Dynamics of Edit War Sequences in Wikipedia

Anamika Chhabra
Indian Institute of Technology Ropar, Rupnagar, India
anamika.chhabra@iitrpr.ac.in

Rishemjit Kaur
CSIR-Central Scientific Instruments Organization, Chandigarh, India
rishemjit.kaur@csio.res.in

S. R. S. Iyengar
Indian Institute of Technology Ropar, Rupnagar, India
sudarshan@iitrpr.ac.in

ABSTRACT

In any collaborative system, cooperation and conflicts exist together. While in some cases these conflicts improve the output, they also lead to increased overhead. This requires examining the dynamics of these conflicts with the help of underlying data. In Wikipedia articles, the conflicts are captured by edit wars which may be examined through the revision history of these articles. In this work, we perform a systematic analysis of the conflicts present in 1,208 controversial articles of Wikipedia captured in the form of edit war sequences. We examine various key characteristics of these sequences and further use them to estimate the outcome of the edit wars. The study indicates the possibility of devising automated coordination mechanisms for handling conflicts in collaborative spaces.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; *Collaborative and social computing design and evaluation methods.*

KEYWORDS

Edit wars, Reverts, Wikipedia, Coordination, Conflicts

ACM Reference Format:
Anamika Chhabra, Rishemjit Kaur, and S. R. S. Iyengar. 2020. Dynamics of Edit War Sequences in Wikipedia. In *16th International Symposium on Open Collaboration (OpenSym 2020), August 25–27, 2020, Virtual conference, Spain.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3412569.3412585

1 INTRODUCTION

The open nature of Wikipedia invites people with diverse demographics to contribute [18, 37]. Generally, diversity of any form has been considered to improve collaborative outcomes and produce content of better quality [6, 19, 23, 25, 31] as it brings in different perspectives and diverse behaviors [7, 8, 10]. A variety of perspectives further instigate more collective outcome as a result of the contributors' interaction [9]. However, diverse perspectives also lead to differences in opinions and a lack of consensus among the contributors [20, 21, 24]. In Wikipedia articles, many a time, contributors bring in contradicting content, which leads to conflicts among them over what should stay in the articles [4]. As the conflicts and disagreements among the participants escalate, complex power plays also happen, further impeding collective resolutions [22]. The sequences of back and forth undoing of each others' contributions during the edits are termed as *Edit Wars*[1] in which contributors try to effectuate their preferred versions of a part of the article. The edit wars reflect the amount of conflict prevalent amongst the participating users and their inability to come out of what is called their *ego chambers* [33]. In most cases, they also defy one of the important policies of Wikipedia requiring contributors to maintain a *Neutral Point of View* (NPOV). These edit wars hinder the overall performance of the system [2] and result in the collaboration being counter-productive [16].

Handling conflicts arising out of differences in perspectives is a persistent concern for collaborative systems [22]. In Wikipedia articles, they have been observed to result in huge coordination costs as they consume considerable amounts of editorial resources [32, 38]. An increase in the number of administrative pages as compared to the article pages further suggests the growing demand for coordination [36]. Additionally, it has been observed by past work that around 40% of all edits in Wikipedia articles involve activities such as policy-building, coordination, consensus-forming etc, which is a substantial amount of indirect work [21]. Due to this, the growth of 'actual knowledge' in the articles is reducing and the amount of time spent in coordination and conflict resolution has been rising. Studies on Wikipedia report that other than diverse perspectives, vandalism is the other main cause of such conflicts, which involves deliberate removal of content or the insertion of false or irrelevant information [35]. The conflicts arising out of vandalism may have far-reaching consequences for the editors as well as the readers as the article stays in an unstable state during these conflicts. The effect is even worse in the case of biographical articles [26].

It takes time and effort to handle edit wars in Wikipedia. In many cases, as we will see, each of these back and forth fights span hundreds of revisions, wasting the time of the administrators who are forced to intervene and handle them. While much work has focused on identifying and visualizing edit wars in Wikipedia [4, 11, 14, 30, 34, 38] there is little prior work on examining such conflicts at a fine-grained level. Alternatively, the existing work is mostly performed at the article-level, while lacking an analysis of individual conflicts. An investigation of the temporal sequences constituting the edit wars may provide actionable insights into devising automatic means to handle such conflicts. The objective of this study, therefore, is to examine the characteristics of the edit war sequences and investigating how they fare with time based on these characteristics.

[1] https://en.wikipedia.org/wiki/Wikipedia:Edit_warring

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
OpenSym 2020, August 25–27, 2020, Virtual conference, Spain
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8779-8/20/08...$15.00
https://doi.org/10.1145/3412569.3412585

It is possible to examine the crucial details of the conflicting content resulting in these wars along with the users involved in them by analyzing the detailed revision history of the articles. These details may help in understanding their dynamics such that they may be handled better and faster. As instance, timely notification of the possible escalation of conflicts may enable prompt measures. Therefore, in this work, we examine the sequences of edit wars prevalent in the most controversial articles of Wikipedia and understand various properties of edit wars, involving the kind of users, polarity and subjectivity of the content etc. Our analysis reveals certain parameters that inform how the fight taking place in an edit war will finally take effect, i.e., which side of the users will finally have their content stay. We examine in detail what properties lead to the initiation of edit wars and what are the signatures of those edit war sequences that finally get nullified, i.e., the sequences where the content of the article remains unchanged before they start and after they stop. Known beforehand, these details may help in controlling fights before they lead to wastage of resources and time.

The results of the study may help in devising more robust Wikipedia policies and enabling the design of automated processes to combat unnecessary conflicts in Wikipedia articles. We propose the usage of automated means for estimating the details of edit wars concerning how they are expected to settle. Such measures may help in controlling long edit wars in time thus reducing human effort for coordination.

2 RELATED WORK

There are certain aspects of conflicts in Wikipedia that have been explored by the past work. Broadly, these works include computing the level of conflict in an article [38], examining different structures of conflicts [34] and building probabilistic models explaining the dynamics of conflicts in Wikipedia articles [13, 17, 28].

Studies have been conducted to automatically examine the controversiality score of the articles [38] and to identify the controversial articles from a given set [30]. Some of these studies involve creating revert maps of the articles where they find them to be significantly different for disputed and non-disputed articles. Chin et al. [11] used content-based classifiers using the data generated from statistical language models applied to the edits to find potential vandalism instances in Wikipedia articles.

Additionally, several agent-based models have been developed to explain the emergence and resolution of conflicts in Wikipedia [28, 29] and social networks [1, 12]. Some of these models are based on *Bounded Confidence model* from opinion dynamics domain where two agents will tend to converge if the difference in their opinion is less than a *tolerance* (agent-agent dynamics) and an agent will make changes in the medium if the difference in his opinion and the medium's opinion is more than a *threshold*, i.e. *inverse bounded confidence model* (agent-medium dynamics).

The network structure of edits in Wikipedia articles has also been examined to derive details about conflicts among users [5, 34]. Through a study of the temporal motifs formed out of the reverts prevalent in the articles, it has been examined that reverters generally have higher status than reverted editors, where the status is proportional to the number of edits the editor has made by the time of the revert under question.

Work has also been done towards visualizing the conflicts present in the articles by examining the reverts, which enables finding the patterns of conflicts in the articles. Viegas et al. [35] built one of the initial tools named *History Flow Visualization* to visualize the articles' revisions. This tool represented each revision of an article as a vertical line with different sections of each revision line colored according to the editor who authored those parts. The edit wars in such a setting were thus visible as zig-zag lines. Suh et al. [32] built *Revert graphs* and used the force-directed layout to visualize the conflicting social structures in the articles. Borra et al. [3, 4] built a platform named *Contropedia* for analysis and visualization of controversies in Wikipedia articles.

Additionally, past work also studies the effect of reverting or banning the editors on knowledge development in the articles. It is observed that banning is destructive and slows down the editing and hence the consensus-reaching process [28]. The effect of reverting the newcomers has also been observed where it is found that reverts have a negative effect on new-comers in terms of reduced co-ordination and overall withdrawal from Wikipedia especially when they are reverted by highly experienced users [15]. However, it was also observed that the quality of work by those new-comers who continue to stay despite being reverted, relatively increases.

The work done so far examines the articles at a high-level to examine, for instance, how controversial they are and how the same may be visualized. Our work rather focuses on the individual instances of conflicts captured in the form of edit war sequences. The study aims to derive low-level signatures of such conflicts that may be useful in improving conflict-management measures.

3 DATA SET AND PRELIMINARY STATISTICS

Wikipedia maintains a list of articles called *Controversial Articles*[2], that are named so as their content is repeatedly reverted or they are the focus of *Article Sanctions*[3] due to their imbalanced nature. They also have high chances of inclusion of content that is not NPOV[4]. We use controversial articles of Wikipedia for this analysis as they are likely to contain more cases of edit wars as compared to non-controversial articles. At the time of this analysis, there were 1,208 articles. We downloaded the detailed revision histories of these articles using KDAP library[5] which provides tools to extract and analyze the data of Wikipedia articles.

To examine the edit war sequences in the articles, we first collected all the reverts in the articles, as the edit wars constitute consecutive reverts.

3.1 Reverts

Revert is a special revision that undoes the effect of one or more edits, thereby, restoring the article content to a state that it was a few revisions ago. Figure 1 shows the schematic diagram of a revert. Here, user U_n reverts, i.e., undoes the changes made by users $U_m ... U_{n-1}$. It should be noted that the content of the article at

[2] https://en.wikipedia.org/wiki/Wikipedia:List_of_controversial_issues
[3] https://en.wikipedia.org/wiki/Wikipedia:General_sanctions#Active_sanctions
[4] https://en.wikipedia.org/wiki/Wikipedia:Neutral_point_of_view
[5] www.github.com/descentis/kdap

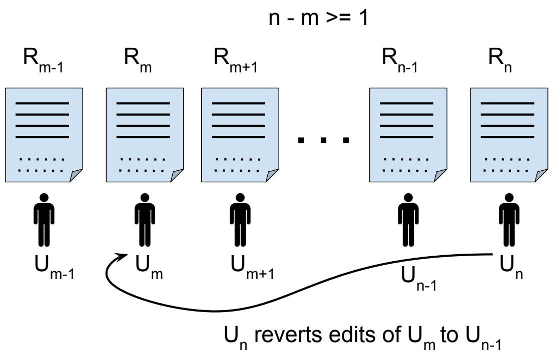
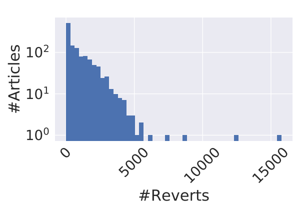


Figure 1: Schematic diagram of reverting in Wikipedia articles

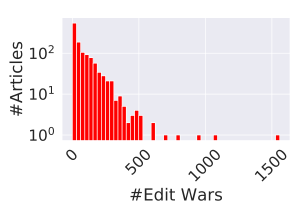
revision R_n exactly matches with the content at revision appearing immediately before the last user that got reverted, which is R_{m-1} in this case.

We collected all the reverts taking place in the articles using MD5 checksum [27, 32] on the revisions' content. It provides a unique hash value for an input string based on its content. Using these hash values, for each revision, we found which previous revision/s its hash value exactly matched, thus finding the reverts that took place in each article. In the example shown in Figure 1, the checksum value of the revision R_n exactly matched with that of revision R_{m-1}. In the subsequent text, we refer to a user who reverts as 'reverter' and the user whose edit gets reverted as the 'reverted user'.

3.1.1 Distribution of Reverts. The number of reverts in the articles ranged from 0 to 15,779 (for 'George W. Bush'). The article on 'Wikipedia' was found to be having 12,380 reverts. Despite the articles being marked as controversial, a total of 119 articles did not have any reverts. This may be due of other reasons such as article sanctions etc that are responsible for an article being put in this list. We, therefore, continued our analysis on the remaining 1,089 articles. Figure 2(a) shows the distribution of articles as per the number of reverts present in them.



(a)



(b)

Figure 2: Distribution of articles with respect to (a) the number of reverts (b) the number of edit wars (Y-axis is log-scaled.)

3.1.2 Anonymous and Registered users: We examined the distribution of anonymous and registered reverters and reverted users across all instances of reverts. It was found that a good proportion of reverts were made by registered users to undo the changes made by anonymous users (i.e. 57.16%) (See Table 3.1.2). On the other hand, there was a relatively small fraction of cases where the reverter was anonymous. A total of 34.29% of the cases involved reverts between registered users.

		Reverted User	
		Registered	Anonymous
Reverter	Registered	34.29%	57.16%
	Anonymous	3.25%	5.28%

Table 1: Proportion of registered and anonymous reverters and reverted users in the reverts

3.1.3 Number of Revisions Reverted. It was observed that in 79.58% cases, the content of one revision (i.e. immediate previous) was reverted, while in the rest, more than one revision's changes were reverted. Further, in the latter bunch of reverts, the number of revisions reverted ranged from 2 to as high as 30,957. In particular, in 1.63% of the cases, the number of revisions reverted was found to be more than 10. Clearly, the reverts involving a large number of revisions indicate cases of vandalism.

3.1.4 Reverting to Own or Others' Content. There were 16.56% cases of reverting to their own content, while 83.43% of the cases of reverts involved reverting to other users' content, indicating the presence of multiple users forming groups fighting for the common content.

3.2 Edit Wars

An edit war is a sequence of reverts where the editors try to override each other's contributions in succession. As an example, *A* <- *B* <- *A* which is the simplest edit war sequence, represents that a change that *A* made to the article, was reverted by *B*, which was followed by *A* reverting the change made by *B*. We refer to the length of an edit war to be the number of reverts taking place in the edit war sequence. The length of the sequence *A* <- *B* <- *A* would thus be 2.

By examining the reverts, we collected the details of all edit war sequences present in the articles. A total of 91.64% of the articles had at least one edit war. The article 'George W. Bush' had the highest number of edit wars, viz., 1,559, followed by the article 'Wikipedia' having 1,091 edit wars. In total, we found 96,734 instances of edit wars. Figure 2(b) shows the distribution of the number of edit wars in the articles.

3.2.1 Types of Edit Wars. We came across many types of edit wars based on the number of users involved as well as the number of revisions that were reverted in the constituting reverts. We can, therefore, define the types of edit wars as follows:

(1) **Single-revision versus Multiple-revision Edit Wars** We define an edit war sequence to be single-revision edit war if each constituting revert restored the article state to the immediate previous revision, i.e., only one revision's changes were reverted in each constituting revert. An example is shown in Figure 3(a). On the contrary, we refer to an edit war as multiple-revision edit war if at least one constituting revert reverted the article state to revisions that are not

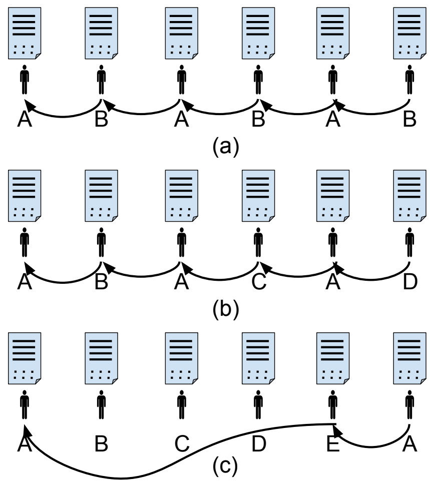


Figure 3: Types of Edit Wars: (a) Single-revision/Alternating users Edit wars (b) Multi-user Edit Wars (c) Multiple-revisions Edit Wars

immediate previous, thus undoing the content of many users, as shown in Figure 3(c).

(2) **Alternating-user versus Multi-user Edit Wars** We refer the edit wars where only two editors are involved in the entire edit war sequence as alternating-user or two-user edit wars. Figure 3(a) is an example of such edit wars, and these are the most common type of edit wars found in Wikipedia articles. On the other hand, the sequences that involve more than two users are referred to as multi-user edit wars as shown in Figure 3(b). Here, the initial reverter (i.e., B) is replaced by new reverters (i.e., C and D here), thus taking up the job of further reverting A's contribution.

The above types of edit wars are not mutually exclusive and in practice, a combination of these may exist. For instance, single-revision multi-user edit wars are a common type of edit wars observed.

3.2.2 Nullified vs Non-nullified Edit Wars. In addition to the categorization of edit wars based on the number of revisions and users, we classify them based on their final outcome. Consequently, we refer to an edit war as a *nullified* edit war, if it ultimately does not take effect in the article. Alternatively, if the initial reverter's revert is nullified at the end of the war, resulting in the article having the same content at the end of the war as it was before the start of the war. As an example, at the end of edit war A <- B <- A <- B <- A that consists of 4 reverts, the content of the article remains the same as it was before the edit war started, i.e., A's contribution stays. Clearly, in the case of single-revision reverts, the even-length reverts would be nullified reverts. On contrary, the reverts such as A <- B <- A <- B <- A <- B are referred to as *non-nullified* edit wars where the initial revert made by B to A still remains by the end, i.e., the edit war takes effect in the article. Hence, in the case of single-revision reverts, the odd-length reverts would be non-nullified reverts.

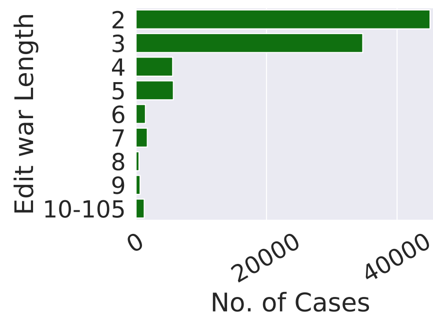


Figure 4: Distribution of length of edit wars. There were 1, 270 **edit wars with length ranging from** 10 **to** 105**.**

3.2.3 Length of Edit wars. Figure 4 shows the distribution of length of edit wars. There were 45, 051 edit wars with length 2. Further, there were 1, 270 edit wars with length more than 10. The length of the largest edit war was found to be 105, which we examine in detail later in Subsection 4.2.

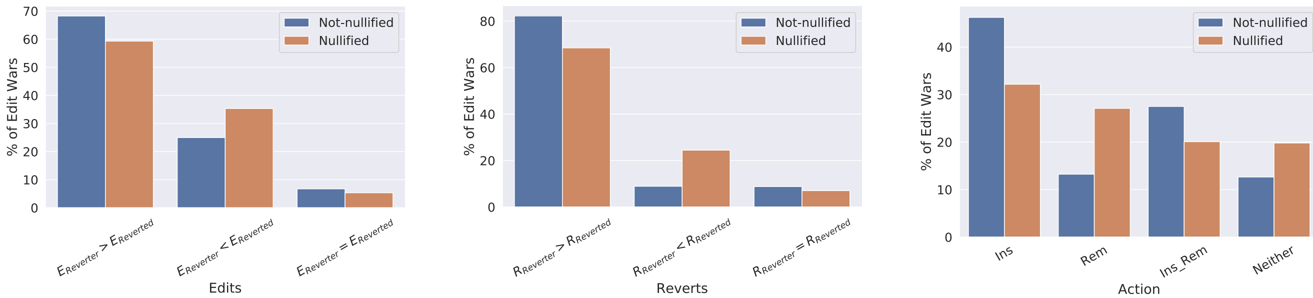
4 EXAMINING EDIT WAR SEQUENCES

In this section, we report the analysis carried out to examine various characteristics of edit war sequences, to get an idea of what kind of sequences are generally nullified. Apart from a general analysis of the edit war sequences, we also examine in more detail the longest edit war sequence to get more insights as well as see an instance of the formation of conflicting groups of users in the articles.

4.1 Analysing Sequences' Properties

We examined a total of 33, 142 single-revision (multi-user) sequences for various properties concerning the characteristics of the users involved as well as the conflicting content. These sequences consisted of 19, 453 (58.69%) non-nullified cases and 13, 689 (41.31%) nullified cases indicating that more proportion of edit wars were non-nullified, i.e. they took effect in the articles.

4.1.1 Edits. To check whether the edits made by less-contributing users are reverted more often or not as the final outcome of the edit war, we examined the number of edits made in the article by all the participating users. We then compared the edits of the initial reverted user and the initial reverter for the nullified and non-nullified edit war sequences. As shown in Figure 5(a), it was observed that when the initial reverter had more edits than the initial reverted user, it was a little more likely that the corresponding edit war would be non-nullified, i.e., the initial revert of the edit war will take effect at the end. On the contrary, as the second column in the Figure shows, when the initial reverter had lesser edits than the initial reverted user, it was more likely that the corresponding edit war was nullified, i.e., the effect of this initial revert will be nullified by the end. The difference in the likelihood in both cases, however, was not very high. Further, no visible pattern was observed when the initial reverter and the reverted user had a comparable number of edits.



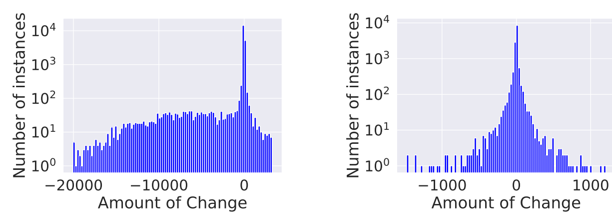
(a) (b) (c)

Figure 5: (a) Proportion of edit wars where the number of edits of the reverter was more, less or equal to that of the reverted user in case of nullified and non-nullified edit war sequences. (b) The proportion of edit wars where the number of reverts of the reverter was more, less or equal to that of the reverted user in case of nullified and non-nullified edit wars. (c) The proportion of cases where the initial reverted user had inserted, removed, done both or done neither for nullified and non-nullified edit war sequences. ('Neither' implies to cases where the changes were done only at the formatting level and not at content level)

4.1.2 Reverts. Just like edits, a similar behavior was observed for the number of reverts that the initial reverter and the reverted user had made in the article. That is, as shown in Figure 2(b), it was observed that when the initial reverter had more reverts than the initial reverted user, there was a little higher chance of the corresponding edit war being non-nullified. The converse was observed for the cases where the number of reverts of reverter was lesser than the reverted user. Further, no significant difference was observed when the edits as well as reverts of the initial reverter and the reverted user were comparable. The observation echoes the findings of Tsvetkova et al. [34] where they report that the users who revert frequently tend to revert users who revert rarely.

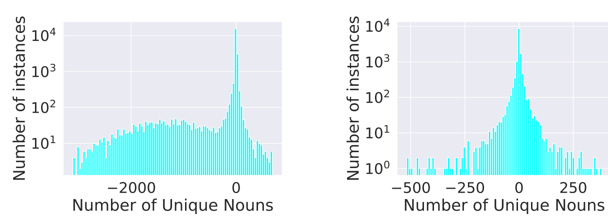
4.1.3 Insertion/removal of content. The edit wars may happen over either insertion or removal of content or both. Therefore, we examined whether the initial reverted user had removed some content from the article or had inserted some new content or done both in their edit, which had led to the start of the edit war. We identified the conflicting content as the difference of the content of the article after the revision made by the reverted user and the revision made by the reverter. Figure 5(c) shows the data from the perspective of the initial reverted user. We see that when the initial reverted user had inserted some content, then it was more likely that the edit war was non-nullified than when he had removed content.

We also examined the chances of the edit war being nullified or non-nullified with respect to the amount of content removed or inserted. The negative values in Figure 6 represent the removed content in terms of the number of words while the positive values represent the inserted content. We see that in non-nullified cases of edit wars where the content was removed, the amount of this removed content was very high. In the case of nullified cases, no such pattern concerning the amount of content inserted/removed was observed. This points to a possibility of a large amount of content being removed from the articles - vandalism being one of the likely reasons - thus making the edit wars non-nullified. A similar trend was observed when instead of just words, the number of nouns inserted/removed was observed (See Figure 7).



(a) Non-nullified (b) Nullified

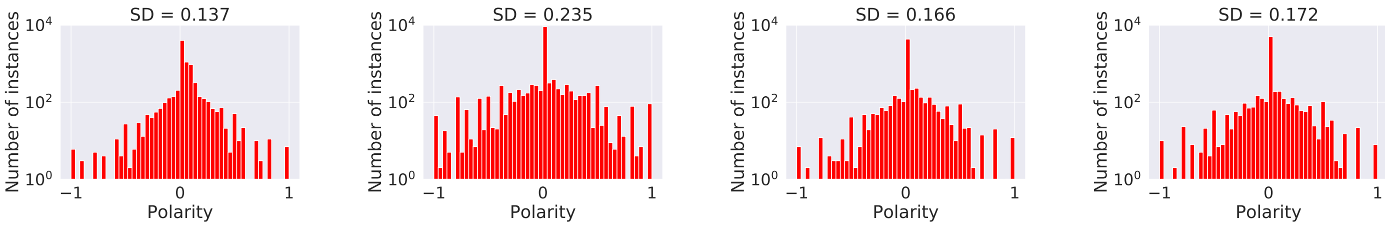
Figure 6: Amount of change made in the articles (in words). Negative amount represents removal and positive amount represents insertion. (Y-axis is log-scaled.)



(a) Non-nullified (b) Nullified

Figure 7: Amount of change (in nouns). Negative amount represents removal and positive amount represents insertion. (Y-axis is log-scaled.)

Additionally, it was found that in non-nullified edit wars, 11.25% cases involved the insertion of content consisting of *bad/swear words* intended to compromise the integrity of the content, while



(a) Non-nullified Edit Wars (Removal by Reverted User)

(b) Non-nullified Edit Wars (Insertion by Reverted User)

(c) Nullified Edit Wars (Removal by Reverted User)

(d) Nullified Edit Wars (Insertion by Reverted User)

Figure 8: Polarity of the inserted/removed content in different types of edit wars (Y-axis is log-scaled.)

2.73% of nullified edit wars contained an insertion of such words. We took the list of bad words banned by Google for this analysis. This list also contains some words such as ‘stupid’, ‘kill’, ‘opium’, ‘gay’, and ‘ugly’ as well, which explains the reason for the presence of a small percent of nullified edit wars despite consisting of bad words’ insertion.

4.1.4 Polarity and Subjectivity. Many a time, edit wars happen when the editors keep posting content of opposite polarity. We computed the polarity and subjectivity of the conflicting content in the edit war sequences. It was found that in the non-nullified edit wars, the content inserted had a large number of cases with high positive or negative polarity, as shown in Figure 8(b). In other words, the users who were reverted had inserted content that was highly polar, hence it was reverted through the edit war. This was confirmed by a high SD (0.235) of the polarity values obtained for the non-nullified cases where insertion happened as compared to the rest of the cases, i.e. non-nullified cases where removal happened (SD = 0.137) (Figure 8(a)), nullified cases where removal happened (SD = 0.166) (Figure 8(c)) and nullified cases where insertion took place (SD = 0.172) (Figure 8(d)).

In addition to this, in the cases of non-nullified edit wars, the subjectivity was high, both for inserted as well as removed content. This indicates that the user who was reverted, had either removed or inserted some content that was highly subjective thus leading to the edit war. The average values of subjectivity for the content inserted and removed in non-nullified cases were 0.226 and 0.240 respectively. However, in the case of nullified cases, the average subjectivity values for inserted and removed cases were 0.180 and 0.193 respectively. This indicates that whenever subjective content is inserted into the articles, or the subjective content added by other users is removed, it leads to non-nullified conflicts.

4.2 Longest Edit War Sequence

We now examine in detail the longest edit war sequence that was found in the article on *Mustafa Kemal Ataturk*, the first president of the Republic of Turkey. It contained 105 reverts, 107 revisions and involved 20 users. It was a non-nullified edit war. It was mostly a single-revision, multi-user edit war, except that two of the reverts involved jumping to two revisions back instead of the immediate previous revision. We study this sequence to get more insights into the conflicting groups present in edit wars sequences. Figure 9 shows a snapshot of the revision history of the article consisting of a part of the edit war sequence showing back and forth insertion and removal of 85, 794 bytes from the article. Figure 10 further shows the entire sequence of reverts along with the users who were supporting positive content about the president as well as the users supporting the inclusion of the negative content about him. This edit war started when a user named *hairyHarry88* removed the entire content of the article and added a paragraph criticizing Mustafa by adding many negative details about him. This was reverted by a user named *ClueBot*, which was further reverted by *Hairyharry* multiple times. Over the course, many other users supporting these two users respectively participated and the edit war finally ended by the initial reverter, i.e., *ClueBot*, nullifying the inclusion of the negative details about the president, thereby, retaining the content that was in the article before the edit war had started. The numbers in the Figure 10 represent the revisions of the edit war in order. A follow-through of these numbers shows how the edit war initiated by two users *Hairy Harry* and *Cluebot* was soon taken over by other users, thus giving rise to two groups of opposing views about the president. Finally, the positive details about the president stayed in the article.

Following are a few details about this edit war:

(1) The size of the content of the article kept fluctuating between 3, 882 bytes to 89, 676 bytes (See Figure 11(a)). Here, the smaller sized content was not a subset of the larger one. Rather, it was an altogether new content writing completely different facts about the president. The users adding the smaller sized content were writing negative details while the larger-sized content comprised of positive details about the president.
(2) The subjectivity of the content also kept fluctuating between 0.44 and 0.33, with the negative details about the president being more subjective.
(3) No particular behavior was observed concerning the total number of edits made by the users supporting the contradicting views about the president. As shown in Figure 11, there were both high as well as low contributing users supporting the two extreme views.
(4) All the 105 reverts of the edit war took place within two days starting from the evening of 19 April 2008 to the evening of 21 April 2008.

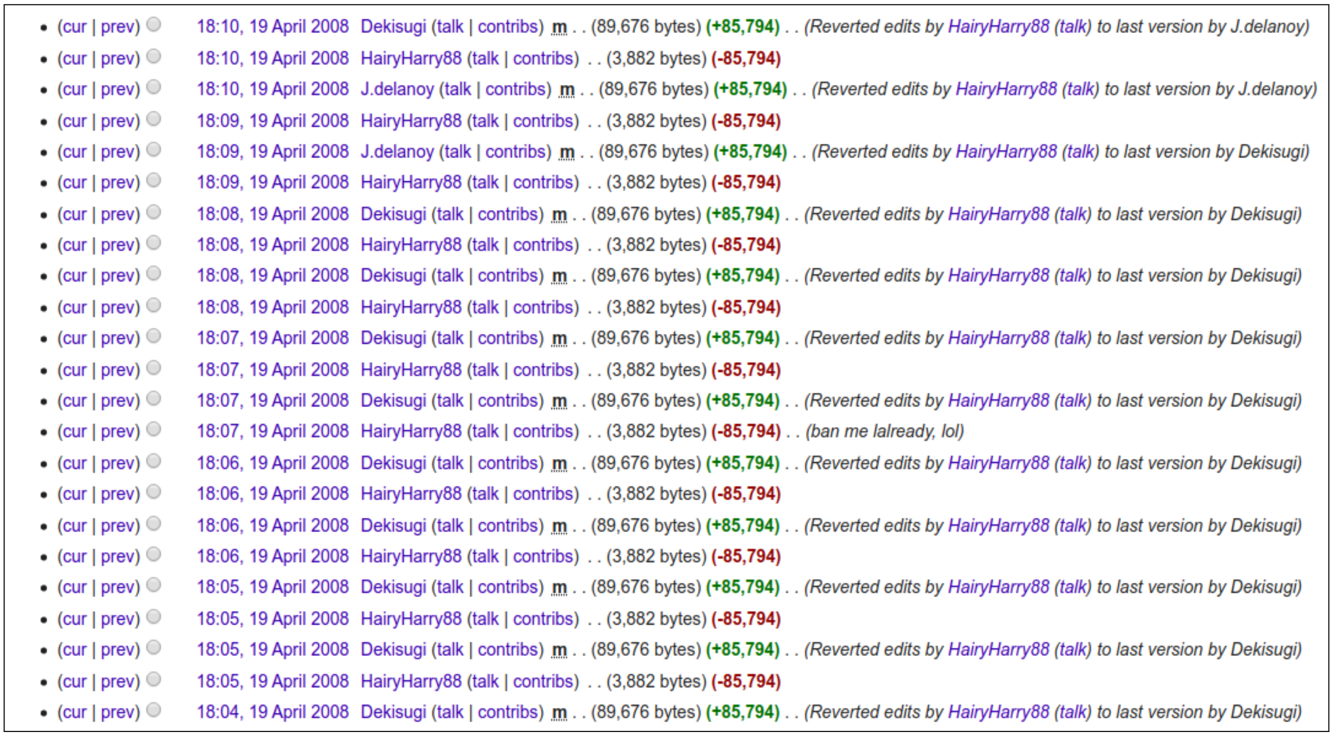

- (cur | prev) 18:10, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by J.delanoy)
- (cur | prev) 18:10, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:10, 19 April 2008 J.delanoy (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by J.delanoy)
- (cur | prev) 18:09, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:09, 19 April 2008 J.delanoy (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)
- (cur | prev) 18:09, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:08, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)
- (cur | prev) 18:08, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:08, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)
- (cur | prev) 18:08, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:07, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)
- (cur | prev) 18:07, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:07, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)
- (cur | prev) 18:07, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794) . . (ban me lalready, lol)
- (cur | prev) 18:06, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)
- (cur | prev) 18:06, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:06, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)
- (cur | prev) 18:06, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:05, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)
- (cur | prev) 18:05, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:05, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)
- (cur | prev) 18:05, 19 April 2008 HairyHarry88 (talk | contribs) . . (3,882 bytes) (-85,794)
- (cur | prev) 18:04, 19 April 2008 Dekisugi (talk | contribs) m . . (89,676 bytes) (+85,794) . . (Reverted edits by HairyHarry88 (talk) to last version by Dekisugi)

Figure 9: A screenshot showing a part of the revision history of the article 'Mustafa_Kemal_Ataturk'

(5) Additionally, it was interesting to see that the number of reverts by users in this particular edit war was either one less or exactly equal to their number of edits. This shows that many of these users were only contributing in the form of reverts.

5 ESTIMATING PROPERTIES OF EDIT WAR SEQUENCES

Handling conflicts leads to extra overhead and coordination costs apart from disrupting the collaborative efforts. Early detection of conflicts might help in saving the article being in an unstable state for long as well as conserving the coordination efforts. As an example, by using the initial signatures of edit wars, if it is predicted whether they will settle soon or may lead to lengthy back and forth changes in the article, necessary intervention measures may be taken. This might help in controlling them early, thus saving unnecessary time. Moreover, using their properties, it may also be possible to predict which side of the two conflicting groups they are likely to be inclined to eventually. These predictions may be made based on various parameters of these edit wars, such as the past contribution of the users involved in terms of the number of edits, reverts and the characteristics of the content that is being fought over etc. All these parameters are known in the initial stages of an edit war. They may, thus, help in devising automated means to control edit wars in the articles, reducing the need for the involvement of administrators for coordination.

5.1 Edit Wars' Outcome

The problem of estimating the edit wars' outcome is framed as a binary classification task, where a target set of edit wars needs to be classified as nullified or non-nullified edit wars. We use Support Vector Classification with a non-linear kernel for this purpose. Each edit war is associated with a feature vector, containing the following parameters:

- ***ins_rem_both_none*** *(*0/1/2/3*)*: Whether the initial reverted user had inserted, removed, done both or done neither in terms of number of words.
- ***ins_polarity***: Polarity of the inserted content. It ranges from -1 to 1.
- ***rem_polarity***: Polarity of the removed content. It ranges from -1 to 1.
- ***ins_subj***: Subjectivity of the inserted content. It ranges from 0 to 1.
- ***rem_subj***: Subjectivity of the removed content. It ranges from 0 to 1.
- ***whose_edits_more***: 1 if number of edits of reverted user are more, 0 if number of edits of the reverter are more.
- ***whose_reverts_more***: 1 if number of edits of reverted user are more, 0 if number of edits of the reverter are more.
- ***bad_words_present***: 1 if *bad words* are present in the conflicting content and 0 if they are not present.

The target variable is: ***N_NN***, that tells Whether the edit war is finally going to be nullified or non-nullified.

We compute the above features for 33, 142 samples of edit wars containing 19, 453 instances of the non-nullified (NN) class and

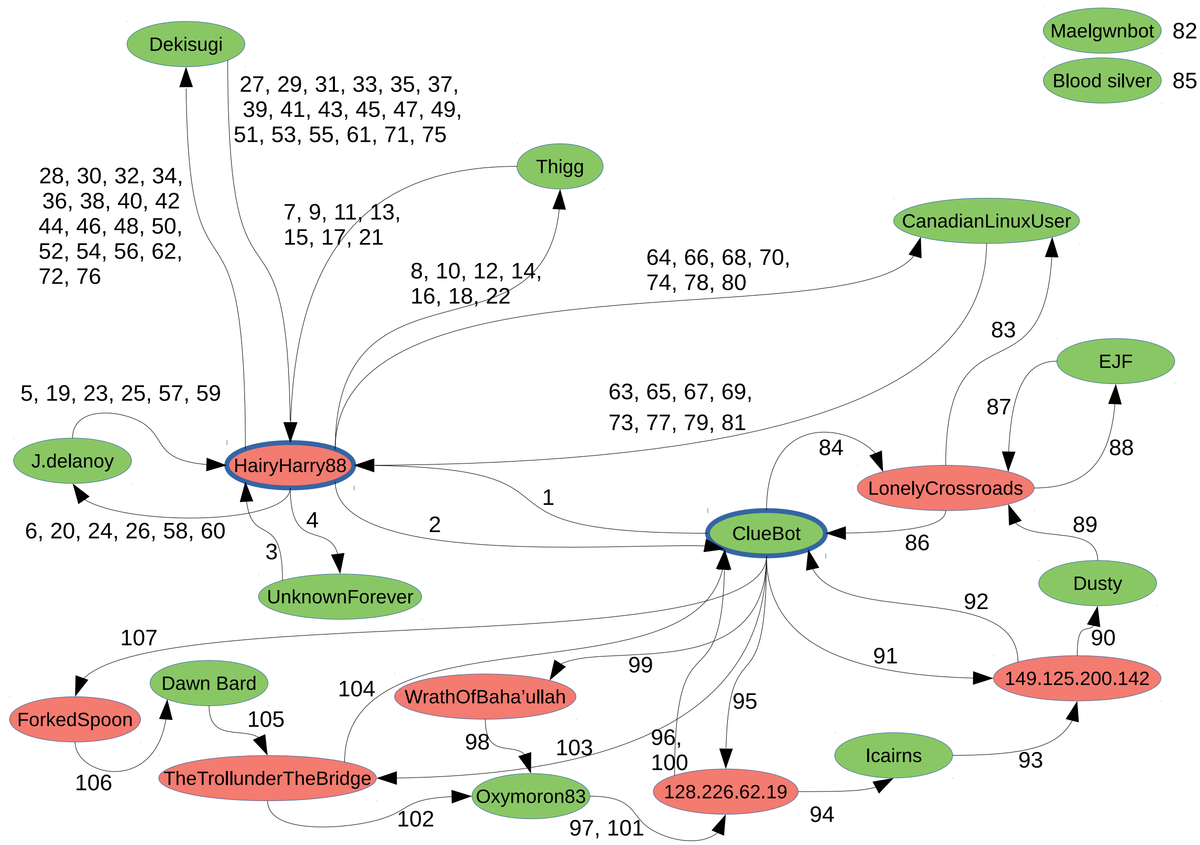


Figure 10: Sequence of reverts in the article *Mustafa_Kemal_Ataturk*. The numbers on the edges represent the revisions in order. No reverts happened in revisions 82 and 85. The green and red nodes represent two groups of users having opposing views about the president.

Class	Precision	Recall	F1-score
0 (N)	0.61	0.42	0.50
1 (NN)	0.67	0.81	0.73

Table 2: Classification Evaluation Parameters

Mean Absolute Error	Median Absolute Error	RMSE
0.848	0.273	1.653

Table 3: Regression Evaluation Parameters

13, 689 instances of the nullified (N) class. The classifier predicts the edit wars' outcome with 66% accuracy. The other evaluation parameters are shown in Table 2.

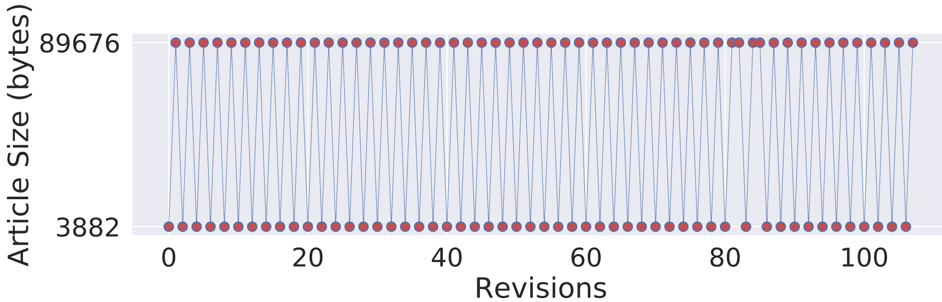
5.2 Edit Wars' Length

Further, using the same samples of edit wars, we use Support Vector Regression to estimate the length of edit wars. The target variable is ***len_ew*** whose value ranged between 2 and 105 as noted previously. The model predicted the edit wars' length with a root mean square error of 1.653. Other evaluation parameters of the model are as shown in Table 3.

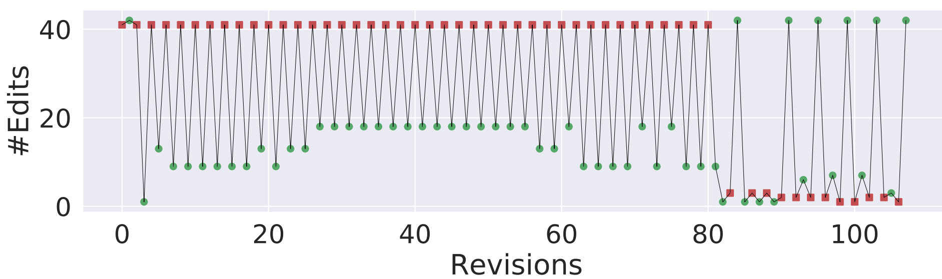
Estimation of length of an edit war sequence based on its initial signatures may be a valuable information for the administrators. This may signal timely intervention that may help avoid the edit wars going on and on and thus saving the article stay in an unstable state for long periods of time.

6 DISCUSSION

Our study shows that it is possible to predict certain behaviors of the conflicts taking place in Wikipedia articles. The features used above are the ones that can be computed in the beginning stages of an edit war. These features may provide details about the prospective outcomes of these edit wars. For example, such models may predict that based on their features, how long an edit war is expected to go on, thus flagging them in the system, prompting administrators involvement early enough. In the analysis, we have used the samples taken from the edit war sequences present in the controversial articles. However, if the samples contain only those edit war sequences where the administrators took appropriate actions, such sequence samples may be used to train the model, thus



(a)



(b)

Figure 11: (a) Variation in the article size during the longest edit war. The size of the article kept fluctuating between 3882 **bytes and** 89, 676 **bytes dut to edits of the users shown as red and green in Figure 10 respectively. (b) The number of edits by the users involved in the longest edit war in the order of their contribution. The plot shows that there were both high as well as low contributing users in the two groups trying to add positive and negative details about the president respectively.**

automatically handling conflicts in terms of nullifying them or not nullifying them, without the involvement of these administrators. Employing the above measures may help in combating conflicts on time which is essential for any collaborative environment.

There are a few limitations or future works of this study. For example, in this work, we focus only on those conflicts that involve the complete revert of the article content to an older state. We do not study partial reverts involving revert of a part of the changed content which is a future work of this study. Also, in this work, we examine single-revision multi-user edit wars. As an extension of this study, we plan to work on other types of edit wars that were observed, as shown in Figure 3.

7 CONCLUSION

This work investigates the dynamics of edit wars by examining the individual instances of edit war sequences. The analysis reveals that different parameters concerning the conflicting content as well as the characteristics of the users involved may divulge future details about them. These details may be used to control them in time through automated measures, thus reducing human effort for coordination. We believe that this empirical study will serve as a basis for more in-depth investigations leading to automatic conflict detection and resolution tools for Wikipedia as well as other large-scale collective and collaborative problem-solving projects.

8 ACKNOWLEDGEMENTS

This work was supported in part by WOS-A, Department of Science and Technology, India [SR/WOS-A/ET-1058/2014] and CSRI, Department of Science and Technology, India [SR/CSRI/344/2016].

REFERENCES

[1] Daron Acemoglu and Asuman Ozdaglar. 2011. Opinion dynamics and learning in social networks. *Dynamic Games and Applications* 1, 1 (2011), 3–49.

[2] Ofer Arazy, Lisa Yeo, and Oded Nov. 2013. Stay on the Wikipedia task: When task-related disagreements slip into personal and procedural conflicts. *Journal of the American Society for Information Science and Technology* 64, 8 (2013), 1634–1648.

[3] Erik Borra, Esther Weltevrede, Paolo Ciuccarelli, Andreas Kaltenbrunner, David Laniado, Giovanni Magni, Michele Mauri, Richard Rogers, and Tommaso Venturini. 2015. Societal controversies in Wikipedia articles. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 193–196.

[4] Erik Borra, Esther Weltevrede, Paolo Ciuccarelli, Andreas Kaltenbrunner, David Laniado, Giovanni Magni, Michele Mauri, Richard Rogers, Tommaso Venturini, et al. 2014. Contropedia-the analysis and visualization of controversies in Wikipedia articles.. In *OpenSym*. 34–1.

[5] Ulrik Brandes, Patrick Kenis, Jürgen Lerner, and Denise van Raaij. 2009. Network analysis of collaboration structure in Wikipedia. In *Proceedings of the 18th international conference on World wide web*. ACM, 731–740.

[6] Anamika Chhabra and SRS Iyengar. 2017. How Does Knowledge Come By? *arXiv preprint arXiv:1705.06946* (2017).

[7] Anamika Chhabra, SR Iyengar, JS Saini, and Vaibhav Malik. 2020. Activity-selection Behavior and Optimal User-distribution in Q&A Websites. *Proceedings of the international conference on Computational collective intelligence (ICCCI'2020)-In press* (2020). https://www.researchgate.net/publication/343553414_Activity-selection_Behavior_and_Optimal_User-distribution_in_QA_Websites

[8] Anamika Chhabra, SRS Iyengar, Poonam Saini, and Rajesh Shreedhar Bhat. 2015. Presence of an Ecosystem: a Catalyst in the Knowledge Building Process in

Crowdsourced Annotation Environments. *Proceedings of the 2015 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2015)* (2015).

[9] Anamika Chhabra and SR Sudarshan Iyengar. 2018. Characterizing the Triggering Phenomenon in Wikipedia. In *Proceedings of the 14th International Symposium on Open Collaboration*. ACM, 11.

[10] Anamika Chhabra and S RS Iyengar. 2020. Activity-selection Behavior of Users in StackExchange Websites. In *Companion Proceedings of the Web Conference 2020*. 105–106.

[11] Si-Chi Chin, W Nick Street, Padmini Srinivasan, and David Eichmann. 2010. Detecting Wikipedia vandalism with active learning and statistical language models. In *Proceedings of the 4th workshop on Information credibility*. 3–10.

[12] Abhimanyu Das, Sreenivas Gollapudi, and Kamesh Munagala. 2014. Modeling opinion dynamics in social networks. In *Proceedings of the 7th ACM international conference on Web search and data mining*. ACM, 403–412.

[13] Simon DeDeo. 2016. Conflict and computation on wikipedia: A finite-state machine analysis of editor interactions. *Future Internet* 8, 3 (2016), 31.

[14] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying controversy on social media. *ACM Transactions on Social Computing* 1, 1 (2018), 1–27.

[15] Aaron Halfaker, Aniket Kittur, and John Riedl. 2011. Don't bite the newbies: how reverts affect the quantity and quality of Wikipedia work. In *Proceedings of the 7th international symposium on wikis and open collaboration*. ACM, 163–172.

[16] P Sol Hart and Erik C Nisbet. 2012. Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research* 39, 6 (2012), 701–723.

[17] Gerardo Iñiguez, János Török, Taha Yasseri, Kimmo Kaski, and János Kertész. 2014. Modeling social dynamics in a collaborative environment. *EPJ Data Science* 3, 1 (2014), 7.

[18] Dariusz Jemielniak. 2014. *Common knowledge?: An ethnography of Wikipedia*. Stanford University Press.

[19] Aparna Joshi and Hyuntak Roh. 2009. The role of context in work team diversity research: A meta-analytic review. *Academy of management journal* 52, 3 (2009), 599–627.

[20] Rishemjit Kaur, Ritesh Kumar, Amol P Bhondekar, and Pawan Kapur. 2013. Human opinion dynamics: An inspiration to solve complex optimization problems. *Scientific reports* 3 (2013), 3008.

[21] Aniket Kittur, Bongwon Suh, Bryan A Pendleton, and Ed H Chi. 2007. He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 453–462.

[22] Travis Kriplean, Ivan Beschastnikh, David W McDonald, and Scott A Golder. 2007. Community, consensus, coercion, control: cs* w or how policy mediates mass participation. In *Proceedings of the 2007 international ACM conference on Supporting group work*. 167–176.

[23] Elizabeth Mannix and Margaret A Neale. 2005. What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological science in the public interest* 6, 2 (2005), 31–55.

[24] Mathias Wullum Nielsen, Sharla Alegria, Love Börjeson, Henry Etzkowitz, Holly J Falk-Krzesinski, Aparna Joshi, Erin Leahey, Laurel Smith-Doerr, Anita Williams Woolley, and Londa Schiebinger. 2017. Opinion: Gender diversity leads to better science. *Proceedings of the National Academy of Sciences* 114, 8 (2017), 1740–1742.

[25] Scott E Page. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies-New Edition*. Princeton University Press.

[26] Reid Priedhorsky, Jilin Chen, Shyong (Tony) K Lam, Katherine Panciera, Loren Terveen, and John Riedl. 2007. Creating, destroying, and restoring value in Wikipedia. In *Proceedings of the 2007 international ACM conference on Supporting group work*. 259–268.

[27] Ronald Rivest and S Dusse. 1992. The MD5 message-digest algorithm. (1992).

[28] Csilla Rudas, Olivér Surányi, Taha Yasseri, and János Török. 2017. Understanding and coping with extremism in an online collaborative environment: A data-driven modeling. *PloS one* 12, 3 (2017), e0173561.

[29] Csilla Rudas and János Török. 2018. Modeling the Wikipedia to Understand the Dynamics of Long Disputes and Biased Articles. *Historical Social Research/Historische Sozialforschung* 43, 1 (163 (2018), 72–88.

[30] Hoda Sepehri Rad, Aibek Makazhanov, Davood Rafiei, and Denilson Barbosa. 2012. Leveraging editor collaboration patterns in Wikipedia. In *Proceedings of the 23rd ACM conference on Hypertext and social media*. ACM, 13–22.

[31] Feng Shi, Misha Teplitskiy, Eamon Duede, and James A Evans. 2019. The wisdom of polarized crowds. *Nature Human Behaviour* (2019), 1.

[32] Bongwon Suh, Ed H Chi, Bryan A Pendleton, and Aniket Kittur. 2007. Us vs. them: Understanding social dynamics in wikipedia with revert graph visualizations. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*. IEEE, 163–170.

[33] Cass R Sunstein. 2018. *# Republic: Divided democracy in the age of social media*. Princeton University Press.

[34] Milena Tsvetkova, Ruth García-Gavilanes, and Taha Yasseri. 2016. Dynamics of disagreement: Large-scale temporal network analysis reveals negative interactions in online collaboration. *Scientific reports* 6 (2016), 36333.

[35] Fernanda B Viégas, Martin Wattenberg, and Kushal Dave. 2004. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 575–582.

[36] Fernanda B Viegas, Martin Wattenberg, Jesse Kriss, and Frank Van Ham. 2007. Talk before you type: Coordination in Wikipedia. In *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. IEEE, 78–78.

[37] Dennis M Wilkinson and Bernardo A Huberman. 2007. Cooperation and quality in wikipedia. In *Proceedings of the 2007 international symposium on Wikis*. 157–164.

[38] Taha Yasseri, Robert Sumi, András Rung, András Kornai, and János Kertész. 2012. Dynamics of conflicts in Wikipedia. *PloS one* 7, 6 (2012), e38869.