the political average. Political controversies and theories alternate above and below the combined average. Political schisms and activism are the categories falling below the unpolitical category. The graphs contain steep rises and falls as the revert ratio is altered. It also contains periods where the reverts are not increased. The biggest category which is the combined political category generally have few spikes.

Graph 4.5 Number of revert across the number of revisions for average political and control sample.

Graph 4.6 Number of revert across the number of revisions per category
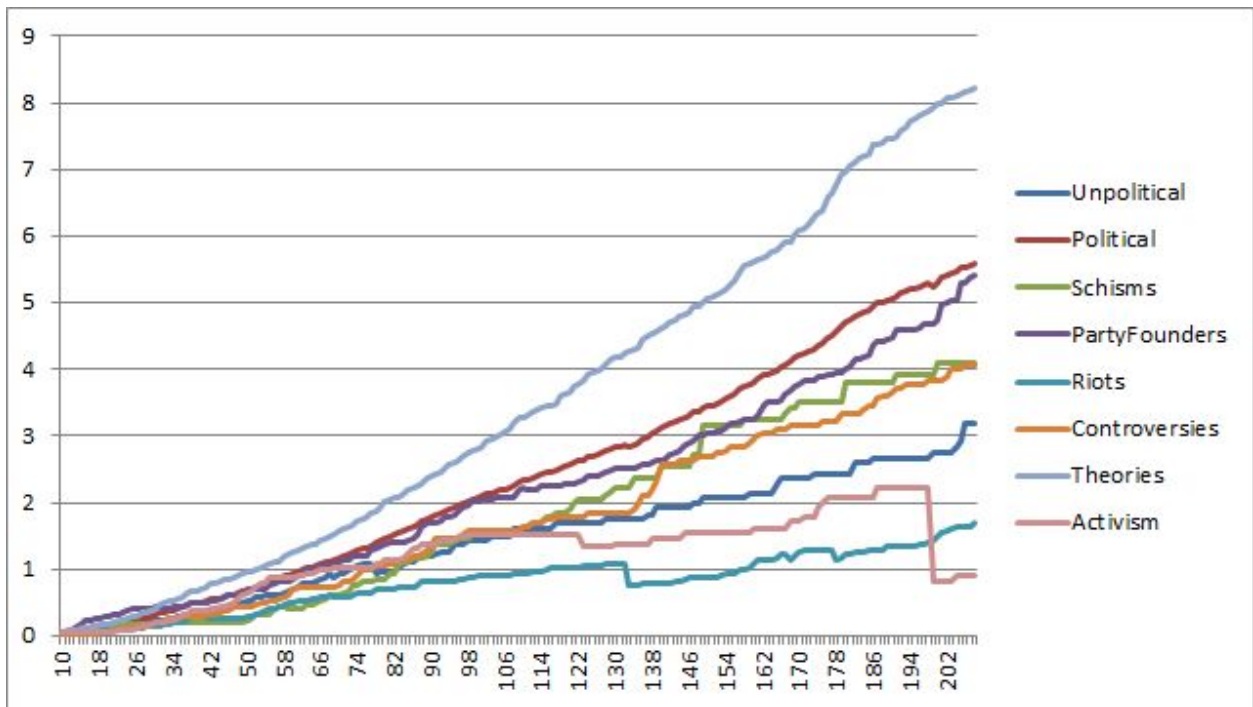
Graph 4.7 Number of revert across the number of revisions per category in the early phase

5 Discussion

5.1 Result analysis

The respective contributor groups showed very little difference in respect to distributions of edits once accounting for the difference in sample sizes. It can be concluded however that the political articles to a larger degree were reverted by the Casuals and Wikipedians. This is interpreted as the political articles on average attracted more contributors with slightly above average interest. This lead to a less dominant Expert group. By leaning on Dahlberg's work we can conclude that the norms reflected in the reverts become less polarized since the difference in reverts between the Lurkers and the Experts is bridged by the Casuals and Wikipedians.

The organization of a political article is different from the average article. There is a more diverse group which reverts the political articles so they comply with Wikipedias guidelines. The many possible issues which might be raised in controversial political articles are closely watched by other groups than the experts. This supports Keegan et al.'s theories that the contributors creates different organizations which can write and maintain the article.

The revert graph exhibits spikes which can be interpreted as two cases. A sudden rise can be interpreted as an article with a low revert ratio that is excluded. The reason for this exclusion is that it does no longer meet the required threshold with regard to the number of revisions. A drop can similarly be interpreted as an article with a high revert ratio that is excluded. The spikes are apparent since the sample sizes are relatively small. The spikes also grow more apparent when the number of revisions increases since there are fewer articles to mitigate the loss of the excluded articles.

Graph 4.6 shows that political articles have a much higher amount of reverts on average in comparison to the number of reverts of non political articles. The graph also shows that the later stages in some cases evens out for periods of time, notably the Controversies category at 2000 edits as well as the unpolitical sample at 1400 revisions. This is partly due to the small sample size. If the line would have represented more articles there would still have been an increase as the combined sample implies. The evening out implies that an article would reach a level where the new contributions are not reverted. This implies that the article reaches a stage where it is uncontroversial. This complies with Mattus study which concluded that the controversy of an article stopped for periods of time. However since the articles are few it would be speculative to assume this is always the case.

5.2 Limitations

Due to limitations in how users identified only by ip addresses are logged at Wikipedia several users from the same geographical area can appear as one user, or the inverse with one "ip user" appearing as several users upon commuting a small distance. To limit the source of error we have removed any non-uniquely identified users, but due to the prevalence of these contributors on Wikipedia they remain a major source of error.

The articles show differences with regard to the number of revisions as well as reverts. This is a potential source of error since big articles have a different impact on the results than small articles. This result is mainly notable in the revert comparisons. The reason for this being that big articles with many revisions have more reverts even after norming with the revision size. This due to small articles not having attracted a high degree of attention. Since the randomized reference sample has a lower mean number of revisions this impact might be reinforced.

A related source of error is that the article with the most reverts within a given category is an outlier with a significantly higher mean number of reverts in said category. This might be interpreted that any following results with regard to reverts are non-generalizable. However even though the number of reverts might be partly concentrated to a small number of articles the other articles still have more articles with a higher rate of reverts on average seeing as the spikes in graph 4.6 each represent the exclusion of an article. If the removed article had an extraordinary rate of revert this would cause a major drop when excluding the articles. This can be seen in some cases, notably in the categories , activism, riots, schisms and theories. The remaining reverts are however still the majority of the reverts. Note that this analysis is only applicable when the article with a high revert ratio has fewer revisions than the remaining articles. Otherwise the article is still in the sample and remain as a source of error.

Another possible source of error is that measuring reverts doesn't catch more subtle forms of disputes not being subject to an outright revert according to Wikipedia's guidelines. A long running controversy conducted without any vandalism could go unnoticed in our report. Irregular users not using the revert function but rather blank out vandalism in a revision also go undetected.

5.3 Future research

The limitation of not measuring other forms of differing opinions than reverts could be remedied by a lexical analysis of the comments accompanying most revisions. The presence of certain words denouncing another contributor could for example suggest an ongoing controversy.

A broader study across a greater range of Wikipedia categories could potentially show that the findings in this report are not unique to political articles, but could instead be subject to some other variable that they have in common. However, were a broader study to find a reliable

metric or threshold for political user behaviour it could be applied across Wikipedia to see whatever other categories exhibit political user behaviour outside the political portal.

A model tracking contributors across articles could potentially find cliches inside the political portal but also in the rest of Wikipedia. A threshold where a certain distance of an induced subgraph to the rest of Wikipedia constitute political or biased user behaviour could be applied across Wikipedia to find likely sources of bias.

Another avenue of inquiry in this report's model is to investigate which users are rollbacked and by whom. To discern whatever it is the Casual contributors or the Experts who are rollbacked would prove insightful into the social dynamics of Wikipedia. Further questions go along the line are if experts to a greater degree are rollbacked by lurkers or by other experts who thus solve their differences internally.

6 Conclusion

This study has concluded that there were no significant differences in the distribution of users in regard to how many revisions they contribute between political and unpolitical articles. Both revision patterns seem to follow a general pyramid structure, with a majority of the users contributing an occasional revision and a few contributing many revisions. The categories differ however in regard to reverts, a sub group of revisions. The political articles across all categories had a more significant amount of medium frequent contributors joining the process, leading to a more even distribution. This is in contrast to the unpolitical articles where nearly half the reverts were done by the most frequent contributors (the Experts). The number of reverts as a percentage of all revisions were on average higher for political categories. This difference manifested early, before 100 revisions and either stayed level or increased as time passed. Due to the reverts either being used to undo vandalism or used for vandalism this indicates a higher level of controversy in political articles.

Bibliography

Keegan, B., Gergle, D., Contractor, N. (2013). Hot off the wiki: Structures and dynamics of Wikipedia's Coverage of Breaking News Events. *American Behavioral Scientist* 57(5) 595-622

Mattus, M. (2014). The Anyone-Can-Edit Syndrome: Interaction stories of three featured articles on Wikipedia. *Nordicom Review*, 35 (Special Issue): 189-203

Török, J., Iñiguez, G., Yasseri, T., Miguel, M., Kaski, K., Kertész, J. (2013). Opinions, Conflicts, and Consensus: Modeling Social Dynamics in a Collaborative Environment. *Physical Review Letters*, 110

Chen, J., Xu, H., Whinston, A. (2011). Moderated Online Communities and Quality of User-Generated Content, *Journal of Management Information Systems, 28(2), 237-268*

Dahlberg L., (2007). Rethinking the fragmentation of the cyberpublic: from consensus to contestation. *new media & society, Vol9(5):827–847*

Hansen, S., Berente, N., Lyytinen, K., (2007). Wikipedia as Rational Discourse: An Illustration of the Emancipatory Potential of Information Systems*. Proceedings of the 40th Hawaii International Conference on System Sciences*

Brandes, U., Lerner J. (2008). Visual analysis of controversy in user-generated encyclopedias. *Information Visualization vol 7, nr 1 34-48*

Niederer, S., van Dijk, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *new media & society, vol. 12(8) 1368-1387*

Malouf, R., Mullen, T,. (2008). Taking sides: user classification for informal online political discourse, *Internet Research, Vol. 18(2) 177 - 190*

O'Reilly, T. (2007). What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, *Communications & Strategies, Vol(1), 17*

Alexa, (2016). *How popular is wikipedia.org*. Available at: http://www.alexa.com/siteinfo/wikipedia.org [Accessed 11 May. 2016]

Oxford English Dictionary, (2016). *Politics*. [online] Available at: http://www.oed.com [Accessed 11 May 2016]

Wikimedia, (2016a). *API:Categorymembers*. [online] Available at:
https://www.mediawiki.org/wiki/API:Categorymembers
[Accessed 30 April 2016]

Wikimedia, (2016b). *Enwiki dump progress on 20160305*. [online] Available at:
http://dumps.wikimedia.org/enwiki/20160305/enwiki-20160305-stub-meta-history.xml.gz
[Accessed 2 April 2016]

Wikimedia, (2016c). *API:Random*. [online] Available at:
https://www.mediawiki.org/wiki/API:Random
[Accessed 4 April 2016]

Wikipedia, (2016a).*Wikipedia*. [online] Available at:
https://en.wikipedia.org/wiki/Wikipedia [Accessed 11 May 2016]

Wikipedia, (2016b). *Wiki.* [online] Available at:
https://en.wikipedia.org/wiki/Wiki [Accessed 11 May 2016]

Wikipedia, (2016c). *Wikipedia: List of policies and guidelines*. [online] Available at:
https://en.wikipedia.org/wiki/Wikipedia:List_of_policies_and_guidelines [Accessed 11 May 2016]

Wikipedia, (2016d). *Wikipedia:Bots*. [online] Available at:
https://en.wikipedia.org/wiki/Wikipedia:Bots [Accessed 11 May 2016]

Wikipedia, (2016e). *Wikipedia:Types of bots*. [online] Available at:
https://en.wikipedia.org/wiki/Wikipedia:Types_of_bots [Accessed 11 May 2016]

Appendix

Thesaurus

Edit	Changes the content on a Wikipedia article by adding and/or deleting content. Edits can be made by a registered user or through an IP-address. Also known as revision.
Revert	A subset of a revision where an article is 'rolled back' to a previous state. Any successive revisions by the latest contributor are undone.
Edit war	Intense edits and reverts due to a conflict of opinion
Web 2.0	A phenomenon where the content of a web page is created by the users themselves