Wikipedia Voting

NATHAN LANG*, Marquette University, United States
AMY GERAGHTY, Marquette University, United States
OLGA KOZLOVA, Marquette University, United States

Free-content, crowd-sourcing websites continue gaining popularity and utilization. Citing Wikipedia has become a norm due to its vast content and ease of access. Such growth invites scrutiny of self-governing processes by which such entities operate. In this paper, we present a study of the Wikipedia Request for Adminship election process, heralded by Wikipedia as fully transparent consensus building. We demonstrate the emergence of information cascades by graphical analysis. Additionally, using the LDA (Latent DirichletAllocation) method to perform topic modeling, we find a significant degree of subjectivity in voters' comments, despite good consistency in intent and content. We thus argue that the RfA process is not free from bias which often accompanies public sequential process such as online voting.

CCS Concepts: • **Self-governing Voting**; • **Cascading**; • **LDA**; • **Social Graphs**;

ACM Reference Format:
Nathan Lang, Amy Geraghty, and Olga Kozlova. 2019. Wikipedia Voting. 1, 1 (December 2019), 19 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Among the most prominent free-content, crowd-sourcing websites, Wikipedia stands out for a number of reasons. Wikipedia is a multilingual, web-based, free-content encyclopedia project supported by the Wikimedia Foundation and based on a model of openly editable content. [6] Today, the English site of the encyclopedia counts nearly 6 million articles, while the total number of available languages is 302 with over 48 million articles. [6] In 2015, the encyclopedia reported that almost 375 million unique visitors accessed its content. [6] The sheer volume of the information, the crowd-sourcing approach to content creation, and abundant website traffic present unique challenges to keep Wikipedia a verifiable and reliable source. Success of these efforts has been often brought into question. Cases that have been brought forth against Wikipedia range from manipulation of content by groundless deletions of pages to misrepresentation of information to vandalism. [7] In this context, both internal and public policies and procedures of Wikipedia, particularly those subject to self-governance, are an important indicator of quality and integrity.

In this paper we pursue the subject of content management on Wikipedia as this process is the foundation of what the encyclopedia visitors can access on the websites. We will approach this question from the perspective of understanding who gets elected to be Administrator, a Wikipedia formal position empowered with privileges beyond those of lower-rank contributors such as deletion of articles deemed unsuitable, protecting pages (restricting editing privileges to that page), and

*All authors contributed equally on this paper

Authors' addresses: Nathan Lang, Marquette University, Milwaukee, United States; Amy Geraghty, Marquette University, Milwaukee, United States; Olga Kozlova, Marquette University, Milwaukee, United States.

XXXX-XXXX/2019/12-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

blocking the accounts of disruptive users. [6] Wikipedia has established the Wikipedia community, comprised of hundreds of thousands of Editors who spend millions of hours annually editing Wikipedia's content. According to Leskovec's et al. influential work on Wikipedia's community, Wikipedia engages in complex self-governance, characteristic of deliberative processes carried out in public light. Election to the Administrator role is one of the prominent examples of such self-governance.

As we dissect the election process, we are also interested in understanding how the psychological phenomenon of information cascades plays out in the Wikipedia self-governing processes. The presence of this phenomenon is justified by the very nature of the election process – public and sequential. As committed Wikipedia members, known as RfA (Request for Administration) contributors in the election process, publicly elect Administrators, i.e., there is a public record of past votes, it is reasonable to suggest the possibility of an information cascade forming, thus giving earlier votes a more weighted significance.

Information cascades, according to Easley et. al occur when there is a certain confluence of the right ingredients. Firstly, there must be a decision to make. Secondly, people make decisions sequentially and are able to observe previous decisions. Thirdly, possession of private information about the decision is important but people usually can only make inferences about each other's private knowledge as it is not easily accessible. In the case of the RfA process, the situation is complicated, however, by the fact that voting is fully transparent, i.e., RfA contributors can see each other's votes and comments on a given Administrator candidate's application. Thus, the next voting member has access to past votes and, most importantly, additional information in the form of reasoning and justification. Moreover, the next voting member can gauge the amount of support from the most influential members of the Wikipedia community.

The subject of our study is timely and in demand for improved transparency. We have already mentioned the recorded history content manipulation by Wikipedia. In addition, there are numerous recorded complaints targeted at the management of privacy, diversity, etc.to include the effects of users' anonymity, the attitudes towards newcomers, biases in the social structure of the community, in particular, gender bias and lack of female contributors, and the role of the project's co-founder Jimmy Wales, in the community. [9] The abuse of privileges by administrators has been reported as well. Considering these issues, it is worth investigating the structure of the community further to understand how the role of the Administrator influences the content of the encyclopedia.

2 LITERATURE REVIEW

Several studies of the Wikipedia Administrator election process have been conducted in the past. The study by Burke et al., presented at the 2008 Conference on Human Factors in Computing Systems, sought to isolate the most important predictors of the promotion to the status of the Wikipedia Administrator. The researchers found that merely performing a lot of production work was insufficient for "promotion" on Wikipedia. [2] Moreover, candidates' article edits were weak predictors of success. Diverse experience and contributions to the development of policies and WikiProjects were stronger predictors of election success. [2] This study also referenced a secret mailing system by which Wikipedia Administrators coordinated their actions toward others, as described in the previous works by Hastie and Dawes. While we did not find evidence of such a secret system today, our original inquiry into just how much influence individual Wikipedia Administrators may wield in the election procedure becomes more gainful in light of this previous research.

A 2018 paper by Asim et al. posited a question of the significance of personal connections of an applicant for the Wikipedia Administrator role. The study uncovered a number of important phenomena in the election process, particularly that personal contacts, i.e. immediate neighbors (measured by degree centrality) and neighborhood (identified as k-neighbors) of an applicant have a positive effect on the chances of getting elected. [1] Moreover, being connected to other Editors who play the role of knowledge brokers and promote reachability increases the likelihood of success as well. [1] Lastly, higher centrality scores seemed to activate voters to vote [for the applicant] thus suggesting that knowing many current Editors is a strong predictor of success as well. [1] There have been other studies aiming to identify most influential voters. In 2014 Desai et al. published a paper in which the authors made election predictions of high accuracy using users' past votes as features and feature selection to locate "influential" voters in the Wikipedia ecosystem. Using modified Multinomial Naive Bayes on all the voters, they were able to obtain around 97.3% accuracy. From a dataset with 2,761 elections and 6,210 voters, the researchers isolated about 40 most influential voters whose votes yielded 95% accuracy in predicting elections. Thus, one can surmise that a very small subset of the Wikipedia community holds decision making power to shape the editorial brain trust of the encyclopedia. This further strengthens our intent to understand the probability and strength of information cascades occurring during the election process.

In 2010, Leskovec et al. published extensive research on social media self-governance using the Wikipedia Administrator election as an example of public self-regulated deliberation. The authors focused on the phenomena of herding and information cascades by exploring the temporal dynamics of election processes. Some observations are unexpected, particularly in light of the traditional herding theory; some findings loosely align with Desai's discovery of 40 most influential voters whose voting activity predicts that of the rest of RfA contributors. In fact, Leskovec et al. wrote that there was strong evidence suggesting that the success of an election heavily depended on the initial votes, indicating the presence of an information cascade. However, when explored further, the order in which the initial set of positive and negative votes was formed had no significant impact on the outcome of the election. This observation led the authors to the conclusion that the expected sequential conformity was absent, thus defying the principles of the herding theory and information cascades.

Wu and Huberman support this absence by arguing that in social media, people tend to express their opinion more rigorously when it is against the prevailing possible outcome. [10] For example, should the next RfA contributor discover a low fraction of support votes for the candidate they particularly find worthy of the promotion, their motivation to cast a support vote increases, despite the expectation to conform according to the herding and information cascade theories. Now to the question of private information, an essential element of information cascade dynamics. Since the voting process is intended to be fully transparent and public, one may argue that there is a small chance for private information to execute its due diligence on the decision making process. However, it is not necessarily true. Consider other important observations from Leskovec's study which furthered the work of Burt on the importance of the dynamics between absolute value of a phenomenon vs. relative comparisons of self against a peer group. In general, in these observations we see evidence that suggests that RfA contributors do possess critical private information which directs their decision-making in an election.

Of particular significance, for example, is the observation that RfA contributors are particularly critical to candidates whose level of achievement resembles that of their own (Leskovec et al., 2010). For examples, RfA contributors tend to give fewer support votes to those candidates who have a similar number of barnstars, a type of Wikipedia award for hard work and due diligence. This finding is in disharmony with the important psychological phenomenon of homophily which suggests that alike people gravitate towards each other. Nonetheless, another observation by Leskovec et al. suggests the opposite: the more frequently the RfA contributor and the candidate communicated prior to the election, the more likely is the former to cast a support vote. While in alignment with the past research, our study takes a step further by expanding the scope to include text analysis of the voters' comments thus seeking to identify common patterns in the qualitative part of the voting process.

2.1 Research Questions

(1) What are themes and topics for positive and negative comments and how consistent are they?
(2) Can we observe information cascades in voting behavior?

3 DATA METHODS

In 2019, Wikipedia reported 1,149 individuals acting as Administrators. The status is granted based on the process of Requests for Adminship (RfA) defined by Wikipedia. [6] Any registered Wikipedia Editor can comment on a request, and each Editor will assess each candidate in their own way. [6] The process is specifically called out as "consensus building' rather than direct vote, despite the fact that RfA contributors have three clear options to vote for the nominee with results tabulated and posted for public access. No quotas are set for any of the three options of voting, and the result is set by an uninvolved higher order Editor, known as bureaucrat, who closes the election and casts the final decision. This higher order Editor also analyzes each vote and accompanying comments to arrive at the final result; thus, it has been supposed that the rationale and identity of the voter are more important than the vote itself. [4] The final determination is not based exclusively on the percentage of support; in practice most RfAs above 75% pass. The community has determined that, in general, RfAs between 65 and 75% support should be subject to the discretion of bureaucrats. [6]

Our data set is pulled from the Stanford Large Network Dataset Collection. [11] The dataset contains all requests for adminship on Wikipedia between 2003 and 2013. [11] There are 198,275 observations in the dataset, and each observation represents a current editor voting on a candidate seeking administrator status on Wikipedia. The dataset has seven features: DAT, which is a timestamp of when the vote was cast, RES, which is the outcome of the election (-1 means the candidate was rejected and 1 means the candidate was accepted), SRC, which is the voter, TGT, which is the candidate seeking adminship, TXT, which is a comment that can be left explaining the voter's vote, VOT, which is how the SRC voted on the TGT (1 means support, 0 means neutral, -1 means oppose), and YEA, which is the year the election for that candidate started. Typical descriptive statistics were not very useful on this dataset since only three features had numerical values and, of those three features, one was a year, one could only have a value of -1 or 1, and the other could only have a value of -1, 0, or 1. However, some useful statistics about this dataset are that there are 3,497 unique candidates voted on between 2003 and 2013. Of those 3,497 candidates, 1,833 were granted administrator status, while the other 1,664 candidates were rejected. As seen below in Table 1, the number of elections held per year increases until 2007, and after that, it continues to decrease until 2013, which is the last year accounted for in the dataset.

3.1 LDA

The text component of this dataset was the comments that voters were able to leave with their vote of support, neutral, or oppose. Comments ranged from short, one-word statements about how the source voted, such as "Support" to long, extensive comments detailing the reasoning behind the vote of support, neutral, or oppose. To handle NA values in the TXT column, all NaN values were replaced with a space. This way the values could be processed, but they were not affecting the topic modeling. Since these comments were all written by humans, the data had to be cleaned before it was possible to perform any sort of topic modeling on it. The first step in preparing the comments was to remove all characters that were not letters from the comments and make all the letters lowercase. The next step in preparing the comments was to tokenize each comment,

Table 1. Number of Elections Held Per Year

Year	Number of Elections
2003	56
2004	283
2005	530
2006	737
2007	767
2008	453
2009	273
2010	176
2011	109
2012	85
2013	28

	DAT	RES	SRC	TGT	TXT	VOT	YEA
0	4/19/2013 23:13	1	Steel1943	BDD	'''Support''' as co-nom.	1	2013
1	4/20/2013 1:04	1	Cuchullain	BDD	'''Support''' as nominator.--	1	2013
2	4/19/2013 23:43	1	INeverCry	BDD	'''Support''' per noms.	1	2013
3	4/20/2013 0:11	1	Cncmaster	BDD	'''Support''' per noms. BDD is a strong contri...	1	2013
4	4/20/2013 0:56	1	Miniapolis	BDD	'''Support''', with great pleasure. I work wit...	1	2013

Fig. 1. Original Dataset Head

which essentially means each comment will be broken into individual words. After tokenizing the comments, the next step is to remove all the stopwords from the comments. Stopwords are words that will not be helpful in identifying topics of the comments, such as “i”, “a”, “is”, and “the.” [12] The nltk package has a list of stopwords that was imported, and then that list was compared against every word in the comments. [17] Any words that appeared in the stopword list and the comments were removed, so that the comments would be left with words that would be helpful in identifying topics in the comments.

The next step in preparing the comments is stemming the comments, which works to reduce words derived from the common root to a common stem. [12] This is useful for topic modeling because instead of having “contribution”, “contribute”, “contributes”, and all other related words show up separately, all related words will be reduced to a common stem, in this case “contrib.” There are multiple different types of stemmers that can be used to stem words, and two different stemmers were tested to see if either worked better on the comments. A SnowballStemmer and a PorterStemmer were applied to the comments, but there was very little difference in the topics that got pulled using the two different Stemmers, so the PorterStemmer ended up being used to stem the data. [17] After removing all characters that were not letters, changing all letter to lowercase, tokenizing the comments, removing stopwords, and stemming the comments, it is time to rejoin the words, so that each comment is one string, instead of multiple broken up words.

After getting the comments into a usable format, it was time to perform topic modeling on the comments. First, the comments were divided into three separate datasets: one containing all the

	TXT	LAL_TXT	TKN_TXT	STP_TXT	CLN_TXT	CLEAN_TXT	UNSTEMCLEAN_TXT
0	'''Support''' as conom.	support as conom	[support, as, conom]	[support, conom]	[support, conom]	support conom	support conom
1	'''Support''' as nominator.--	support as nominator	[support, as, nominator]	[support, nominator]	[support, nomin]	support nomin	support nominator
2	'''Support''' per noms.	support per noms	[support, per, noms]	[support, per, noms]	[support, per, nom]	support per nom	support per noms

Fig. 2. Cleaned Dataset for LDA

Table 2. Number of Comments in Each Category

Type of Comment	Number of Comments
Positive	144,451
Neutral	12,648
Negative	41,176

positive comments, one containing all the negative comments, and one containing all the neutral comments. The number of comments that fall in each type of vote are shown below in Table 2. The reason for splitting the comments into these three datasets was to perform topic modeling on each dataset to find the most common topics that appear in each set of comments. Now, it is time to perform topic modeling on the positive, neutral, and negative comments. Topic Modeling is an unsupervised learning technique that is used to understand and categorize large corpora of data. [15] The documents, in this case the comments, are clustered into groups based on their characteristics. [14] The end goal of topic modeling is to group the comments so that all comments in the group have the same topic. There are different approaches that can be used to perform topic modeling. After doing research, an LDA, Latent Dirichlet Allocation, was the method chosen to perform the topic modeling. According to Blei, Ng, and Jordan, "Latent Dirichlet Allocation is a generative probabilistic model of a corpus. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words." [16] For some context about the terminology used in LDA, documents are a sequence of words and a corpus is a collection of documents. [16] In Malik's words, "Documents are probability distributions over latent topics. Topics are probability distributions over words." [14] In more general terms, this model will say that documents with similar words have the same topic, and when documents have groups of words occurring together often, they usually have the same topic. [14] To begin topic modeling on the positive comments, the CountVectorizer class from sklearn was used to create a document term matrix, which is a matrix that holds the frequency of words in the documents, or positive comments in this case. [14] [18] In order to find topics within each type of comments, a LatentDirichletAllocation method from sklearn is run on the document term matrix with two parameters: n_components = 10, which is the number of topics to extract, and random_state = 0, which helps. [18] This LDA will pull 10 topics from all the positive comments, and then, the LDA model was manipulated to extract the top twenty words in each topic as well. After the LDA had found ten topics, the dataset with all the positive comments was iterated through and each comment got assigned a topic, based on which of the ten topics that comment had the most similarities with. The purpose of assigning a topic to each of the comments is that once all the comments are assigned a topic, exemplar sentences can be pulled from the dataset to give more context to the topics. From the twenty words assigned to a topic and the exemplar sentences, a broader understanding of the topics can be reached. This method was repeated for the neutral and negative comments as well to extract 10 topics for both neutral and negative comments.

3.2 Social Network Analysis

The other aspect of our dataset is focused on SRC, or source; TGT, or target; as well as columns correlated to the date, what the source voted target, and the result of that targets election. On the basic level we can create a social network and analyze the behaviors of Wikipedia Voters. This network can continuously increase in complexity. By introducing different colors and sizes of edges and nodes, we can introduce a deeper understanding to the network. Unfortunately, our dataset is quite large so to combat this in our network graphs we have decided to take a look into 2013 as it has the least number of elections, but is also the most recent. By focusing on a single year, we are able to individual elections much easier and focus on the importance of single individuals within the network. This allows us to analyze tendencies of voters as well as the outcome and influence around elections.

One of the more important ideas surrounding cascading depends on time frame. We use a focus group to analyze cascading. Our targets consist of only 'Banaticus', 'JasperDeng', 'Piotrus', 'Miniapolis', and 'Pjoef'. These five targets were chosen because they offer interesting findings which will be discussed later on. Miniapolis was the only target in this focus group that was voted to become an administrator. This was strategically chosen because we observed many people approved are nearly all voted yes by everyone. It is important to note that this alone may show some cascading because people follow there predecessors voting styles, but for a graphical standpoint having multiple approved does not show anything interesting.Therefore, for our graphs we tend to focus on cascading through a more voted no lens in our focus group. Also, we implemented a plotting mechanism which only takes active voters for our graphs. We define active as people who voted on at least ten candidates within 2013. Often times, people who vote on a smaller number of candidates have less of an impact on the overall outcome and little power. We were able to observe different graph statistics like degree, degree centrality, and betweenness centrality. These metrics allow us to differentiate the size of nodes within our system.

4 RESULTS

4.1 LDA

Ten topics were found for each the positive comments and each comment in that dataset was assigned a topic that explained that comment best. The amount of comments that fall into different topics can be seen below in Figure 3 for positive comments. In Figure 3, it shows that topic 4 and topic 8 have the most comments assigned to them. The top 20 words of Topics 4 and 8 are shown in Table 3. However, the words that appear in Table 3 are the words that occur with the highest frequency within that topic, and they do not offer much context about what kind of comments fall into that topic. To gain that context and develop a better understanding of the topics, 5 exemplar sentences were pulled from the comments that were assigned to that topic. Topic 4 was the most prevalent topic in the positive comments. The distribution of the 20 most common words in Topic 4, shown in Figure 4, shows that "support" appears about 7x more frequently than the next top word. After diving into the comments associated with Topic 4, many of the comments in this category are simply "Support" with no further explanation as to what makes this editor suitable to become an administrator. Of the five exemplar sentences pulled for Topic 4, shown below in Table 4, only one comment is not just "Support" and that comment refers to the editor being strong with a lot of good experience. Based on all the information collected on Topic 4, such as the top 15 words, the distribution of those words in the topic, and the exemplar sentences, this topic will be simply

Table 3. Top 15 Words in Positive Topics 4 and 8

Topic 4	Topic 8
support	support
strong	good
font	editor
answer	candid
per	look
good	admin
question	great
span	seem
nomin	user
color	like
contributor	make
course	fine
great	mop
excel	excel
user	luck

labelled as support. As for Topic 8, the 15 words that appear most frequently in that topic are shown in Figure 4, and the plot of how often each word appears is in Figure 5. The word "support" is again the top word, however, other words appear fairly frequently, such as "editor" and "good", which is not seen in Topic 4. To gain a better understanding of Topic 8, five exemplar sentences were pulled from the comments assigned to this topic. Using all the relevant information from Tables 3 and 4 and Figure 5, the support comments are slightly more detailed, but not by much. Most of the comments that get assigned to this topic are messages of support that include commenting on the strength of the candidate. To sum this topic up, it will be labelled as strength of candidate.

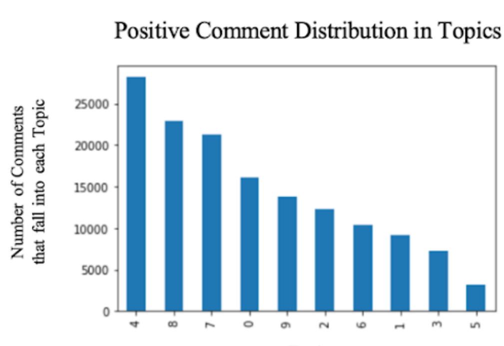
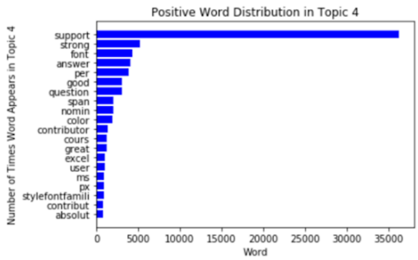
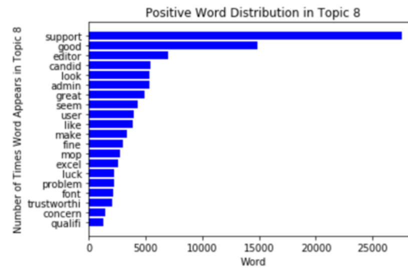


Fig. 3. Number of Postive Comments in Topics

Ten topics were extracted from all the neutral comments, and again, each comment was assigned a topic that matched it best. In Figure 6 below, Topics 5 and 6 are the most frequent topics that are assigned to comments. To find out why voters vote neutral on a candidate, it is necessary to dive into Topics 5 and 6. Table 5 shows the top 15 words in each topic, but it is hard to make sense of a topic from only these 15 words. Figure 7 shows the frequency of the top 20 words in Topic 6.



(a) Positive Word distribution in Topic 4



(b) Positive Word distribution in Topic 8

Table 4. Exemplar Sentences for Positive Topics 4 and 8

Topic 4	Topic 8
Support	Support. Great editor. Supported last time.
Support	Support. Great candidate on all fronts.
Support	No hesitation whatsoever. Very good candidate.
Support per noms. BDD is a strong contributor with thousands of edits and years of experience.	
I have no problem supporting	SO EASY support. Fantastic editor.

This shows which words are more important, or more common, within the topic and what kind of weight each word holds in the topic. Additionally, Table 6 shows some exemplar sentences that represent Topic 6 and offer a better understanding of what comments in Topic 6 look like. Combining all the information from Tables 5 and 6 and Figure 7, Topic 6 can be summed up as experience needed. The second most common topic is Topic 5, and the 15 most common words in that topic can be viewed in Table 5. The breakdown of word distribution in Topic 5 is shown in Figure 8, which shows which words appear very frequently in neutral comments in this topic. Seeing the most common words and the number of times those words appear in a topic offers only so much insight into what the topic actually is. The exemplar sentences, shown below in Table 6, give better context as to what the comments in this topic mean. Using the most common words in the topic, the distribution of those words, and the exemplar sentences, Topic 5 can be labelled as importance of edits over time.

Finally, ten topics of negative comments are extracted via LDA model, and each comment was labelled with one of the topics. From Figure 9 below, it can be seen that Topics 4 and 5 are the most prevalent in the negative comments. In an attempt to understand what makes voters vote oppose on a candidate, Topics 4 and 5 were explored further. From Figure 9 below, Topic 5 gets assigned to over 10,000 comments, and the most common words in Topic 5 are shown in Table 7. Figure 10 is shown below, and it shows the distribution of words that are in Topic 5. Along with Table 7 and Figure 10, Table 8, which shows the exemplar sentences that go with Topic 5, offers a deeper understanding of what kind of comments fit into this topic. All the information offered from Tables 7 and 8 and Figure 10 leads to declaring the topic is edits per month experience. As for Topic 4, the top words in the topic can be seen in Table 7 below, and how those words are distributed within the topic is shown in Figure 11. To get a better understanding of the words that make up a topic, Table 8 shows five exemplar sentences that are representative of the negative comments in that topic. Combining the information from Tables 7 and 8 and Figure 11, the topic can be summed up as lack of some Wikipedia skills.

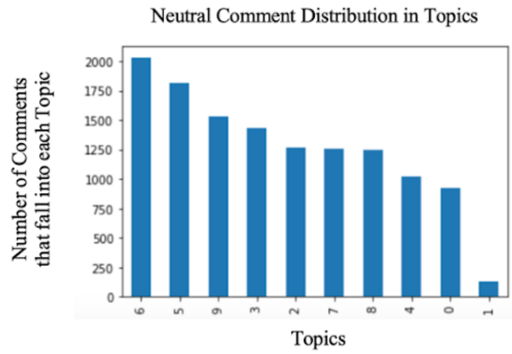
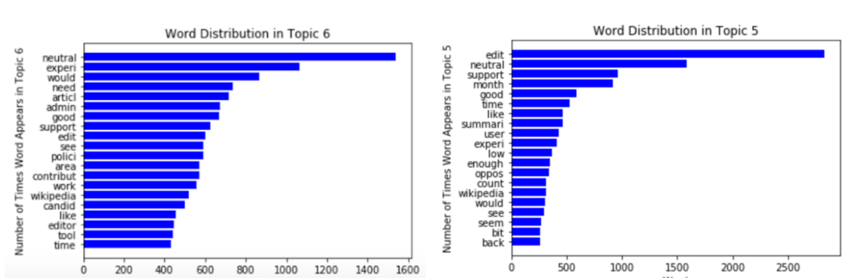


Fig. 5. Neutral Comments per Topic

Table 5. Top 15 Words in Positive Topics 5 and 6

Topic 5	Topic 6
edit	neutral
neutral	experience
support	would
month	need
good	article
time	admin
like	good
summary	support
user	edit
experience	see
low	policy
enough	area
oppose	contribute
count	work
Wikipedia	Wikipedia



(a) Word Distribution in Neutral Topic 6 (b) Word Distribution in Neutral Topic 5

Table 6. Exemplar Sentences for Neutral Topics 5 and 6

Topic 5	Topic 6
Neutral, far too many automated edits for me to support at this time.	Neutral. Looks like a trustworthy user who would make a good administrator in the future. If this candidate gets more experience, I would probably end up in the support column.
Neutral, for now. I would encourage keeping up the good work and trying again in a few months.	Neutral. You have been a net benefit to the project, but would like to see more experience before I support in the future.
Neutral. While I am unhappy with only 3.5 months can anyone highlight where this user's experience has been problematic. Otherwise, it looks as we are rejecting based on edit counts.	Neutral. Many edits but relatively few in areas where the admin tools would be used, and more experience in those areas would be preferred before getting the tools.
Neutral. Although in the past month, you have made a substantial amount of edits, it has only been 1 month. I would like to see more months of edits. I need a longer span of solid edits to decide whether you'll use the tools effectively and not abuse them.	Neutral, great article writer, but you need more policy experience.
Neutral. I'm sorry, but your edits and editing are just too inconsistent for me.	Neutral, more article edits will be better.

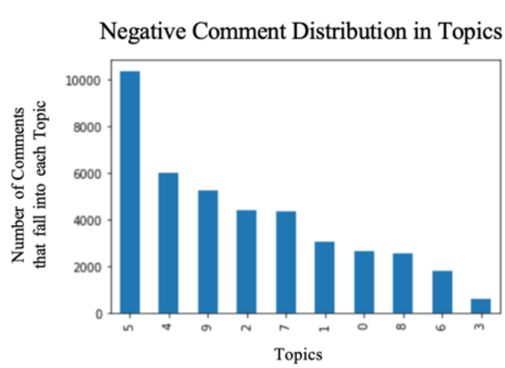


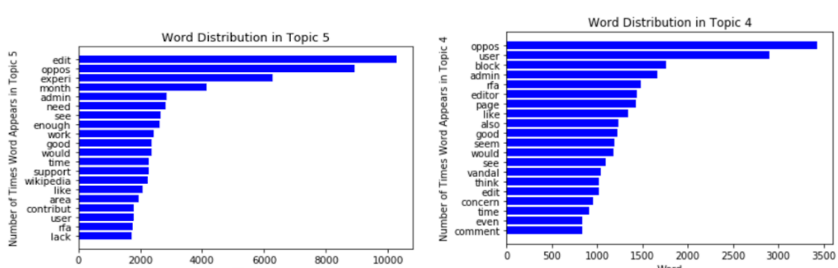
Fig. 7. Negative Comments Count per Topic

4.2 Social Network Analysis

The rise of social network plotting has been rapidly growing in recent years because it allows for researchers to visualize what is going on in a system in many ways simple raw data would not depict. In the following graphs the edge colors are based on the vote that a source passes to a target where Cyan represents 'yes', Orange represents 'no', and Pink represents 'neutral'. Similarly, we assign colors to nodes. A Red node means that the result of their election was denied becoming an administrator while a Green node means they were approved. There are some instances where we introduce pink nodes and this means that they are in the top 15 of voters based on betweenness centrality. Lastly, the size of nodes vary per graph but represent one of the previously mentioned

Table 7. Top 15 Words in Positive Topics 5 and 6

Topic 4	Topic 5
oppose	edit
user	oppose
block	experience
admin	month
rfa	admin
editor	need
page	see
like	enough
also	work
good	good
seem	would
would	time
see	support
vandal	Wikipedia
think	like



(a) Word Distribution in Negative Topic 5 (b) Word Distribution in Negative Topic 4

centrality metrics either betweenness or degree. Varying the size allows us to visually depict the importance of a node. Obviously the largest node will always be the one target nodes, but will be followed by the powerful voters.

As mentioned before having a full network is only useful to an extent. We can see in Figure 9 how the whole network is cluttered with including all voters even those who only vote on a single election. These voters are not that interesting and not much analysis can be done on them with regards to our focus and research questions. This graph has node sizes based on betweenness centrality so of course all the active voters will be centered in the middle as we can observe. So, why not only focus on these active voters? This is exactly what we do in Figure 10 which depicts all of 2013 elections again, but now only contains voters who voted on 10 or more elections in the calendar year.

Now this is much more interesting because we can begin to see how voters are voting as well as clusters surrounding the targets, or candidates rather. Table 7 explicitly shows the top voters based on betweenness centrality. It is apparent that on the top side and central a lot of the targets are green meaning they were approved to become adminstrators, but it is interesting to see how Dirtylawyer-1 and Carrite both seem to infiltrated the mostly yes votes. Many of the reasons seem to be based on experience or lack there of as we have analyzed previously in the LDA. Also, the

Table 8. Exemplar Sentences for Negative Topics 4 and 5

Topic 4	Topic 5
Oppose: The third RFA for only having 2,673 edits is quite worrying.	Oppose for now. Needs more experience.
Oppose, self-acknowledgement that the editor has not and will not be active. The last thing we need is another inactive admin.	Sorry, you've got nowhere near enough experience to gain the community's trust right now. But please try again later.
Oppose, user doesn't use edit summaries and doesn't appear to spend much/any time on talk pages. Certainly a valuable contributor, but being an admin is mostly about conversations, not mainspace edits.	Oppose. Sorry, but 1,184 edits in 22 months is simply NOT enough experience. Not enough experience in the areas candidate wants to use the tools in, plus the candidate's edit summary usage could be better.
Oppose, per Realist2. The last RFA was just held two and a half months ago and content building and activities in Wikispace are more required.	Oppose, come back when you have a few more months of edits under your belt. If possible, ask a more experienced editor to nominate you.
Oppose. User needs to show a greater understanding of Wikipedia policies (usage of cool-down, blocks, etc.)	Only 103 edits made in the last 6½ months, which includes a long layoff. Answers don't demonstrate why they would need the tools at this point.

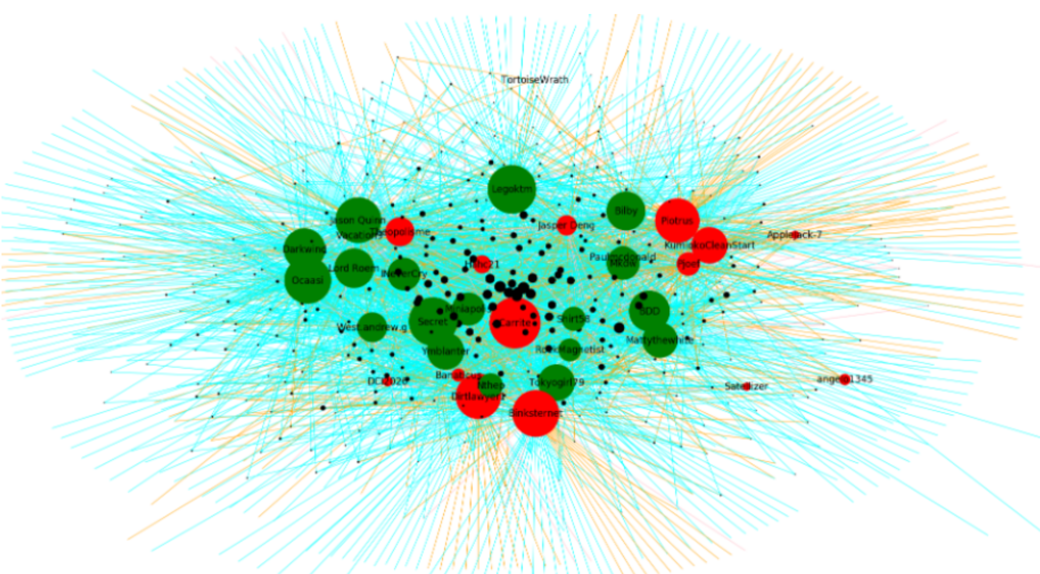


Fig. 9. Whole Network for 2013

green nodes seem to be larger on average. This could signify that people are more likely to vote yes on a good candidate than no on a bad one.

By limiting down the voters we can began to analyze the importance. It is intriguing that the top voters are now much more on the outside and are often times connected to the stragglers that were voted no on the outside. This can signify that the top voters also tend to vote for smaller candidates

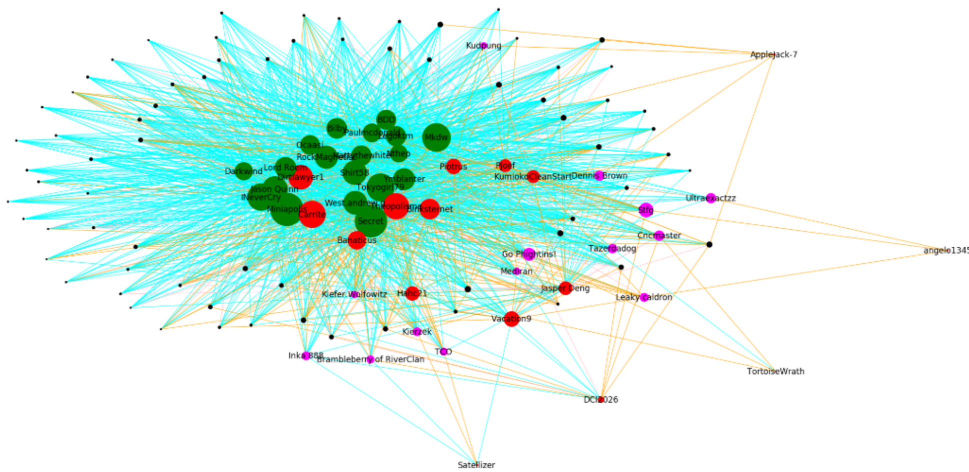


Fig. 10. 2013 Network, Only Active Voters

Table 9. Top Voters Betweenness Centrality

Voters	Betweenness Centrality
Stfg	0.01018339
Go Phightins!	0.007933629
Cncmaster	0.005040844
Ultraexactzz	0.004990045
Dennis Brown	0.004881837
Tazerdadog	0.00419024
Leaky caldron	0.003811369
Kierzek	0.003568781
Inka 888	0.003540863
Brambleberry of RiverClan	0.002992307
TCO	0.002820304
Mediran	0.002388095
Kiefer.Wolfowitz	0.002324188
Kudpung	0.002104615

which even goes further to show that they are invested in the voting process. Taking a look at the far right of the network in Figure 10 at the three lines going to angelo1345. We can see they are connect back to Ultraexacts, Stfg, and Vacation9. There are a few takeaways namely that smaller elections usually end in a denial of adminship. Also, the real fascinating node link is Vacation9 because this suggests that Vacation9 is also an active voter. This opens up a new discussion of users voting on other elections to be more active in the community to in hopes of becoming elected themselves. The more active a voter is it seems the more variance in there votes they have. A lot of the outside nodes on the top left are all blue. They maybe just coming in for backup of larger voters to get somebody approved. Something similar to this happens to politcal elections in the government today. Many politicians just need to get a figure head approval and then will get blind followers.

Now for cascading The most important factor is what happened at time t-1 and t. In our case, we have the timestamp of when user voted on a target. Focusing on a simple example displaying cascading we can see Figure 11. Taking the target of Jasper Deng, it is apparent that almost all the votes in the beginning of his election are blue representing yes. As soon as a single no vote enters the graph there is a lot more that follow.

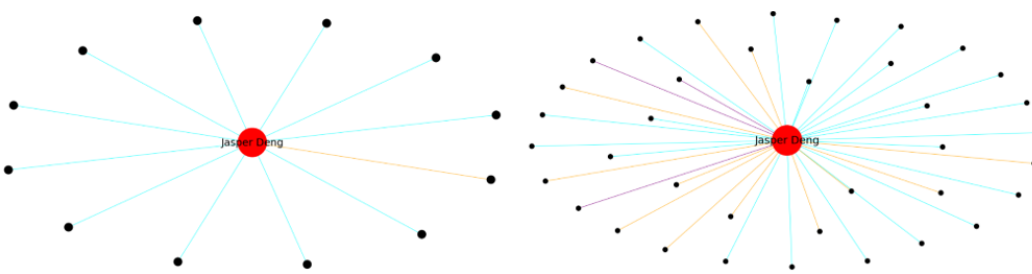


Fig. 11. Jasper Deng at time t-12, time t

For the focus group we had targets consisting of only 'Banaticus', 'JasperDeng', 'Piotrus', 'Miniapolis', and 'Pjoef'. This graph is depicted in Figure 12. We can see how there tends to be a cluster between Piotrus and Pjoef as well as a cluster by Banaticus and Jasper Deng. Miniapolis being the sole approved target is more on its own. It is interesting to see that most of the people who voted yes to Miniapolis tend to vote yes to the others as well. There are exceptions of course, but it also seems that most people who voted no or neutral to the others stuck with that regardless of the target.

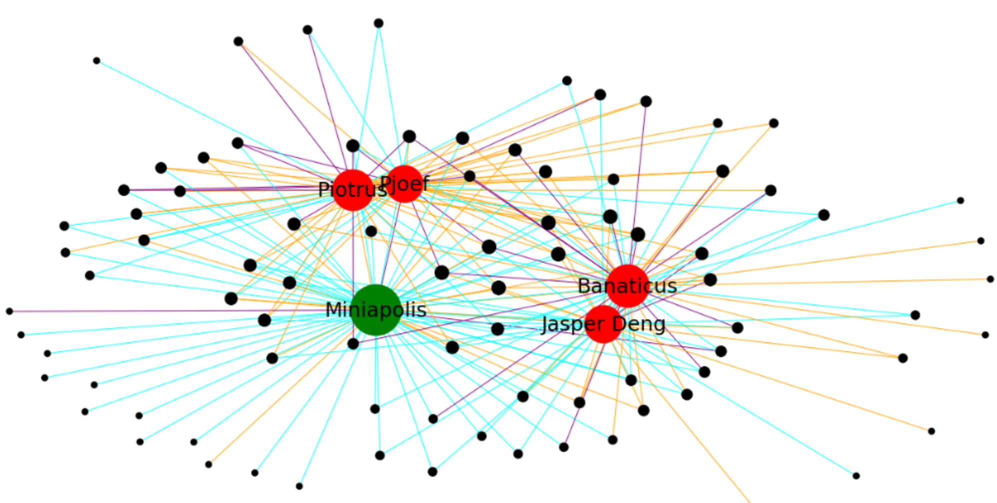


Fig. 12. Jasper Deng at time t-12, time t

5 DISCUSSION

The two most prominent positive topics were Topics 4 and 6. In Topic 4, the majority of the comments were only one word, “Support.” Many of these voters offered no further explanation as to why they supported these users, just that they did. As for Topic 6, different positive adjectives often appeared in these comments, raving about the candidate or their skills as an editor. However, no thorough explanations were left that explained what prompted these voters to support this candidate. One reason for this could be that voters who are supporting a candidate do not feel the need to explain why they support, and it is not as important to defend support of a candidate as it is to defend opposition of a candidate. The topics extracted from the positive topics were fairly standard and not very in depth, due to the lack of explanation left with supporting comments.

As for the neutral topics, Topics 5 and 6 were the most representative of the neutral comments. Topic 6 of the neutral comments contained mostly comments about candidates needing to gain more experience on some or all areas of Wikipedia before the voter would be able to support. Topic 5 was comprised of comments about how important it was to have a large number of edits over a long period of time. This topic shows the importance of edits in applying for administrator role. Also, many of the comments that are assigned to this topic say that the candidate has an inadequate number of months under their belt. It seems that, in general, voters who voted neutral believed that the candidates needed more experience in one way or another to become an administrator.

The topics that dominated the negative comments were Topics 4 and 5. Topic 5 primarily focuses on candidates not having enough edits per month. This seemed like a metric that was important to voters when deciding whether a candidate was fit to become an administrator. Topic 4 focuses on a lack of some Wikipedia skills. These comments all focus on candidates that still lack skills in certain areas of Wikipedia as being an admin comes with many responsibilities across the site. Many of these comments talk about candidates that have enough skill in some areas of Wikipedia, but need to improve in other areas before voters could support the candidate.

From looking at the topics extracted from the positive, neutral, and negative comments, it should be noted that voters who voted neutral or negative on a candidate often wrote their reasoning in the comment section. However, the same cannot be said for voters who supported a candidate, many of the positive comments lacked any explanation of the skills or characteristics that resulted in a vote of support. Part of the reason that neutral and negative comments had more detailed comments could be the voters trying to influence other voters by voicing their concerns. Since all votes and comments can be viewed by other voters, we have to assume that voters are reading other voters’ reasons for not supporting a candidate and that could influence their vote.

The purpose of finding the most frequent topics in the positive, neutral, and negative comments is to give us further insight into what skills and traits voters are looking for in candidates. As stated earlier, positive comments do not offer much explanation as to what makes a good administrator. However, by looking at what skills candidates are lacking from the neutral and negative comments, we can deduce what skills are considered important for an administrator on Wikipedia. From the topics derived from neutral comments, we can see how crucial experience is and how voters think highly of candidates that edit often over a long period of time. The negative topics that were pulled from the comments highlight the need for candidates to have a sufficient number of edits per month and a lack of skills in some areas of Wikipedia. Many voters used the edits per month metric as a measure of the level of activity the candidate had, and voters are adamant about keeping inactive users off of Wikipedia. Being an administrator comes with many responsibilities, including, but not limited to, editing pages, contributing to articles, fighting vandalism, and using the admin tools to delete pages, block users, and protect pages. Many of the voters were concerned that candidates were not well-versed enough in all the different areas of being an administrator. The logic behind this is that if accepted as an administrator, a candidate is given access to the admin tools. If a candidate could not balance the different responsibilities before, they will not be able to handle those duties on top of the responsibilities that come with the tools.

Thus, voters are looking for experienced candidates who contribute to articles, edit pages, and fight vandals and have frequent and consistent edits over a relatively long period of time. Clearly voters are looking for a lot out of candidates, however, this is not unreasonable due to the power the administrators receive when given access to the admin tools.

We were able to split up voters into two categories active and inactive. Active voters were classified as people who voted more than ten times in a calendar year. While this was an arbitrary number chosen, it allowed us to further analyze our graph visually. As we saw in the graph of the whole network, there tend to be many voters who simply vote on a single candidate. There are a few things wrong with this. First, when analyzing a graph it tends to be skewed towards the center of active nodes in the environment. Therefore, our whole network for 2013 is much more difficult as it looks like a cluster of active voters surrounded by the targets they voted for. This is why we chose to split up the graph into active voters specifically.

When using the betweenness centrality to appropriately size nodes, we were able to determine who these prominent voters are. For instance, Stfg in Figure 10 and Table 7 have a betweenness centrality of .01 which was the highest by 20% compared to other voters. This suggests that Stfg is a information broker and bridge. It also means the Stfg participates in a lot of the elections. Its important to note that there are others close to him such as Cncmaster and Tazerdadog. All three of them tend to vote on smaller elections as well. One thing about these active voters especially the top ones that are colored pink are they don't only vote yes like less influential people. We can takeaway that the more active and involved a voter is the more likely they are to vote fairly and continue to vote on elections that are smaller than others.

When looking at Figure 10 we can clearly see that all the green nodes signifying approved targets are clearly grouped together. There are no outside the center targets that are green. This can suggest that cascades do exist and when an election begins and tends to get a lot of votes they are more likely to be green. Cascading maybe happening because the larger an election gets smaller active voters assume the previous ones are correct with there reasoning. People tend to be more passionate about things they dislike. This is very apparent in the responses in yes vs no topics as discussed in the LDA sections. Many time a person that votes yes reasoning will simply be "Support". This response means nothing other than blindly following aka Cascading.

Cascading is further seen in our focus group specifically around Jasper Deng, even reverse cascading can be observed. While we show only two time instances in this paper analyzing every time instance surrounding Jasper Deng election depicts cascading more. His first eight votes were yes and as soon as one no was introduced it was like a plague and quickly turned anything but yes and he was voted no.

6 LIMITATIONS AND FUTURE WORK

We acknowledge that we are working with a data set spanning a decade thus vulnerable to redundancies and duplication, if not tripling, of RfAs and their outcomes. During our analysis, we noticed certain candidates had submitted RfAs multiple times, and it would be interesting to study if and how the voting changes over time for those candidates. Looking at how the votes and comments change for candidates who submit multiple RfAs would offer unique insight into the growth voters are looking for when they originally oppose a candidate.

We also did not link the analysis of information cascades to that of the comments. In future studies, much insight can be derived in analyzing comments of voters casting the first opposing vote and breaking the information cascade. By looking at the comments of each vote once an information cascade had started, research would be able to pinpoint how much the start of the cascade affected subsequent voters. Such research will contribute to the body of work explaining motivation to express opposing views in social media.

Cascading also gets intriguing when trying to determine where a new node entering the graph will go. For our research this would be interesting to determine the probability of a node voting on each of the targets still getting voted on and then determining what is the probability of the new node voting yes, no, or neutral for each target. We tend to think that if they are likely to enter in and vote for a new election they are more likely to vote yes for a mainly yes target than no for a mainly no target. This is because from what we have observed there is less volatility in elections that end in an approval.

7 CONCLUSION

We found supportive answers to both research questions. Firstly, information cascades are apparent in the study of the dynamics of the election process. Moreover, just like in the work by Leskovec et al., the cascade may emerge from a variety of patterns of preceding votes. Secondly, we see good consistency in the content and connotation of voters' comments. For example, the intent of negative comments appears to converge on the issue of the lack of experience expressed as the ratio of edits over time. Positive comments share a common thread of support due to sufficient experience. These findings diverge with the previous research on predictors of successful elections which indicate that the number of edits was a weak predictor. We, however, have concluded that experience measured in the number of edits was a critical attribute of success.

To the bigger question with which we open the paper, i.e., the challenge to Wikipedia to maintain integrity and quality of the content by permitting only most qualified members to manage the content. The analysis of information cascades suggests that there is bias among voters, evident in the form of pursuit of personal interest. The text analysis, on the contrary, points in the direction of the search for objectivity as RfA contributors exchange observations about candidates' traceable experience. However, we posit that that this search is meek as RfA contributors never operate with common thresholds for performance thus inevitably injecting personal bias in the assessment of a candidate. We never see any adopted threshold of performance against which candidates are measured. Thus, our study suggests that Wikipedia may benefit from reviewing policies of the transparent voting process to increase individual accountability of RfA contributors.

REFERENCES

[1] Asim, Y., Niazi, M., and Malik, A. K. Personal vs. Know-How Contacts: Which Matter More in Wiki Elections. Islamabad, Pakistan. (2018). Retrieved November 27, 2019 from https://arxiv.org/ftp/arxiv/papers/1804/1804.07450.pdf

[2] Burke, M., and Kraut, R. Mopping up: modeling wikipedia promotion decisions. San Diego: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work. (2008). Retrieved November 27, 2019 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.141.7271rep=rep1type=pdf

[3] Burt, R. S. Neighbor Networks: Competitive Advantage, Local and Personal. Unpublished. (2009). Retrieved November 27, 2019 from https://faculty.chicagobooth.edu/ronald.burt/research/files/NNintro.pdf

[4] Desai, N., Liu, R., and Mullings, C. Result Prediction of Wikipedia Administrator Elections based on Network Features. Stanford University. (2014). Retrieved November 27, 2019 from http://cs229.stanford.edu/proj2014/Nikhil%20Desai,%20Raymond%20Liu,%20Catherine%20Mullings,%20Result %20Prediction%20of%20Wikipedia%20Administrator%20Elections%20based%20ondNetwork%20Features.pdf

[5] Easley, D., and Kleinberg, J. Information Cascades. In Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press. (2010).

[6] en.wikipedia.com. (2019). Retrieved from Wikipedia pages.

[7] Leitch, T. Knowledge, Authority, and Liberal Education in the Digital Age. Johns Hopkins University Press. (2014).

[8] Leskovec, J., Huttenlocher, D., and Kleinberg, J. Governance in Social Media: A Case Study of Wikipedia Promotion Process. Cornell University. (2010). Retrieved November 27, 2019 from https://www.cs.cornell.edu/home/kleinber/icwsm10-govern.pdf

[9] Simonite, T. The Decline of Wikipedia. MIT Technical Review. (2013). Retrieved November 27, 2019 from https://www.technologyreview.com/s/520446/the-decline-of-wikipedia/
[10] Wu, F., and Huberman, B. A. How Public Opinion Forms. International Workshop on Internet and Network Economics. (2008). Retrieved November 27, 2019 from https://link.springer.com/chapter/10.1007/978-3-540-92185-1_39
[11] Leskovec, J. and Krevl, A. (June 2014). Retrieved December 7, 2019 from https://snap.stanford.edu/data/wikipedia
[12] Allen, M. 2018. 102: Pre-processing data: tokenization, stemming, and removal of stop words (compressed code). (December 2018). Retrieved December 7, 2019 from https://pythonhealthcare.org/tag/nltk/
[13] Boujon, C. 2019. How to list the most common words from text corpus using Scikit-Learn? (September 2019). Retrieved December 8, 2019 from https://medium.com/@cristhianboujon/how-to-list-the-most-common-words-from-text-corpus-using-scikit-learn-dad4d0cab41d
[14] Malik, U. 2019. Python for NLP: Topic Modeling. (April 2019). Retrieved December 8, 2019 from https://stackabuse.com/python-for-nlp-topic-modeling/
[15] Saxton, M. A Gentle Introduction to Topic Modeling Using Python. Retrieved December 8, 2019 from https://theolib.atla.com/theolib/article/view/506
[16] Blei D., Ng, A., and Jordan M. Latent Dirichlet Allocation. Journal of Machine Learning Research 3 (January 2003). Retrieved December 8, 2019 from http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf
[17] nltk, http://www.nltk.org
[18] sklearn, https://scikit-learn.org/stable/
[19] pandas, https://pandas.pydata.org
[20] matplotlib, https://matplotlib.org