# Gender Bias and Vandalism Study: Wikipedia

**Abu Saleh Md Noman**
School of Informatics and Computing
Indiana University Bloomington
amdnoman@indiana.edu

**Abdulwahab Alkharashi**
School of Informatics and Computing
Indiana University Bloomington
aalkhara@indiana.edu

*Abstract-* **Modern online communities are experiencing privacy concerns and gender bias in open source. A lot of online discussion communities nowadays offer users to hide their identity and interact. Such flexibility is understandable, however, they engender threats to the reputation and reliability in collective goods. In this study we are interested in studying one such open source online community: Wikipedia. Wikipedia is the largest open and free online encyclopaedia that is accessible to all. We are especially interested in finding the type of activity that the anonymous users are doing on Wikipedia articles and is there any significant gender bias. Since not a lot of previous work addressed these issues it is important to study the aforementioned issues to build an innate understanding of recent ongoing vandalism of Wikipedia pages and women leaving Wikipedia. The size of Wikipedia is vast (approximately 28,085,962 users, 5,135,196 English articles and 39,169,255 wiki pages) and that makes the study of Wikipedia more interesting. The study reveals ~90% of the vandalism or foul edits are done by unregistered users in Wikipedia thanks to the free and open nature of it. Preliminary studies also indicate women are less represented here and they are subjected more criticism compared to men. We hope the initial findings can engender further investigation and ways to uproot this problem.**

## I. INTRODUCTION

Wikipedia is the largest multilingual, web-based, free-content encyclopedia that is available in the internet. This project is supported by Wikimedia foundation and it is completely open source that means it is editable by anyone. Since its birth in 2001, it has experienced a massive growth in size and now it is invoking 374 million unique visitors monthly as of September 2015 [24]. About 70,000 active contributors are there working on more than 38 million articles in almost 292 languages. As of today there are 5,135,190 articles in English (compared to 3.9 million articles to nearest competitor Baike.com, Chinese wiki). Everyday there are tens of thousands of edits and visits to different Wikipedia pages help it grow and the gigantic size and diversity in Wikipedia data makes it particularly intriguing for researchers interested in mining patterns [1] and trends in the data. People from all backgrounds, cultures and ages can view and edit existing pages that makes the collaboration regardless of their qualification, it is the content not the quantity that matters. The reliability of the edits are reviewed by bots and experienced Wikipedians ensuring the quality of content that pertains.

However, the credence of such contribution are often questionable because of this open nature. While articles or pages evolve over the period of time and the collaboration making them more comprehensive, they are vulnerable to misinformation, errors, vandalism, foul edits, hacks etc.. Every Wikipedia article has an 'Edit' and 'View History' button (Figure 1(a)) that allows user to make necessary changes, summarize them and publish. They can make minor edits (spelling correction, grammatical changes) that require no further attention or they can make major edits which require further scrutiny by the community. Some pages are semi-protected indicated by the lock sign but still can be edited by placing an edit request. Depending on the access method the users can be classified into two categories:

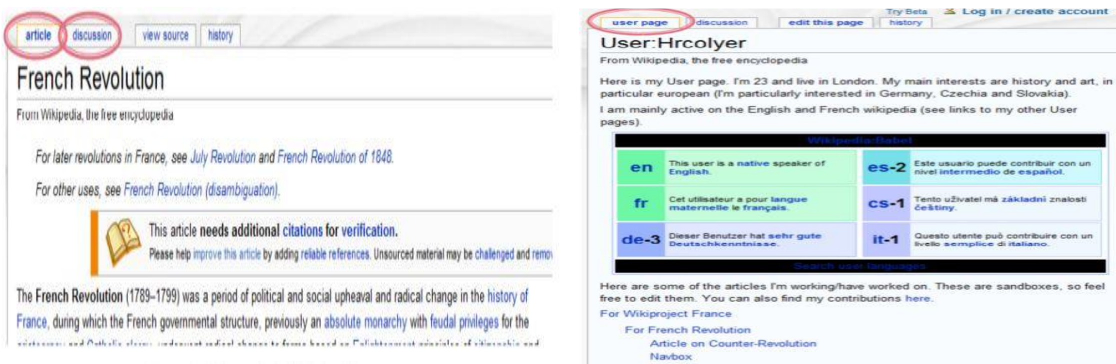



Figure 1. (a) A typical Wiki page (b) A typical user page (image taken from Wikipedia)

Registered user (those who have individual user page (Figure 1(b)) and account), unregistered user (do not have an account and identified by IP address). Both type of users can edit, talk and discuss before making edits. Anonymous users prefer to hide their identity and that engenders privacy issue under this study. Another intriguing challenge is the narrow diversity of contribution in Wikipedia in terms of gender. It is observed that only 13% of the Wikipedia contributors are female (2009 Wikimedia Foundation survey). This has led to inadequacy in 'female' oriented topics and women leaving Wikipedia. Wikimedia Foundation has targeted to improve the women contribution to 25%. To achieve this, we need to answer a critical question: if women are getting reverts in Wikipedia overall, compared to men? It has been observed that female Wikipedians are contributing significantly less. We need to investigate if their contribution are less reliable or poor quality or the community is particularly harsh community reaction. Bias?

This require exhaustive study of massive Wikipedia dataset to unearth the association of different kind of user with type of edits, quality of edits [2] and Wikipedia revision study for gender detection. Precisely, We want to analyze the edit history [5] of different Wikipedia articles and find out interesting facts that can be extracted from it i.e., types of editors/contributors, recognizing editor characteristics from edits they make [3,4], community reaction to female Wikipedians, coverage of female oriented topics etc.. Wikipedia is built on online collaboration and information sharing among users. Since it is free and open to all, there are myriads of talking, editing and corresponding changes which are going on every day. The main motivation behind this project idea is in recent years female Wikipedians are getting more reverts than male Wikipedians and it is observed that there are plenty of foul editing, vandalism that is taking place. To prevent this we need to identify what kind of editors are doing these and which types of articles are susceptible to these malicious activities. Besides, another important aspect of any social media study is gender issue. Wikipedia has long been identified as male dominant media. So this study will briefly address the fact that if there is really any gender bias in Wikipedia or is it just a random phenomenon? So in this research we have three research questions:

- What anonymous and registered people are doing (type of edits)?
- Are anonymous users causing more vandalism?
- Is there any apparent gender bias?

Previous works addressed different aspects of Wikipedia study on editor behavior, influences, personality traits etc. [6, 7, 8] but above mentioned questions were not well answered according to our knowledge.

***Research challenges.*** This work addressed various challenges for making the study possible. First, the dataset of Wikipedia dump is of titanic scale. The English wiki dump produces terabytes of edit history which was cumbersome to handle. Second, the results of analysis are often misleading and it is difficult to determine a comprehensive metric for the study. Third, due to the nature and size of data the presence of bias in study is difficult to catch.

***Contribution.*** The main contributions of this study are:

- **Detecting** type of edits by type of users.
- **Identifying vandals** and the reliability of their contribution.
- **Detecting** gender bias.
- **Feasibility study** of early prediction of vandalism on articles.

The remainder of this paper describe these contributions in detail. Section II deals with overview of our approach, section III describes the design of the study (data and methods), section IV reports the evaluation results, section V is devoted for discussion, section VI shows some related works on this issue followed by the conclusive remarks.

## II. OVERVIEW OF APPROACH

The purpose of this study is to identify anonymous and female user behavioral pattern in Wikipedia edits and corresponding feedback received. We can do so by formulating our problem space and goals again in a nutshell: Wikipedia is the best online encyclopedia available which is free and open to edit for all. It is very easy to make an edit on an existing article. Registered users as well as anonymous users (using their


Figure 2. (a) Anonymous user in shadow (b) Vandalism example (image source: https://en.Wikipedia.org/wiki/Wikipedia:IPs_are_human_too)

IP address, Figure 2(a)) can modify any article anytime (except few exceptions). Wikipedia stores all the edit history and the entire snapshot of the article is saved. Each article has a talk page and discussion page. People can talk and discuss before and after making an edit. Every registered editor has his own page where he can update his personal information and interest. Every user page also associate with a talk page which reflects the topics he is interested in. Using distance metric between two consecutive versions of articles, we can analyze how much changes are made, the type of edits and type of editors. Further digging into things might reveal the characteristics of the editor, clustering them into groups and/or take preventive measures against foul edits/vandalism (Figure 2(b)).

The approach adopted in this paper to answer our questions can be divided into following subcategories:

### *A. Measuring Community Reaction on Female Wikipedians*

Community feedback is of extreme importance in this study. We have to analyze if the community is particularly harsh, flexible or lenient towards female users. To measure this, we can count the number or percentage of times their edits are getting reverted. It might turn out that these edits are more reliable and are less likely to be reverted. So it is not hard to judge from this metric what is the ratio of reverts for each user type. A deeper analysis might reveal if there is any community bias towards the female users and if the tolerance of community towards female users edits are evolving over time.

### *B. Understanding Anonymous User Behavior*

In order to better understand the behaviors of anonymous users, it is important to analyze the percentage of articles where they are contributing. Later we focused on several questions and tried to answer them such as:

• How do anonymous users edit? For example, if they prefer to stay on topic or they stray. So we need to compare the similarity between the current and previous edits made by anonymous users. We also have to measure the type of edits that they are making such as if they contain negative emotions/words etc..

• How anonymous users react to community reaction? If they remain silent or attempt to make further edits and what is the consequence.

### *C. Predicting Edits by Female Users*

By training a Machine Learning classifier(bag-of-words strategy) with existing articles those are already edited by female users, we can predict if a future unseen article is going to be edited by an female user. This way, NLP techniques could help find if female contributors edit any particular type of articles and if any particular word/term in edits force reverts. This approach is still ongoing so it might not be reflected in the results or evaluation section for the time being.

## III. DATA AND METHODS

The data is collected from the huge data dumps those are already available in Wikimedia sites (https://dumps.wikimedia.org/enwiki/). The dumps for different language versions of Wikipedia are kept separate. It is important to note that the data are available in XML format (some previous version have SQL and HTML dumps too but are out of date) and needed to be transformed into more readable format. So I had to execute a Java tool that is already available called *mwdumper* that converts the XML to SQL. Some other tools such as mwdump.py, ImportDump.php, xml2sql (https://meta.wikimedia.org/wiki/Data_dumps/Tools_for_importing) are also available. But considering the huge size of English wiki dump, *mwdumper* is the best solution that generates the script without getting crashed. For data preprocessing, Wikipedia provides a set of tool such as DiffDB, MediaWiki API, Wiki

Edit History Analyzer, Wikiprep (http://wikipapers.referata.com/wiki/List_of_data_processing_tools) etc.. For this study *Wiki Edit History Analyzer* was used which processes MediaWiki revision history and produces summaries of edit actions performed. Basic edit actions include insert, delete, replace, and move; high-level edit actions include spelling correction, wikify, etc. WikiXray Python Parser is another tool written in Python programming language that helped me expedite some analysis. Data visualization is another important aspect of this study to get an insight of the edit history. Several interesting tools those were used in different part of the study include: HistoryFLow, Listen to Wikipedia, StatMediaWiki, Wiki Explorer (http://wikipapers.referata.com/wiki/List_of_visualization_tools) etc.. After initial analysis, data was loaded into a MySQL database and relevant results were extracted. Some recent dumps are also available in JSON format too. Loading JSON files in MongoDB and analyzing with pyMongo can greatly reduce the processing time. We are planning to continue our work using this latter technique.

There are three main reasons behind why Wikipedia dataset was chosen for this particular study: First, wiki dataset is publicly available in chunks and it is ideal for comprehensive longitudinal study. Second, Wikipedia data offers the diversity unlike anything else. All data are representative of sample drawn from diverse population. Third, the dataset is apparently harmless since it contains no personal information about user. For the sake of the study and brevity of analysis, the dataset used was 'enwikisource-20160305-pages-meta-history.xml' (https://dumps.wikimedia.org/enwiki/) combined with another dump of Bengali Wikipedia 'bnwikibooks-20160407-pages-meta-history.xml' from another timeframe. This was done to introduce more diversity in the data. The whole dataset was divided into 50 different tables under a predefined schema. The three main tables are given by:

- *user* – gives the names and total number of edits for each used in the dataset;
- *page* – provides information about the Wikipedia pages in the dataset;
- *revision* – saves the revision of each page with user id, comments and timestamp.

The remaining tables describe the page/user categories and their relations. Table pagecategory gives information about page categories. Similarly, table category provides information about user categories. Table 1 summarizes the dataset:

| Dataset | #Revisions | #Registered Users | #Anonymous Users |
|---|---|---|---|
| 20160305 | 158112 | 5741 | 1840 |
| 20160407 | 140399 | 3938 | 1467 |

Table 1: Dataset Summary

Some study were performed based on random articles using the 'Random article' option provided in the Wikipedia webpage navigation pane. For that study 150 random articles were chosen from wide range of topics. Later, the study extracted features from edits those were reverted eventually. A Machine Learning classifier is trained (bag-of-words strategy) and later tested to predict if an article is likely to get edited by anonymous users. It might be the case such edits can be reliable. But if not, then it can act as a precursor to preventing foul edits and vandalism. Since this study is incomplete it is not explained in detail. It is kept for future analysis.

The metric used for vandalism study is the percentage of posts marked as vandalism, therefore,

$$\frac{revisions\ marked\ as\ vandalism}{total\ number\ of\ revisions} * 100\%$$

Reliability of user edit is measured by retention rate of articles given by,

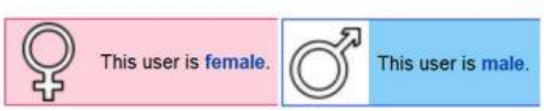
$$\frac{number\ of\ character\ retained}{total\ number\ of\ characters} * 100\%$$

On a different study, we analyzed if we can detect vandalism in Wikipedia edits using NLP text classification strategy. Let $E = \{e_1, \dots, e_n\}$ denote a set of edits, where each edit $e$ comprises two consecutive revisions of the same document d from Wikipedia, say, $e = (d_t, d_{t+1})$. Let $F = \{f_1, \dots, f_p\}$ denote a set of vandalism indicating features where each feature $f_i$ is a function that maps edits onto real numbers, $f_i : E \to R$. Using F an edit e is represented as a vector $e = (f_1(e), \dots, f_p(e))$; E is the set of edit representations for the edits in E. Given a vandalism corpus E which has a realistic ratio of edits classified as vandalism and well-intentioned edits, a classifier c, $c : E \to \{0, 1\}$, is trained with examples from E. c serves as an approximation of $c^*$, the true predictor of the fact whether or not an edit forms a vandalism case. Using F and c one can classify an edit e as vandalism by computing $c(e)$. For ML vandalism study I used the WEBIS-VC07-11 [33] corpus (940 human-assessed edits from which 301 edits are classified as vandalism).

***Gender Data***

There are several ways to get gender information of a wiki user:

- They can specify the gender in the account preference setting. It is accessible via Wikipedia API.
- They can place a gender userbox on their User page.



- From the discussion page description (It requires text processing).

For our analysis, we queried the API to get gender information and also cross-checked with US census dataset. We try to identify gender impact in terms of revering edits and posting in specific subject. A brief description of the approach is as follows:

*Data preprocessing*

The Wiki pages have a massive quantity of data dumps stored in the form of Wiki source and metadata embedded in XML. Wikipedia articles contain typically a free text, yet also contain different types of structured information, such as infobox templates, categorization information, images, geo-coordinates, links to external Web pages and links across different language editions of Wikipedia.

*Revision API*

This API offers revisions for a given page, or at latest revision for each of several pages. Pages are indicated either by PageID or titles parameter, whereas individual revisions are identified by RevID parameter. Each request used in this API is utilized by the PHP structure [29].

*Gender API*

Many commercial and non-commercial APIs offer gender deduction by given name. The commercial API mostly restrict requests with a certain number of daily limit, while the open source API has no limits. We extracted the data of top Wiki users by the number of edits and use python gender-guesser API [30] to detect the gender based on username. The API returns a gender value if known or uses a probabilistic model to determine whether a given name is mostly male or female. If the name is not listed within the database of names, it will return an unknown value. However, this approach is problematic in terms of detecting nicknames, which most users chose to not disclose their names or gender information.

*NLP supervised classification*

Classification [31] is the task of selecting the correct class label for a given input. It is built based on training corpus containing the correct label for each input. The classifier will determine the gender of names ending in a, e and i to be female, while names ending in k, o, r, s and t to be male. It will also provide an accuracy value of the predicted label name.

*Wiki Quarry*

This is a public querying interface that allows researchers to run SQL based queries, a set of live databases of public Wikimedia Wikis [32]. Quarry is designed to make running queries against database easy. Quarry also provides a means for researchers to share and review each other's queries. Users of Quarry are required to agree to Labs Terms of use.

## IV. RESULTS

This section is devoted to the findings of the analysis performed on the dataset mentioned above. The results are depicted using graphs and tables. The analysis can be grouped according to different dataset and corresponding research questions. So following subsections will try to elucidate the results:

### *A. Type of Edit*

At first I analyzed the type of article both type of users are targeting. The result is depicted in Figure 3.

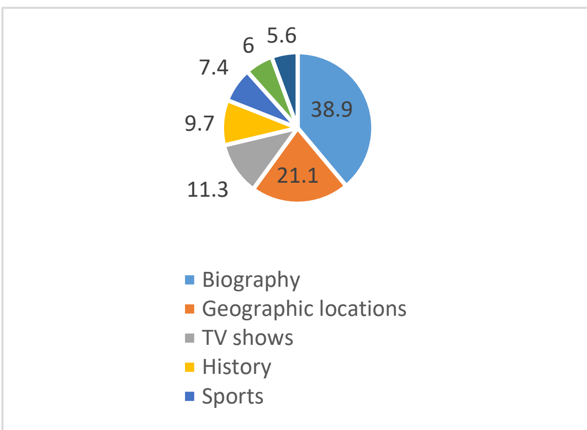



Figure 3. Type of articles

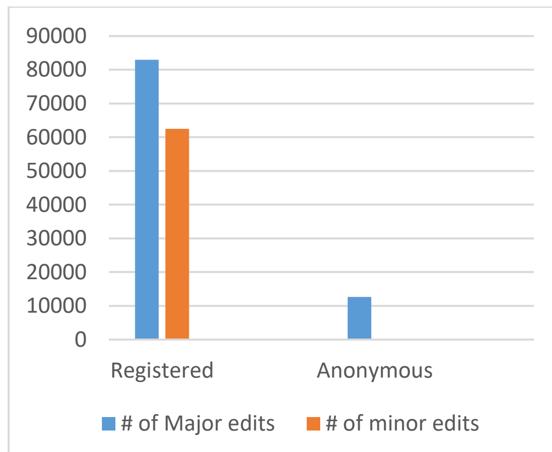



Figure 4. Type of edits

It turns out that majority of the articles targeted by registered users are related to Biography (38.9%), followed by Geographic location (21.1%), TV shows, History, Sports, Politics etc.. Intuitively the targeted articles for unregistered users turn out to be the same too with somewhat different percentage. The category of edits by both types of users are reported in Table II:

TABLE II

| 20160407 dataset | # of Major edits | # of minor edits |
|---|---|---|
| Registered | 83000 | 62475 |
| Anonymous | 12672 | 1 |

It clearly shows number of edits done by registered users are dominating than that of the IP users and mostly the type of edits are major edits. The number of edits per year by each user type was not surprising as well (Figure 5): registered users dominated unregistered users here too.

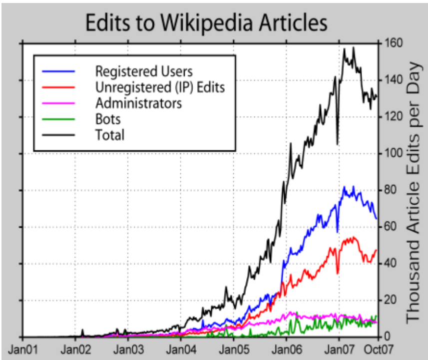



Figure 5. Number of edits by users

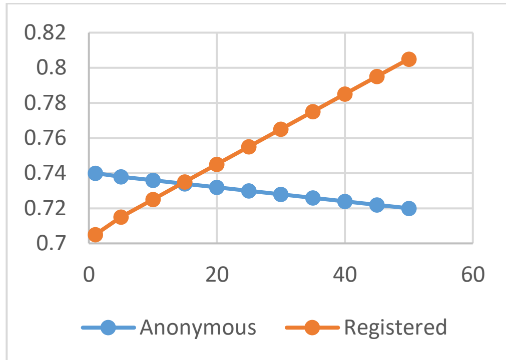



Figure 6. Reliability by user

Figure 6 shows the reliability of edits by user type in terms of retention rate defined previously. It shows that when number of edits are getting higher, the reliability of anonymous users drop while the reliability of registered user increase. This result is congruous to the study presented in [3].

## *B. Vandalism Study*

The direction of the research shifted towards vandalism study at this point because it was one of the most important research questions. Vandalisms are often caused by lack of knowledge, attention seeking attitude, personal grudge etc.. It is important to understand not every malicious act is considered as vandalism. Things like abuse of tags, illegitimate page creation, spam external linking, trolling etc. are considered few of different types of vandalisms. The vandalism study was performed on different dataset: for randomly selected articles and querying the API.

For randomly selected articles the manual analysis yielded following results:

TABLE III

| Year | #of Articles | # of Edits | # of Vandalism | #Caused By Anonymous Users |
|---|---|---|---|---|
| 2010 | 40 | 450 | 34 | 25 |
| 2011 | 50 | 397 | 29 | 18 |
| 2012 | 35 | 869 | 59 | 45 |

On an average, ~90% of times the vandalism are caused by anonymous users as predicted. However, study on user pages yielded interesting result. Out of 10 randomly generated user page, the ratio (% of vandalism done by registered to anonymous) returned was 53:47. This might be indicative of the fact that, anonymous user tend to target main article pages while registered users are main culprit for vandalisms in user pages. The analysis on full dump querying the revision API yielded the following:

| Dataset | # of Edits | # of Vandalism | #CausedBy Anonymous Users |
|---|---|---|---|
| 20160407 | 140399 | 9567 | 7653 (~80%) |
| 20160305 | 158112 | 13345 | 9745 (~78%) |

The chi-squared test yielded:

| | Vandalism | Non-Vandalism | |
|---|---|---|---|
| **Registered** | 1914<br>*8013.12*<br>(4642.29) | 105018<br>*98918.88*<br>( 376.06) | **106932** |
| **Anonymous** | 7653<br>*1553.88*<br>(23939.50) | 13083<br>*19182.12*<br>(1939.27) | **20736** |
| | **9567** | **118101** | **127668** |

$\chi^2$ = 30897.111, df = 1, $\chi^2$/df = 30897.11 , P($\chi^2$ > 30897.111) = 0.0000

expected values are displayed in *italics*
individual $\chi^2$values are displayed in (parentheses)

For 1 degree of freedom the critical value of chi-squared statistics is 3.84. Since $\chi^2$>3.84, we can reject the Null hypothesis that they are unrelated. So vandalism is dependent on user type at 5% level of significance.

Findings of the automatic vandalism detection:

We used the following featureset upon close examination of the datset and consulting prior work [33]:

| Feature *f* | Description |
|---|---|
| char distribution | deviation of the edit's character distribution from the expectation |
| char sequence | longest consecutive sequence of the same character in an edit |
| compressibility | compression rate of an edit's text |
| upper case ratio | ratio of upper case letters to all letters of an edit's text |
| term frequency | average relative frequency of an edit's words in the new revision |
| longest word | length of the longest word |
| pronoun frequency | number of pronouns relative to the number of an edit's words (only first-person and second-person pronouns are considered) |
| pronoun impact | percentage by which an edit's pronouns increase the number of pronouns in the new revision |
| vulgarism frequency | number of vulgar words relative to the number of an edit's words |
| vulgarism impact | percentage by which an edit's vulgar words increase the number of vulgar words in the new revision |
| size ratio | the size of the new version compared to the size of the old one |
| replacement similarity | similarity of deleted text to the text inserted in exchange |
| context relation | similarity of the new version to Wikipedia articles found for keywords extracted from the inserted text |
| anonymity | whether an edit was submitted anonymously, or not |
| comment length | the character length of the comment supplied with an edit |
| edits per user | number of previously submitted edits from the same editor or IP |

Using 10 fold cross validation and Logistic regression and Random forrest classifier the findings are listed below:

| **Algorithm** | **Recall** | **Precision** |
|---|---|---|
| Baseline | 0.61 | 0.74 |
| LR | 0.75 | 0.77 |
| RF | 0.71 | 0.73 |

As we can see our approach outperforms slightly the existing baseline classifier. But incorporating more training data and more nuanced understanding as well as feature engineering can improve it further.

*C. Gender Bias*

The number of edits since Wikipedia has started is 415,826,115 edits, and 2,759,335 registered users. Only 16,903 (3,003 females and 13,900 males) who disclose their gender information. It is apparent in the results that female editors tend to contribute less. I took a sample number of users to test tje experiment from top 5000 Wikipedians by number of edits, the number of females is still significantly less than males for all edits (as shown in Fig.7). We can see that only 16% are female editors and 84 % male editors.

The women scientists Wiki project (Wikipedia:WikiProject Women scientists. (n.d.). Retrieved December 13, 2016, from https://en.Wikipedia.org/wiki/Wikipedia:WikiProject_Women_scientists) is an example of how would females increase gender diversity among the community. We ran a query test to see edits per month since the beginning of Wikipedia. As we can see in Fig. 8, that the number of edits

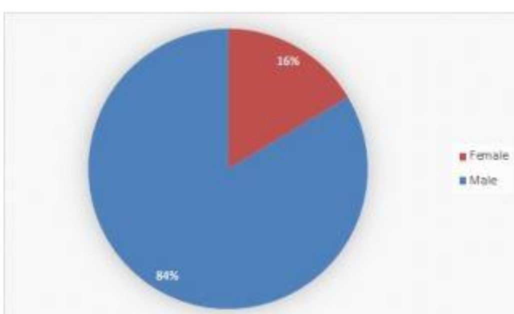



Figure 7. Top Wikipedians by gender

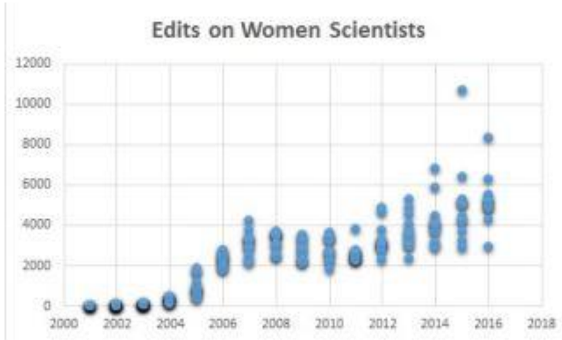



Figure 8. Edits per month

on this particular scenario. This could lead to the fact that feminism related subjects may encourage females to be inclusive in the community and contribute more in the future. The level of mutual understanding what each gender like to post about seem to be very ambiguous. Yet, some Wikipedians particularly females shared their expression stating that the fact that their edit is externally insulting and unpleasant. Although we run a query to see revert history among topics, it is unclear what exactly cause the reversion of an edit.

*Manual Analysis*

On a separate analysis, randomly 50 articles were picked. Then each article were manually checked: out of total 1137 edits 195 were made by female users (verified gender from their account page). The registered users are considered for this study presumably. Of these edits, 45 got reverted eventually. Later, 10 female Wikipedians were chosen randomly. Following similar strategy, it was observed that they make almost 50% less contribution than 10 randomly chosen male editors. 29% of these female editor's edits got reverted compared to 19% of that of the male editors.

*Full Dump Analysis*

To further investigate the scenario, 2010 January English Wiki dump was loaded into database. 112,778 total users made 65m+ edits. The usernames were cross checked with US census dataset. The census dataset contains individual names associated with a probability of him/her being male/female. There was a 13% match (maybe because most users do not use their username). Of them 16% were recognized as female. Consulting the database yielded 33% of female editors edits getting reverted compared to 23% of males.

*Analysis based on API*

Wikipedia allows to access it programmatically via API query and it can be analyzed using any scripting language. We did our analysis using python. We randomly picked 20 articles and analyzed its revision history by API:Revisions and filtered out the editors name. Then we queried the API: Users to find the gender of the user from their account settings. It turned out only 19% of the total edits were done by women who reported their gender. And out of total reverts, 24% got reverted compared to 18% of males.

To check if the reversion rate depends on gender we performed a Chi-squared test which yielded the following result:

| | **Revert** | **Not Revert** | |
|---|---|---|---|
| **Male** | 23052<br>*23344.29*<br>( 3.66) | 105018<br>*104725.71*<br>( 0.82) | **128070** |
| **Female** | 7209<br>*6916.71*<br>( 12.35) | 30737<br>*31029.29*<br>( 2.75) | **37946** |
| | **30261** | **135755** | **166016** |

$\chi^2$ = 19.581, df = 1, $\chi^2$/df = 19.58 , P($\chi^2$ > 19.581) = 0.0000

expected values are displayed in *italics*
individual $\chi^2$values are displayed in (parentheses)

For 1 degree of freedom the critical value of chi-squared statistics is 3.84. Since 19.58>3.84, we can reject the Null hypothesis that they are unrelated. So reversion is dependent on gender at 5% level of significance.

Reversion rate in early stage of tenure:

| **Edit #** | **Male** | **Female** |
|---|---|---|
| | | |
| 1 | 5.2% | 7.1% |
| 2-5 | 4.45% | 5.88% |
| 6-9 | 3.34% | 4.56% |

We are trying to analyze if this has any co relation with female contributors stopping edits and leaving Wiki. We are still trying to develop an understanding of measuring the edit quality. Then we will be able to determine if women really make low quality edits or they fall victim to gender bias.

## V. DISCUSSION

The results of this study are fairly conclusive. Registered users, as expected, account for most edits while anonymous users cause most vandalism. It turns out not all anonymous users are vandals though. Another important finding was reliability of edits made by these users. It varied gradually over a period of time and proved that anonymous users with less number of edits are in fact more reliable than registered users. This could attributed to the fact that such unregistered users are experts on specific fields and do not bother about reputation in such community.

There is a significant gender skew in Wikipedia in terms of gender too. The statistical significance of the analysis was not tested though. It will be really interesting to see if there really is a bias against female editors or if their edit quality is not up to mark and hence get reverted. Also, advanced NLP techniques could help find if female contributors edit any particular type of articles and if any particular word/term in edits force reverts.

Although our analysis gives us a hint about gender bias in Wikipedia it is not comprehensive since the analysis was done on small scale. A few other questions like if the gender bias is shrinking, what is women editors motivation and what type edits they are making and what kind of changes would they like on Wikipedia has not been answered. But this can be a cornerstone to revealing the reasons behind community being harsh on women and eventually their exit from Wikipedia. The study of Wikipedia data was challenging yet fruitful but it suffered from some limitations such as: First, due to huge nature of dataset exhaustive study could not be performed. So the restriction on dataset might introduce bias in the study. Second, the selection of random articles were made under the assumption that the randomness of drawing the sample is purely random without any guarantee. Third, there was no statistical test performed to test the hypotheses and approximate a confidence level. Fourth, the gender information is purely self-reported. There is no way of verifying what the user's actual gender is. The analysis is also based on the assumption that the non-reported user group behave the same way as the reported group.

## vi. RELATED WORK

### A. *Wikipedia Vandalism Study:*

Adler et. al [18] presented the results of an effort to integrate three of the leading approaches to Wikipedia vandalism detection: a spatio-temporal analysis of metadata (STiki), a reputation-based system (WikiTrust), and natural language processing features. They examined in detail the contribution of the three approaches, both for the task of discovering fresh vandalism, and for the task of locating vandalism in the complete set of Wikipedia revisions. Potthas [19] presented results of a new approach to detect destructive article revisions, so-called vandalism, in Wikipedia. They discussed the characteristics of vandalism as humans recognize it and develop features to render vandalism detection as a machine learning task.

Adler et. al [20] presented using the full set of features computed by WikiTrust, they have been able to construct classifiers that identify vandalism with a recall of 83.5%, a precision of 48.5%, and a false positive rate of 8%, for an area under the ROC curve of 93.4. Using these classifiers, they have implemented a simple Web API that provides the vandalism estimate for every revision of the English

Wikipedia. A statistical language model, constructing distributions of words from the revision history of Wikipedia articles was presented in [21]. As vandalism often involves the use of unexpected words to draw attention, the fitness of a new edit when compared with language models built from previous versions may well indicate that an edit is a vandalism instance.

### *B. Wikipedia Gender Effect*

The authors of [26] addressed the gender inequality issue of Wikipedia and showed that Wikipedia has a substantial editor gender gap and it is shrinking. Their study also revealed some other facts such as: females are more likely to be involved in social and community-oriented areas, females tend to avoid controversial topics, they are more likely to get more reverts in early tenure, they are less likely to get blocked etc. Another study [27] revealed that, based on some explicit assumptions, the proportion of female US adult editors was 27.5% higher than the WMF/UNU-MERIT survey reported (22.7%, versus 17.8%), and that the total proportion of female editors was 26.8% higher (16.1%, versus 12.7%). (Reagle & Rhue, 2011) found evidence of gender bias in Wikipedia coverage of biographies. Wagner et. al [28] adopted computational method to assess gender bias on Wikipedia and concluded that while women on Wikipedia are covered and featured well in many Wikipedia language editions, the way women are portrayed starkly differs from the way men are portrayed.

## vii. CONCLUSION

Wikipedia's free and open nature has given it the utmost popularity. In this study I did not want prove this concept wrong but to make Wikipedia a better place for Wikipedians it is necessary to understand what different types of users are doing and the fellow community is reacting to that. In this study it was evident that unregistered users mostly cause the vandalisms and there seem to have significant gender bias. This study can be a first step to solving existing issues with gender skew and ways to addressing them. This study engendered a lot of new horizon yet to be explored; few of these include:

- Content and quality of women editors edits.
- If high reversion rate in early stage of tenure is causing female Wikipedians to stop editing and leave Wiki.
- Classifying the vandals to detect the vandalism ahead of time.
- Other privacy issues in Wikipedia i.e. leak user private information from user page and factors influencing privacy loss.

## REFERENCES


[1] Maass, D. (2013). Data Mining Revision Controlled Document History Metadata for Automatic Classification.

[2] Daxenberger, J., & Gurevych, I. (2013, October). Automatically Classifying Edit Categories in Wikipedia Revisions. In*EMNLP*(pp. 578-589).

[3] Anthony, D., Smith, S. W., & Williamson, T. (2009). Reputation and reliability in collective goods the case of the online encyclopedia Wikipedia.*Rationality and Society*,*21*(3), 283-306.

[4] Lieberman, M. D., & Lin, J. (2009, May). You Are Where You Edit: Locating Wikipedia Contributors through Edit Histories. In*ICWSM*.

[5] Wierzbicki, A., Turek, P., & Nielek, R. (2010, July). Learning about team collaboration from Wikipedia edit history. In*Proceedings of the 6th International Symposium on Wikis and Open Collaboration*(p. 27). ACM.

[6] Tsikerdekis, M. (2013). The effects of perceived anonymity and anonymity states on conformity and groupthink in online communities: A Wikipedia study. *Journal of the American Society for Information Science and Technology*, *64*(5), 1001-1015.

[7] Panciera, K., Halfaker, A., & Terveen, L. (2009, May). Wikipedians are born, not made: a study of power editors on Wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 51-60). ACM.

[8] Amichai-Hamburger, Y., Lamdan, N., Madiel, R., & Hayat, T. (2008). Personality characteristics of Wikipedia members. *CyberPsychology & Behavior*, *11*(6), 679-681

[9] Lucas, M. M., & Borisov, N. (2008, October). Flybynight: mitigating the privacy risks of social networking. In *Proceedings of the 7th ACM workshop on Privacy in the electronic society* (pp. 1-8). ACM.

[10] Fang, L., & LeFevre, K. (2010, April). Privacy wizards for social networking sites. In *Proceedings of the 19th international conference on World wide web* (pp. 351-360). ACM.

[11] Madden, M. (2012). Privacy management on social media sites. *Pew Internet Report*, 1-20.

[12] Narayanan, A., & Shmatikov, V. (2009, May). De-anonymizing social networks. In *Security and Privacy, 2009 30th IEEE Symposium on* (pp. 173-187). IEEE.

[13] Szongott, C., Henne, B., & von Voigt, G. (2012, June). Big data privacy issues in public social media. In *Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference on* (pp. 1-6). IEEE.

[14] Tsikerdekis, M. (2013). The effects of perceived anonymity and anonymity states on conformity and groupthink in online communities: A Wikipedia study. *Journal of the American Society for Information Science and Technology*, *64*(5), 1001-1015.

[15] Leskovec, J., Huttenlocher, D. P., & Kleinberg, J. M. (2010, April). Governance in social media: A case study of the Wikipedia promotion process. In *ICWSM*.

[16] Rad, H. S., & Barbosa, D. (2012, August). Identifying controversial articles in Wikipedia: A comparative study. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration* (p. 7). ACM.

[17] Kittur, A., Chi, E., Pendleton, B. A., Suh, B., & Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World wide web*, *1*(2), 19.

[18] Adler, B. T., De Alfaro, L., Mola-Velasco, S. M., Rosso, P., & West, A. G. (2011). Wikipedia vandalism detection: Combining natural language, metadata, and reputation features. In *Computational linguistics and intelligent text processing* (pp. 277-288). Springer Berlin Heidelberg.

[19] Potthast, M., Stein, B., & Gerling, R. (2008). Automatic vandalism detection in Wikipedia. In *Advances in Information Retrieval* (pp. 663-668). Springer Berlin Heidelberg.

[20] Adler, B., De Alfaro, L., & Pye, I. (2010). Detecting Wikipedia vandalism using wikitrust. *Notebook papers of CLEF*, *1*, 22-23.

[21] Chin, S. C., Street, W. N., Srinivasan, P., & Eichmann, D. (2010, April). Detecting Wikipedia vandalism with active learning and statistical language models. In *Proceedings of the 4th workshop on Information credibility* (pp. 3-10). ACM.

[22] Harpalani, M., Hart, M., Singh, S., Johnson, R., & Choi, Y. (2011, June). Language of vandalism: Improving Wikipedia vandalism detection via stylometric analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2* (pp. 83-88). Association for Computational Linguistics.

[23] Rizoiu, M. A., Xie, L., Caetano, T., & Cebrian, M. (2015). Evolution of Privacy Loss in Wikipedia. *arXiv preprint arXiv:1512.03523*. Computational Linguistics.

[24] https://en.Wikipedia.org/wiki/Wikipedia:About

[25] https://en.Wikipedia.org/wiki/Wikipedia:IPs_are_human_too

[26] Lam, S. T. K., Uduwage, A., Dong, Z., Sen, S., Musicant, D. R., Terveen, L., & Riedl, J. (2011, October). WP: clubhouse?: an exploration of Wikipedia's gender imbalance. In *Proceedings of the 7th international symposium on Wikis and open collaboration* (pp. 1-10). ACM.

[27] Hill, B. M., & Shaw, A. (2013). The Wikipedia gender gap revisited: characterizing survey response bias with propensity score estimation. *PloS one*, *8*(6), e65782.

[28] Wagner, C., Garcia, D., Jadidi, M., & Strohmaier, M. (2015). It's a man's Wikipedia? assessing gender inequality in an online encyclopedia. *arXiv preprint arXiv:1501.06307*.

[29] Auer, Sren, Christian Bizer, Geori Kobilarov, Jens Lehmann, Richard Cyganiak,

and Zachary Ives. "Dbpedia: A nucleus for a web of open data." In The semantic
web, pp. 722-735. Springer Berlin Heidelberg, 2007.
[30] https://pypi.python.org/pypi/gender-guesser/
[31] Manning, Christopher D., and Hinrich Schtze. Foundations of statistical naturallanguage processing. Vol. 999. Cambridge: MIT press, 1999.
[32] https://meta.wikimedia.org/wiki/Research:Quarry
[33] Potthast, Martin, Benno Stein, and Robert Gerling. "Automatic vandalism detection in Wikipedia." *European conference on information retrieval*. Springer, Berlin, Heidelberg, 2008.