Exploring The Motivations Behind Voting For Wikipedia Adminship*

MD ROMAEL HAQUE, Marquette University,
MICHAEL FUSS, Marquette University,
MATTHEW J RUBIO, Marquette University,

This paper describes the motivations working behind voting for Wikipedia adminship. We used publicly available data set which was created by Jure Leskovec for Stanford University. The set is a request for amidships for Wikipedia. Throughout our analysis we first observed how voting community developed over time through social network visualization. Then we applied topic modeling into voters' comments and contextualize that with thematic analysis in search of the motivations. Motivations for supportive voters include trust, organizational commitment and personal relationship. On the other hand motivations for non-supportive comments include doubt and experience valuation. later we connected our results with broader implications of "Unified theory of technology use" around voters' behavioral patterns.

CCS Concepts: • **Data Mining → Social Network Analysis**; *Topic Modeling*; Behavioral Pattern.

ACM Reference Format:
MD Romael Haque, Michael Fuss, and Matthew J Rubio. 1. Exploring The Motivations Behind Voting For Wikipedia Adminship. *Proc. ACM Hum.-Comput. Interact.* 1, 1, Article 1 (January 1), 13 pages.

1 INTRODUCTION

Social network analysis is an approach and set of techniques used to study the exchange of resources among actors (i.e., individuals, groups, or organizations). One such resource is information. Regular patterns of information exchange reveal themselves as social networks, with actors as nodes in the network and information exchange relationships as connectors between nodes. Just as roads structure the flow of resources among cities, information exchange relationships structure the flow of information among actors. Social network analysis assesses information opportunities for individuals or groups of individuals in terms of exposure to and control of information. By gaining awareness of existing information exchange routes, information providers can act on information opportunities and make changes to information routes to improve the delivery of information relationships. [20]

Information relationships indicate what kinds of information are being exchanged, between whom, and to what extent. The pattern of relationships between actors reveals the likelihood that individuals will be exposed to particular kinds of information, and the likelihood of their considering that data to be authoritative. Patterns of forwarding and receipt describe networks that

*Data Mining Class Project - Fall 2019

Authors' addresses: MD Romael Haque, Marquette University, ; Michael Fuss, Marquette University, ; Matthew J Rubio, Marquette University, .

2573-0142/1/1-ART1 $15.00
https://doi.org/

show how information moves around an environment, and how actors are positioned to facilitate or control the information flow.

Even though social network analysis can put great visualizations of information relationships, sometimes that need to be inferred more broadly. And to find a hidden reasons or motivations for that relationship, topic modeling is a great way to achieve that goal. Topic modeling is a type of statistical modeling for discovering the abstract "topics" that occur in a collection of documents. It is a type of algorithm that scans a set of documents, examines how words and phrases co-occur in them, and automatically learns groups or clusters of words that best characterize those documents. These sets of words often appear to represent a coherent theme or topic [26]..

This paper describes the motivations working behind voting for Wikipedia adminship. We used publicly available data set which was created by Jure Leskovec for Stanford University [1]. The set is a request for amidships for Wikipedia. Throughout our analysis we first observed how voting community developed over time through social network visualization. Then we applied topic modeling into voters' comments and contextualize that with thematic analysis in search of the motivations. Motivations for supportive voters include trust,organizational commitment and personal relationship. On the other hand motivations for non-supportive comments include doubt and experience valuation. later we connected our results with broader implications of "Unified theory of technology use" around voters' behavioral patterns.

2 RELATED WORK

We did a background study around social network visualization and community detection, topic modeling - their history and applications in various aspects. We also did a brief background research on "Social Influence Theory" to explain the voters' motivational patterns.

2.1 Social network visualization and community detection

Visual analysis of social networks is an integral component of the field of social network analysis [19]. Visualizing community structures present in social networks and identifying people who play important roles within a network can reveal interesting information specially by exploiting the temporal evolution of relationships. Social networks can exhibit temporal dynamics in a number of ways. The instances in the data may appear and disappear over time whereby different time windows may exhibit different characteristics. For example, a person might change his affiliation with a business organization by joining a different business enterprise and developing new social ties within this new environment. Moreover, the relationships may represent events and associations that are significant at a particular point of time, such as new job opportunities, or the establishment of a new business organization [7, 10, 22, 32]. If this is the case, then the temporal dimension associated with these events play a key role to capture important information. These visualizations have attracted much interest as images of social networks have provided investigators with new insights about these networks [20].

Community detection in social networks has attracted lots of attention in the domain of sociology. A more generic formalism for the term community is the term cluster. Sociologists use the term community [6] as compared to the statistical and data mining domain where people use the term cluster [35] to refer to the same concept. A cluster might not necessarily represent a community but throughout this paper, we use the terms interchangeably to refer to the same concept. Several surveys [8, 25] are available addressing the clustering or community detection problem. Some approaches [5, 21, 31] have performed better than the others for the discovery of communities in social networks. Researchers have also shown interest in discovering changing clusters in dynamic data [27] and clustering the evolving data streams [2]. However, these techniques are either insufficient or inefficient to characterize the changes in community structures. Since the interactions taking place between individuals can be characterized by a single relationship (for example: a weighted edge), interactions between communities inherit a number of ways that can establish an interaction between two communities over the passage of time. Since most of the existing techniques are adapted to handle changes occurring in individuals rather than communities, the goal of our approach is clearly different from others.

These systems perform well to exploit the temporal dimensions of a dynamic network focusing on changes and transitivity of individuals or their relationships. The system helps to discover structural changes in the entire network by studying the evolution of communities and the goals are clearly different from the other systems presented in this section.

2.2 Topic modeling

A topic model is a kind of a probabilistic generative model that has been used widely in the field of computer science with a specific focus on text mining and information retrieval in recent years. Since this model was first proposed, it has received a lot of attention and gained widespread interest among researchers in many research fields. So far, besides text mining, there also have been successful applications in the fields of computer vision [17], population genetics, and social networks [26].

The origin of a topic model is latent semantic indexing (LSI) [16]; it has served as the basis for the development of a topic model. Nevertheless, LSI is not a probabilistic model; therefore, it is not an authentic topic model. Based on LSI, probabilistic latent semantic analysis (PLSA) [23] was proposed by Hofmann and is a genuine topic model. Published after PLSA, latent Dirichlet allocation (LDA) proposed by Blei et al. [11] is an even more complete probabilistic generative model and is the extension of PLSA. Nowadays, there is a growing number of probabilistic models that are based on LDA via combination with particular tasks. Latent Dirichlet Allocation (LDA) is an example of topic model and is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distribution [11].

Topic models have many applications in natural processing languages. Many articles have been published based on topic modeling approaches in various subject such as Social Network, software engineering, Linguistic science and etc. There are some works that have focused on survey in Topic modeling. In [14], the authors presented a survey on topic modeling in software engineering field to specify how topic models have thus far been applied to one or more software repositories. They focused on articles written between Dec 1999 to Dec 2014 and surveyed 167 article that using topic modeling in software engineering area. They identified and demonstrate the research trends in mining unstructured repositories by topic models. They found that most of studies focused on only a limited number of software engineering task and also most studies use only basic topic models. In [15], the authors focused on survey in Topic Models with soft clustering abilities in text corpora and investigated basic concepts and existing models classification in various categories with parameter estimation (such as Gibbs Sampling) and performance evaluation measures. In addition, the authors presented some applications of topic models for modeling text corpora and discussed several open issues and future directions.

2.3 Planned behavior and user influence

There are several theories around how users' behavior got influenced. One of the highly evaulated and well recognized is Kelman's "Social Influence Theory" [29]. The central theme of social influence theory, as proposed by Kelman, is that an individual's attitudes, beliefs, and subsequent actions or behaviors are influenced by referent others through three processes: compliance, identification, and internalization. Kelman posited that social influence brings about changes in attitude and actions, and that changes may occur at different "levels". This difference in the level of changes can be

Social Network Statistics	
Nodes	10,835
Edges	159,388

Table 1. Network description of the data set

attributed by the differences in the processes through which individuals accept influence. Social influence is a pervasive force in human social interaction. In many social encounters, individuals modify their opinions, attitudes, beliefs, or behavior towards resembling more those of others they interact with. Individuals are socially influenced because they are persuaded by convincing arguments [13], because they seek to be similar to others [3], because they are uncertain about a decision and follow the lead of others [9], or because they feel social pressure to conform with social norms [18, 37].

Despite much research, social influence remains one of the most puzzling social phenomena. On the one hand, empirical studies across a variety of areas have documented how social influence reduces differences between people, as has been found in experiments on conformity [4], research on small group behavior [24], persuasion [13], innovation diffusion [30], the influence of mass media [28] or online social networks [12]. On the other hand, there is a long-lasting debate about the complex dynamics that social influence in interpersonal interactions generates on the collective level (Mason, Conrey and Smith 2007). For one thing, while assimilation seems to be the predominant pattern in interpersonal interactions, people may not only influence each other to become more alike, but also sometimes reject attitudes or behavior of those they interact with, and even seek to become more different from them [33]. However, there is much uncertainty about the exact conditions and mechanisms that elicit assimilation or differentiation in interpersonal influence [34], and about how these processes recombine in generating opinion dynamics at the macro-level of groups, organizations, or societies at large.

2.4 Research problem

Our research problems involves around whether or not voters got influenced by "Social Influence Theory". Through the course of time, do voters' motivation to vote for a particular candidate change seeing what others are voting for. If not then what are the reasons that voters to vote in such a transparent platform - vote for Wikipedia adminship.

3 METHODS

We laid out the methods in several sections. First we described the data set, then we describe how we pre-process the data set. Lastly, we gave a detail description of methodologies of our analysis. The analysis has been divided into two parts - social network visualization and topic modeling.

3.1 Description of the data set

We used publicly available data set which was created by Jure Leskovec for Stanford University [1]. The set is a request for adminships for Wikipedia. A request can be submitted by a candidate or by another member in the community. The votes casted are supporting, neutral, or opposing. This set contains 11,381 users forming 189,004 voter vote pairs. There is a total of 198,275 votes. This data set also contains comments of why the voter choice whether to support, be neutral, or oppose the votee. network description has been provided in table 1.

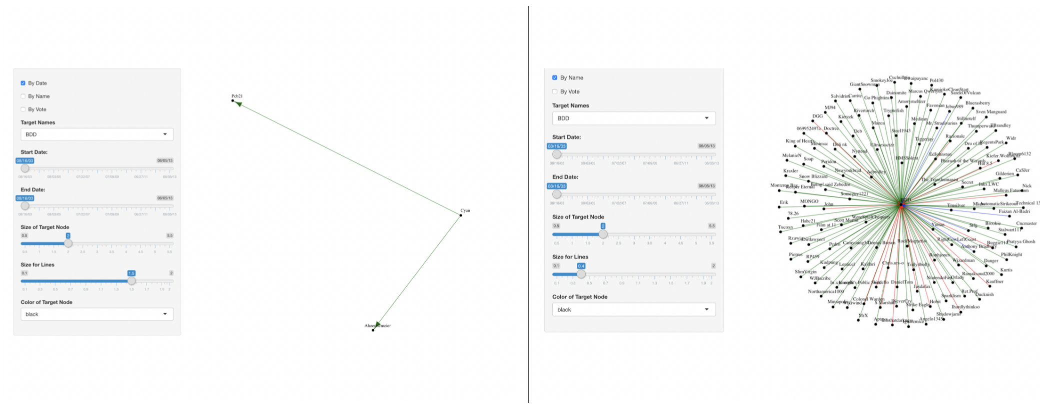

Fig. 1. **On the right hand side, the graph shows by date. This is the first date of the data set. On the left hand side, this is BDDs' personal network of everyone that had voted and how each person voted.**

3.2 Data preprocessing

The data set was composed of seven different columns: source, target, vote, result, year, date, and text. Target is the noted to be the person that is already in the network and voting to have the target join the network. Vote is noted to be what the source voted for the target. Possible votes are positive one, zero, and negative one. Positive one means that the source agrees for the target to join, zero means that the source is neutral if the target joins, and negative one means that the source does not agree with the target joining the network. Result is noted to be whether the target was voted into the network or not and is represented the same way as the votes column. Year is the year in which the vote was given. Date is a time stamp noted as "d m,yyyy" and also time of the day that it was posted. Text is then composed of whether the source voted to support, neutral, or negative and the reason why they had voted that way.

After gathering the data set and understanding it, cleaning the data was next. The first step was to pick the columns that contained the most information. The columns that were most valuable where source, target, vote, and date. Date had been parsed and manipulated to "mm/d/yyyy". After selecting the columns that will be used to visualize the network, date had to be manipulated. Taking the time stamp out of the date column to make it easier to work with when visualizing the data.

3.3 Analysis

The analysis consist of two part.
1. Social network visualization
2. Topic modeling

3.4 Social network visualization

Next, we tried to find a way to visualize the set, but wanting the visualization to be dynamic to see how the graph changes over time. The way this can be accomplished is by using R and using the built in Shiny app package. The interactive app will allow for users to visualize the network by date, name, and vote probability which can be seen in figure 3. The first three check boxes allow the user to interact and choose what is viewed in the right-hand side.

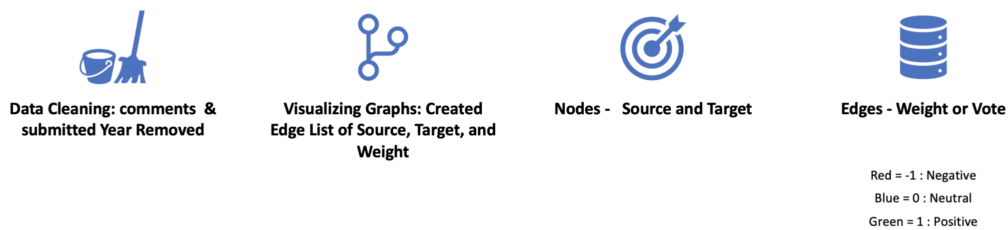

Fig. 2. **An overview of the methodology for social network visualization**

Visualizing the network by date allows for seeing the network from a selected start date and end date. Figure 1 is the control panel for viewing the graph. It allows the user to pick the range of dates, the name of the person of interest, the overall and daily probability of the votes. In figure 1, sliders are used to pick the start and end date. A directed graph is then shown in the area to the right of the selectors. One can visually see how sources and targets are connected to each other. In figure 1, connections are noted by green, blue, and red connections. These colors symbolize whether the source voted positive, neutral, or negative. Another way to visualize the network is by target. This allows to see the targets individually overall network which in the selectors, is the "Target Names". Here the targets full network will be show and how the person voted on them. Lastly, see the probability, as seen in figure 1 of a person being added to the network. Using the sliders, one can visually see how the probability changes are time goes on. While being able to see the overall frequency of the votes, one can also view the daily frequency over time. An overview of the methodology for social network visualization has been described in figure 2.

3.5 Topic modeling

Text documents require preprocessing before mining techniques can be applied. Preprocessing techniques include tokenization such as removing of numbers, punctuation marks, special characters etc. and stop word removal. While removing stop words, standard dictionary stop words and domain specific words can be removed.

For our project, we used of of the most popular method - LDA. LDA is a generative probabilistic model for collections of discrete data such as text corpora. It is a three-level hierarchical Bayesian model that can infer probabilistic word clusters, called topics, from the document-word matrix cite. LDA models each document as a mixture of topics and the model generates automatic summaries of topics in terms of a discrete probability distribution over words for each topic, and further infers per-document discrete distributions over topics [].

LDA is a distinguished tool for latent topic distribution for a large corpus. Therefore, it has the ability to identify sub-topics for a technology area composed of many patents, and represent each of the patents in an array of topic distributions. With LDA, the terms in the set of documents, generate a vocabulary that is then applied to discover hidden topics. Documents are treated as a mixture of topics, where a topic is a probability distribution over this set of terms. Each document is then seen as a probability distribution over set of topics. We can think of the data as coming from a generative process that is defined by the joint probability distribution over what is observed and what is hidden.

We propose a structured mioxed method approach to the analysis of large document collections. It is conducive to validating and enhancing the meaningfulness of individual topics generated by

Fig. 3. **An overview of the methodology for topic modeling and thematic analysis**

Steps	Description
Step A	Each topic model-based analysis begins by creating a large textual qualitative data set, i.e. a large document collection to be analyzed.
Step B	The given document collection is subject to two subsequent processes of de-contextualization. First, the documents are pre-processed into a term-document matrix. This pre-processing results in a loss of word context as it treats documents as bags of words. Second, the topic model itself further detaches words from their original context in order to arrive at topics suitable to model the entire document collection. At this stage, quantitative topic modeling would end, in most cases, and the resulting topics may be given ad-hoc labels.
Step C	The topic model components form the basis of our proposed design. Using both qualitative and quantitative methods, we propose an approach to the validation and enhancement of individual topics found in the document collection. In particular, this approach targets the re-contextualization of the previously decontextualized data. Compared to a purely quantitative topic modeling approach, it allows us to draw more meaningful inferences.
Step D	The research process finally yields recontextualized output, namely the key themes of the document collection. As detailed below, these themes are presented by the developed final pattern codes.

Table 2. **A step by step description of the methodology for topic modeling and thematic analysis**

topic models, and hence to fathoming the meaning of the analyzed document collection at large. The research process consists of four steps, which are shown in Figure 3 and Table 2.

4 RESULT

We crafted our results in two ways. First we described the community and voter's behavioral analysis through a certain time frame by building a dynamic social network visualization and then we reasoned the patterns of behavior through a mixed method analysis of topic modeling and thematic analysis.

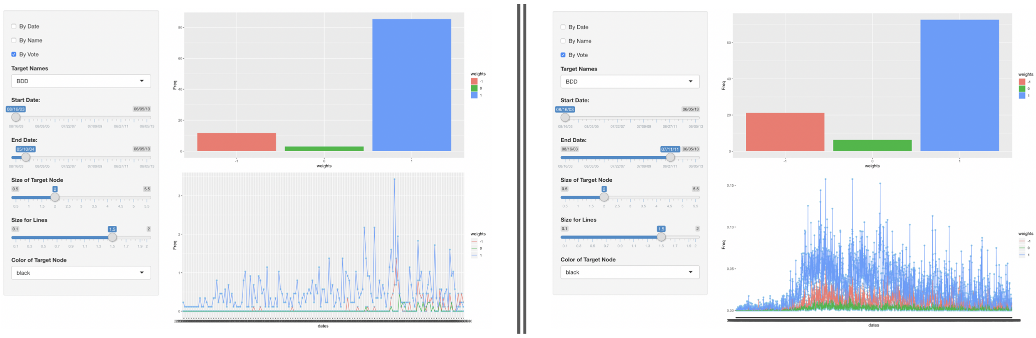

Fig. 4. **On the left hand side, this show the probability of a negative, neutral, and positive vote in the first few months. The right hand side shows the probability of a negative, neutral, and positive votes towards the end of the data set. In both left and right side, they show the daily probability over time.**

4.1 Visualizing the community

When the network was completely constructed there was a few things that was visible from the network: how people inserted into the network reacted and how the network changed over time. When looking at the network, one can see the different people inserted into the network. One can see the people that were inserted into the network and who was not. Many people were voted upon within the network, but not all were inserted into the network. People that were not inserted into the network, one can see that they had not voted on other people in the network. They had been voted on, but they had never voted on someone else over time. These people were not important to the network. They were inserted into the network but nothing else can from that. Looking at the people that were inserted into the network, they used the power that was given to them. They were invited into the network and used the power to vote other people trying to be inserted into the network, but something interesting came from looking at this trend overtime.

When looking at the network as it changes overtime, one can see that it doesn't matter when one enters or is voted upon. As the network start off, there are only two people that are voted into the network. As the network expands by a month, new people try to join the network and are voted on outside the main network. A month after that, the people from outside of the network are now inserted into the overall network. This shows that overtime, people being inserted into the network are not significate. One reason why the network expands overtime is because people that are inserted where voted on by a source already in the network. This means that the full network is created with people accepted to the network and not accepted to the network. This can be seen in the graph because people not inserted to the graph do not vote on others. People that had been accepted into the used that power granted to them to have new members join the network.

When looking at the graphs it is hard to see who had been voted in and not been voted into the network but using the vote check box. One can see the probability of the overall votes given. As seen in figure 4 and figure 5, the probability of a person getting a yes vote is not much different from later in time. There is not a drastic change of probability in any of the categories. They all stay relatively constant during the time span.

Positive Topics	Probability Score	From Greatest Probability to Least (First line is Probability and Second line is the word)					
Topic 1	0.305015	.053 edit	0.027 articl	0.024 work	0.023 excel	0.020 editor	0.019 user
Topic 2	0.012504	.081 answer	.66 question	.035 reason	.032 think	.029 admin	.029 tool
Topic 3	0.012500	0.182 font	0.085 color	0.071 style	0.066 span	0.052 face	0.051 user
Topic 4	0.012501	0.096 good	0.084 user	0.048 tool	0.043 trust	0.039 great	0.033 luck
Topic 5	0.012503	0.051 user	0.024 talk	0.021 delet	0.021 page	0.017 nomin	0.015 articl

Table 3. **Topic Modeling in supportive comments. (Top five topics with top five words have been displayed)**

4.2 Motivations behind being supportive

In order to find the motivations behind the supportive behavior, we first did the topic modeling using LDA. The results are being shown in table 3. Then using the top words we extracted the sentences from the comments and find the topics which helped us in finding the broader themes of motivations behind being supportive behavior.

4.2.1 Trust. One of the themes that generated from the supporting comments is trust. People voted as they trust the candidates. Their trust has been gained from several sources. First, every candidate had to submit their answers about previous experience in editing, their vision for the future and how they will help the Wikipedia community. Many voters supported them by reading those agendas. They put their trust seeing what they have already accomplished and what they are capable to do. Second, they put their trust because they know their work. The voters have seen those candidates work previously and they judged around that and put their trust. Lastly, they put their trust because they personally know that person. Even though, some voters didn't know what they have done or capable of doing, they instantly trust that person as they know him.

> *"My father told me you can always trust a WM grad. In all seriousness, I have run into BDD at AfD several times and have long thought he was a solid admin candidate" - random voter*

4.2.2 Organizational Commitment. Another broad theme generated from the multiple iterations of thematic analysis is the organizational Commitment. Organizational commitment can be referred as, in a general sense, the stakeholders'/users' psychological attachment to the organization and the organizational objectives. The users in our case the voters tend to overlook the personal relations or other thought process that might cloud them in a fair judgement. They judge by the fact around what will be the best for the organization like the below quote from the random voter -

> *"I nominated him because he voiced a desire to delete articles with bad sourcing. He needs admin tools to delete articles, and also deal with editors who insist on breaking rules on sources" - random voter*

Here a voter expressed his/her interest in voting for a candidate as he/she deemed that the candidate posits very interesting future for the organization - Wikipedia. These commitment builds

Negative Topics	Probability Score	From Greatest Probability to Least (First line is Probability) (Second line is the word)					
Topic 1	0.887468	0.013 edit	0.012 oppos	.010 experi	0.010 articl	0.009 candid	.008 qualifi
Topic 2	0.594967	0.034 time	0.026 oppos	0.024 candid	0.023 concern	0.021 strong	0.017 mistak
Topic 3	0.012504	0.012 oppos	0.009 admin	0.009 readi	0.008 notnow	0.007 edit	0.007 good
Topic 4	0.012502	0.028 front	0.026 face	0.013 trebuchet	0.013 user	0.009 oppos	0.009 head
Topic 5	0.012502	0.068 oppos	0.019 neurtal	0.015 concern	0.015 issu	0.011 candid	0.011 comment

Table 4. **Topic Modeling in non-supportive comments. (Top five topics with top five words have been displayed)**

upon duties and values, and the degree to which an stakeholder thinks about an organization out of a sense of obligation and duties.

4.2.3 Personal Relationship. The third theme that came out from the analysis is the personal relationships. Personal relationships puts great contribution on the voting decisions. By analyzing the text, it has been come up multiple times how candidates are getting votes as the voters know them. As the Wikipedia promotes transparent voting behavior, candidates are not being coded. Voters who know the candidates tend to vote for them. Like the below comment we can see from a voter

> *"He is not an admin already? Wow. From all my experiences with him at the IRC, Legoktm has been certainly very much suitable for this adminship. Also, Anonymouse explains very well why" - random voter*

Personal relationships works a little bit different than trust. Even though some of the personal relationships generate trust but whether voters trust them ot not some of the votes got influenced by their relationship with the candidates.

4.3 Motivations behind being non-supportive

In order to find the motivations behind the supportive behavior, we first did the topic modeling using LDA. The results are being shown in table 4. Then using the top words we extracted the sentences from the comments and find the topics which helped us in finding the broader themes of motivations behind being non-supportive behavior.

4.3.1 Doubt. From the non-supportive comments, first thing that came out as a prominent theme is they doubt the capabilities of the candidates. As explained before, every candidate has a very detail profile and based on that voters get to decide whether or not they are gonna vote for that person. Such non-supportive comments can be represented by below comments -

> *"Does not trust himself with the tools" - random voter*

These really simple comments put out the strong trends around mistrust or doubt about the candidates. As most of the Wikipedia comments are very descriptive and to the point, these abrupt comments nudge around the direction that voters don't know the candidate or any previous work. They just don't believe in their capabilities.

4.3.2 Experience. Most constructive comments lead to this theme. Voters who are deeply linked with the Wikipedia community knows what kinds of experience is required for the betterment of the community and what is not required. Voters who value the experience put out really thoughtful comments.

> *"I had a brief look at your edit history and saw CSD-A7 and A1 tags placed within a minute of article creation, as well as vandalism warnings given for run of the mill mistakes like a badly formatted image insertion into an infobox" - random voter*

The valuation of experience is probably the very positive way of how voters think about the candidates. We have seen several comments which indicates voters know the candidates but they are not satisfied with the performance expectancy and they didn't vote. Like the comments below -

> "*Well, it actually looked quite promising for a moment and I personally know you. as you have no blocks and over 2000 article edits, but I have to agree with the other editors and say that your answer to Q1 shows that you have no experience in admin areas" - random voter*

From our analysis it's prominent that voter's behavior doesn't get influenced by other voters. It's get motivated from trust and experience of the candidates.

5 DISCUSSION

In our discussion, we have connected our results with broader implications from "Unified theory of acceptance and use of technology (UTAT)" [38] and our results suggest the feasibilty around practical implications of combining topic modeling and thematic analysis.

5.1 Voting Behavior Patterns

During crafting our research problem, we tried to connect voting behavior with "Social Influence Theory". Through our investigation it has been revealed part from the "Organizational Commitment" [38]. Organizational commitment is the bond users/stakeholders experience with their organisation. Broadly speaking, users/stakeholders who are committed to their organisation generally feel a connection with their organization, feel that they fit in and, feel they understand the goals of the organisation. The added value of such employees is that they tend to be more determined in their work, show relatively high productivity and are more proactive in offering their support [38]. But It can broadly connected to the "Unified theory of acceptance and use of technology (UTAT)" [36].

The UTAUT suggests that four core constructs (performance expectancy, effort expectancy, social influence and facilitating conditions) are direct determinants of behavioural intention and ultimately behaviour, and that these constructs are in turn moderated by gender, age, experience, and voluntariness of use [36]. It is argued that by examining the presence of each of these constructs in a "real world" environment, researchers and practitioners will be able to asses an individual's intention to use a specific system, thus allowing for the identification of the key influences on acceptance in any given context.

5.2 Practical implications of combining topic modeling and thematic analysis

We address the challenge of analyzing the thematic composition of large document collections using quantitative topic models, which usually implies an erosion of contextual meaning. To address this challenge, we did a structured mixed-methods research (MMR) approach. This design draws on qualitative coding and quantitative hierarchical clustering to validate and enhance topic models through re-contextualization. the proposed MMR approach mitigates shortcomings of both quantitative research, such as the problematic choice of the "correct" amount of topics included in a model, and de-contextualization of textual data, as well as qualitative research, such as deficiencies regarding validity, reliability, and replicability. This particular study paved ways of future work may consider validating and refining the proposed MMR approach, as well as creating software environments that support researchers in implementing such an approach.

6 CONCLUSION

We conclude our study stating that as it was a learning curve for all us, we tried our best to imply the visualization and topic modeling in extracting the motivations behind voters' decisions in Wikipedia adminship. Our study revealed an interesting fact how voters are getting influenced by self personality trait rather than getting influenced by each other.

REFERENCES

[1] 2017. Wikipedia elections network dataset – KONECT. (April 2017). http://konect.uni-koblenz.de/networks/elec

[2] Charu C Aggarwal, Jiawei Han, Jianyong Wang, and Philip S Yu. 2003. A framework for clustering evolving data streams. In *Proceedings of the 29th international conference on Very large data bases-Volume 29*. VLDB Endowment, 81–92.

[3] Ronald L Akers, Marvin D Krohn, Lonn Lanza-Kaduce, and Marcia Radosevich. 1995. Social learning and deviant behavior: A specific test of a general theory. In *Contemporary Masters in Criminology*. Springer, 187–214.

[4] Solomon E Asch. 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied* 70, 9 (1956), 1.

[5] David Auber, Yves Chiricota, Fabien Jourdan, and Guy Melançon. 2003. Multiscale visualization of small world networks. In *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No. 03TH8714)*. IEEE, 75–81.

[6] John A Barnes. 1969. Graph theory and social networks: A technical comment on connectedness and connectivity. *Sociology* 3, 2 (1969), 215–232.

[7] Michael Baur, Marc Benkert, Ulrik Brandes, Sabine Cornelsen, Marco Gaertler, Boris Köpf, Jürgen Lerner, and Dorothea Wagner. 2001. Visone Software for visual social network analysis. In *International Symposium on Graph Drawing*. Springer, 463–464.

[8] Pavel Berkhin. 2002. Survey of clustering data mining techniques. Accrue Software. *Inc. TR, San Jose, USA* (2002).

[9] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy* 100, 5 (1992), 992–1026.

[10] Mustafa Bilgic, Louis Licamele, Lise Getoor, and Ben Shneiderman. 2006. D-dupe: An interactive tool for entity resolution in social networks. In *2006 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 43–50.

[11] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[12] Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature* 489, 7415 (2012), 295.

[13] Hermann Brandstätter. 1982. *Group decision making*. Vol. 25. Academic Press.

[14] Tse-Hsun Chen, Stephen W Thomas, and Ahmed E Hassan. 2016. A survey on the use of topic models when mining software repositories. *Empirical Software Engineering* 21, 5 (2016), 1843–1919.

[15] Ali Daud, Juanzi Li, Lizhu Zhou, and Faqir Muhammad. 2010. Knowledge discovery through directed probabilistic topic models: a survey. *Frontiers of computer science in China* 4, 2 (2010), 280–301.

[16] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American society for information science* 41, 6 (1990), 391–407.

[17] Li Fei-Fei and Pietro Perona. 2005. A bayesian hierarchical model for learning natural scene categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2. IEEE, 524–531.

[18] Leon Festinger, Stanley Schachter, and Kurt Back. 1950. Social pressures in informal groups; a study of human factors in housing. (1950).

[19] LC Freeman. 2004. The Development of Social Network Analysis: a Study in the Sociology of Science.,(Empirical Press: Vancouver, BC, Canada). (2004).

[20] Frédéric Gilbert, Paolo Simonetto, Faraz Zaidi, Fabien Jourdan, and Romain Bourqui. 2011. Communities and hierarchical structures in dynamic social networks: analysis and visualization. *Social Network Analysis and Mining* 1, 2 (2011),

83–95.
[21] Michelle Girvan and Mark EJ Newman. 2002. Community structure in social and biological networks. *Proceedings of the national academy of sciences* 99, 12 (2002), 7821–7826.
[22] Jeffrey Heer and Danah Boyd. 2005. Vizster: Visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 32–39.
[23] Thomas Hofmann. 2001. Unsupervised learning by probabilistic latent semantic analysis. *Machine learning* 42, 1-2 (2001), 177–196.
[24] Michael A Hogg. 2018. *Social identity theory.* Stanford University Press.
[25] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. 1999. Data clustering: a review. *ACM computing surveys (CSUR)* 31, 3 (1999), 264–323.
[26] Shuhui Jiang, Xueming Qian, Jialie Shen, Yun Fu, and Tao Mei. 2015. Author topic model-based collaborative filtering for personalized POI recommendations. *IEEE transactions on multimedia* 17, 6 (2015), 907–918.
[27] Panos Kalnis, Nikos Mamoulis, and Spiridon Bakiras. 2005. On discovering moving clusters in spatio-temporal data. In *International Symposium on Spatial and Temporal Databases*. Springer, 364–381.
[28] ELIHU KATZ and FELD LAZARs. 1955. PF Personal influence. (1955).
[29] Herbert C Kelman. 1958. Compliance, identification, and internalization three processes of attitude change. *Journal of conflict resolution* 2, 1 (1958), 51–60.
[30] M Rogers Everett. 1995. Diffusion of innovations. *New York* 12 (1995).
[31] Adam J Schwarz, Alessandro Gozzi, and Angelo Bifone. 2009. Community structure in networks of functional connectivity: resolving functional organization in the rat brain with pharmacological MRI. *Neuroimage* 47, 1 (2009), 302–311.
[32] Zeqian Shen, Kwan-Liu Ma, and Tina Eliassi-Rad. 2006. Visual analysis of large heterogeneous social networks by semantic and structural abstraction. *IEEE transactions on visualization and computer graphics* 12, 6 (2006), 1427–1439.
[33] Muzafer Sherif and Carl I Hovland. 1961. Social judgment: Assimilation and contrast effects in communication and attitude change. (1961).
[34] Károly Takács, Andreas Flache, and Michael Mäs. 2016. Discrepancy and disliking do not induce negative opinion shifts. *PloS one* 11, 6 (2016), e0157948.
[35] Robert C Tryon. 1939. Cluster analysis. Edwards Brothers. *Ann Arbor, Michigan* (1939), 122.
[36] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly* (2003), 425–478.
[37] Wendy Wood. 2000. Attitude change: Persuasion and social influence. *Annual review of psychology* 51, 1 (2000), 539–570.
[38] Darwish Abdulrahman Yousef. 2017. Organizational commitment, job satisfaction and attitudes toward organizational change: A study in the local government. *International Journal of Public Administration* 40, 1 (2017), 77–88.

Received February 0000; revised March 0000; accepted June 0000