Networks effects of firehosing during the 2016 United States presidential elections

Hugo Souza-Almeida[1], **Humberto Torres Marques-Neto**[1]

[1]Departamento de Ciência da Computação - Instituto de Ciências Exatas – Pontifícia Universidade Católi

deSouza@protonmail.com,humberto@pucminas.br

Abstract. *The United States 2016 presidential election was quite controversial in many ways. Various different events, such as an alleged russian intervention in the electoral process, hacking of the Democratic National Congress private servers, publishing of said material on wikileaks, a wide usage of bots on social media and dubious content being propagated ad nauseum. This paper seeks to analyze the network effects of the 'firehose of lies' tactic used by candidates, that consists of increasing the tempo of information being propagated without necessarily having a commitment to truth, utilizing state of the art topic analysis techniques to identify the primary focus areas of political candidates and the predominant topics on social media. Our findings indicate that while candidates can shape social media discourse, the collective mindset of individuals on these platforms is independent, suggesting that candidates cannot solely dictate the topics of discussion.*

Resumo. *A eleição presidencial de 2016 nos Estados Unidos foi bem controversa em diversas formas. Vários acontecimentos diferentes, como uma alegada intervenção russa no processo eleitoral, invasão dos servidores privados do congresso nacional democrático, publicação de vazamentos no wikileaks, uso extensivo de robôs em mídia social e conteúdo duvidoso sendo propagado extensivamente. Este artigo procura analizar os efeitos em rede da tática 'firehose de mentiras' utilizada por candidatos, que consiste em aumentar a velocidade de propagação de informação sem um comprometimento com a verdade, utilizando técnicas de análise de tópicos de última geração para identificar as áreas de foco principal dos candidatos políticos e os tópicos predominantes nas redes sociais. Nossas descobertas indicam que, embora os candidatos possam moldar o discurso nas redes sociais, a mentalidade coletiva dos indivíduos nessas plataformas é independente, sugerindo que os candidatos não podem ditar sozinhos os tópicos de discussão.*

1. Introduction

Candidates for political office often face a barrage of hostility on social media platforms. This abuse disproportionately targets individuals who identify as minorities, exacerbated by entrenched systemic biases. They also have to endure stereotyping [Nawabdin 2021], and are negatively portrayed by the media, even more when they address systemic biases

[Sorrentino 2021]

The 2016 U.S. election was notably contentious, marked by malicious interference from international actors on social media, driven by agendas divergent from the interests of the American public and liberal democracies at large. The intense contestation of the election extended beyond the political arena, as the online polarization of discourse impacted personal relationships, influencing not only the electoral landscape but also the broader social context.

The 'firehose of falsehood' [Paul and Matthews 2016] strategy was highly deployed and involves rapidly disseminating information at a pace that outstrips the ability to verify its accuracy. It is a direct descendant from soviet active measures and is a threat to democracy everywhere. This approach leverages the fact that producing incorrect or deceptive content is quicker than conducting thorough fact-checking.
Such misinformation tends to have a more significant impact on candidates outside the societally dominant demographic groups, and the fact that elections in liberal democracies are usually won by small margins, a small advantage such as propagating identity-based falsehoods could become standard practice and affect minorities disproportionately.
Candidates used social media in provocative ways, without consideration for consequences [Savoy 2018], and analyzing how they can effectively control the narrative on social media, thereby influencing public perception, could be instrumental in fostering a more just and equitable society, as well as a less toxic social environment for all.

This study aims to leverage practical data via topic analysis utilizing state of the art technology to examine the cascading impact of inundating social networks with copious amounts of information, and in this way attempt to quantify the effectivity of the technique known as 'firehose of falsehoods' [Paul and Matthews 2016] in shaping political discourse in social media and influencing elections.

The larger goal of this study is to help understand prejudice based discourse and social media polarization, and foster a safer, more equitable society. This study analyzes a Twitter post database from the 2016 U.S. Presidential Election, focusing on quantitative metrics like tweet and word counts. It employs BERTopic, an advanced topic modeling algorithm, to identify discussion topics and examine if candidates influenced debate discussions. The analysis includes keyword extraction using KeyBERT and employs tools like Pandas and Plotly for data handling.

This study observed that in the 2016 U.S. Presidential Elections Twitter discussions mirrored and diverged from the debate topics. The first debate saw a 50% increase in mentions of "Obama," following candidate Trump's emphasis on the term, and a similar rise in discussions about Trump and Clinton. Mainstream and key issues like women's rights were also prominent, despite not being central to the debate. In the second debate, while Trump tried to refocus on Obama and respond to social media trends like women's rights, there was a notable rise in conversations about transgender and LGBT rights, which were not primary debate topics. The third debate led to an immediate increase in Twitter activity, especially on WikiLeaks and women's rights, the latter being consistently addressed by Clinton but not by Trump. This pattern indicates a complex interplay between debate content and online discourse, where strategic emphasis by candidates and social media trends both shaped the conversation.

The structure of the article includes several sections: one detailing related works, another describing the methodology, followed by a section presenting the actual results and their analysis, a section dedicated to the conclusion, and finally, a section discussing potential future works.

2. Related Works

In their research, Caetano, Marques-Neto et al, focused on the analysis of Twitter user data during the 2016 U.S. Presidential election. Their study, detailed in Caetano, Marques-Neto et al, examined political homophily [Caetano et al. 2018b] - the tendency of individuals with similar political views to associate with each other on social networks. This concept is crucial in social network analysis as it can reveal the existence of ”echo chambers” or ideological bubbles. They investigated various types of Twitter interactions, including follows, mentions, retweets, both in unidirectional and reciprocal forms, as well as multiplex – multiple types of relationships or interactions that can exist between the same set of nodes in a network – connections and friendships among users with similar political views. Their findings indicated significant levels of homophily among negative sentiment users, supporters of Donald Trump, and supporters of Hillary Clinton across all scenarios. Notably, they observed an increase in homophily levels in cases of reciprocal interactions, similar speeches, or multiplex connections. Additionally, an adapted version of this study was published in a Brazilian journal [Caetano et al. 2017] in Portuguese in 2017. The ’echo chambers’ observed in this study directly affect the topics discussed on twitter.

Caetano, J.A. and Marques-Neto explored the behavior of Twitter users during a political campaign [Caetano et al. 2018a], as detailed in their 2018 publication. They focused on analyzing the language patterns in tweets, identifying which users gained more popularity during the campaign, and examining how the candidates’ tweets potentially influenced the mood variations of Twitter users, as reflected in the content of their shared messages. Generously, they provided the data they collected for the research, namely the database in this and the previous study, contributing significantly to the findings presented in this paper.

Sorrentino and colleagues [Sorrentino 2021] delved into how the U.S. media portrayed Hillary Clinton’s approach to gender issues during the 2016 presidential campaign. The study, employing a Critical Discursive Psychology approach, analyzes media content from February to November 2016. The study highlights the complex challenges female politicians face in addressing gender and inequality. The media often criticized Clinton for either playing the victim or not adequately representing women’s issues. This portrayal reflects a postfeminist worldview where gender inequalities are seen as outdated issues. The study notes that the media’s framing of Clinton’s gender orientation had implications for understanding gender, gender inequality, and power relations. The criticism and negative portrayal of female leaders who publicly address gender issues underscore the delicate balance women in power must maintain when discussing gender and sexism. This reflects a broader societal challenge where acknowledging gender inequalities and advocating for change remains a contentious and risky endeavor for women in leadership positions. The delicate situation women and other minorities are situated can and will be exploited by candidates utilizing the 'firehose of falsehood' technique in order to set the tune of the discussions and control the narrative.

Oliveira, Kasai and Marques-Neto conducted a study on the impact of various and repeated influence tactics [Oliveira et al. 2022], as outlined in their 2022 publication. Their research focused on understanding how Twitter users are affected by information susceptibility and the thresholds at which they adopt new information. The 'firehose of falsehood' technique is one influence technique, and this study also tries to understand how Twitter users are affected by underhanded tactics.

Oliveira, Kasai and Marques-Neto examined the extent to which Twitter users are susceptible to information and the point at which they are likely to adopt new information [Oliveira et al. 2020]. The 'firehose of falsehood', by flooding users with new information, tries to convince susceptible users of inaccurate information.

Pew Research Center did a post election study [Pew Research Center 2018] on registered voters backgrounds and their voting record in the 2016 United States Presidential Election. In March 2018, a modest gender gap emerged among Trump supporters: men gave Trump higher ratings than women. Older Trump voters, particularly from the Silent Generation, rated him the highest. Trump voters without a four-year college degree consistently rated him higher than those with a college degree or more. The article underscores the complexity of Trump's support base, revealing both steadfast loyalty and changing perceptions among different voter segments. The article focuses on the enduring "warm" feelings Donald Trump's voters held for him, even more than a year into his presidency. Confirmation bias might play a significant role in how much users are susceptible to erroneous information, as used in the 'firehose of falsehood' techhnique.

Kennedy Et Al [Kennedy et al. 2020] critically examined the performance of pre-election polls during the 2016 U.S. Presidential Election, focusing on whether the polls failed and why. Its conclusion was that although national polls were accurate in predicting the popular vote, state-level polls, particularly in the Upper Midwest, underestimated support for Trump. Several factors contributed to this, including the over representation of college graduates in the polls and late-deciding voters. The 'Shy Trump' theory was found to have little evidence, and the study suggests that the observed polling errors in 2016 do not indicate a systematic bias in U.S. polling. The 2016 United States Presidential election saw activists in social media questioning the integrity of the pre-election polls, which was one major aspect of the 'fake news' narrative employed by Trump.

The article by Fatemah Nawabdin [Nawabdin 2021] examines the impact of gender stereotypes on Hillary Clinton's electability in the 2016 U.S. Presidential Election. Using data from the 2016 American National Election Survey and statistical models, the study explored how voters' perceptions of Clinton's masculine and feminine traits influenced their voting behavior and feelings towards her. The study found that masculine personality traits were more influential in winning the presidency than feminine traits. Voters' positive perceptions of Trump's masculine traits significantly increased their likelihood of voting for him and having warmer feelings towards him. The findings suggest that while women can achieve high political offices, as exemplified by Kamala Harris in the 2020 election, gender stereotypes remain an influential factor in shaping electoral outcomes. Falsehoods targetting gender stereotypes can, this way, be very effective for undermining an opponent's political viability.

Savoy analyzed [Savoy 2018] the style and rhetoric of Hillary Clinton and Donald Trump during the 2016 US presidential election. He examined their oral and written communication forms, providing insights into their use of pronouns and other linguistic features. The study indicated that Clinton's rhetoric employed more cognitive words while negative emotions and exclusive terms occurred more frequently in Trump's verbiage. His oral communication differed significantly from his written speeches, suggesting the involvement of ghostwriters in the latter, while Clinton's was more consistent. The article also notes that Trump's campaign used social media in a provocative way without any real consideration for consequences. Utilizing social media without caring for the ethical consequences involved might make a campaign more susceptible to using underhanded tactics, such as the 'firehose of falsehood', and this study indicates that the Trump campaign was very intentional about its communication style.

Paul and matthews [Paul and Matthews 2016] discuss the Russian "Firehose of Falsehood" propaganda model, which is characterized by high-volume, multichannel messaging that is often contradictory and lacks a clear source. It discusses how these strategies are rooted in Soviet-era techniques but have adapted to modern technology and the current information environment, utilizing the Internet, social media, and a variety of media outlets. The article emphasizes the challenge in countering such propaganda due to its nature and the psychological aspects behind its effectiveness, and offers potential options for countering it. These options include increasing media literacy, promoting alternative narratives, and imposing costs on Russia for its propaganda activities. The article emphasizes the need for a comprehensive, coordinated response to this propaganda model. Part of the coordinated response involves understanding when and how candidates use this tactic in an electoral context, and this study attempts to understand the electoral context of the 2016 United States presidential elections and its usage of the tactic.

This paper seeks to identify the 'firehose of lies' [Paul and Matthews 2016] technique in an electoral context, in order to prevent gender or other identity based inequalities [Sorrentino 2021] such as gender stereotypes [Nawabdin 2021] and ascertain how influence tactics [Oliveira et al. 2022] swayed [Caetano et al. 2018a] susceptible users [Oliveira et al. 2020] in the electoral outcome [Pew Research Center 2018] of the 2016 United States Presidential Elections by analyzing dominant topics both in the debates and on social media, outside homophilic user groups [Caetano et al. 2018b] and thus in the general population.

3. Methodology

Utilizing the official Twitter API, Caetano et al. [Caetano et al. 2017] embarked on an extensive data collection project, amassing a dataset that included tweets, user profiles, and contact networks. The period of data acquisition spanned from January 1st to November 30th, 2016, a timeframe strategically chosen to encompass critical political events in the United States, including the three televised debates between presidential candidates Donald Trump and Hillary Clinton (occurring on September 26th, October 9th, and October 19th) and the election day on November 8th.

The Twitter API facilitated the retrieval of up to 200 of the most recent tweets published by each user, albeit with a limitation of 300 requests per 15-minute interval. The initial step in data collection involved identifying 'seed users,' defined as individuals who actively engaged with content related to the U.S. presidential campaign. This identification was achieved through the API's streaming method, enabling real-time tweet collection. A user was classified as a seed user if they retweeted content from either of the candidate's accounts at least once between August 1st and November 30th. The underlying hypothesis was that retweeting a candidate's tweet indicated not only the user's engagement with the content but also their participation in or promotion of political discourse on the platform.

All users meeting this criterion during the collection period were designated as seed users. For each of these users, the researchers gathered their Twitter profile information, tweet timeline (from January 1st to November 30th), and contact network (including followers and friends). The data collection extended to users within each seed user's contact network, up to two hops from the origin. The focus was on users with English set as the default language in their profiles and who had posted at least 200 tweets. The final dataset comprised data from a total of 115,664 users, including 37,468 seed users and 78,196 users within their networks.[Caetano et al. 2018b].

We then performed a quantitative analysis, in order to ascertain the characteristics of the database, such as character count, characters per tweet, word count, number of tweets and number of users. Then we performed text analysis, and utilized word embeddings through BERTopic [Grootendorst 2022].

BERTopic is an algorithm designed for topic modeling, a natural language processing (NLP) task aimed at identifying and clustering similar topics within a collection of documents. It combines the capabilities of BERT (Bidirectional Encoder Representations from Transformers) embeddings with classical topic modeling algorithms to enhance the performance and accuracy of topic detection.

BERT a powerful and versatile model for transfer learning developed by Google, to convert text data into numerical representations (embeddings). These embeddings capture the contextual relationships between words in a document.
BERTopic leverages transformers and c-TF-IDF to create dense clusters allowing for easily interpretable topics whilst keeping important words in the topic descriptions. BERTopic was utilized to determine whether any candidates influenced the direction of discussions through firehosing in the periods of three days before, on the day of, and three days after the most watched debates of the 2016 elections, the exact days being shown in Table 1.

This analysis involved correlating these influences with repeated keywords in each debate, as captured on various TV stations' websites and analyzed using KeyBERT [Grootendorst 2020], a Python library used for extracting keywords and keyphrases from text also built on BERT.

Table 1. Elections Analyzed

Debate	Day 1	Debate day	Day 7
First debate	23/09/2016	26/09/2016	29/09/2016
Second debate	06/10/2016	09/10/2016	12/10/2016
Third debate	17/10/2016	20/10/2016	23/10/2016

Additionally, open-source libraries like Pandas and Plotly, alongside the Python programming language, were be employed to conduct this analysis.
The source code is available on github [Souza-Almeida 2023].

Table 2. Electoral Database

Database	Tweets	Users	Seed Users	Size (GB)
2016	65,329,481	4,334,627	115,664	122,5

Table 2 describes the 2016 electoral database we analyzed, in terms of tweet number, number of users, seed users and size, and demonstrates the size and how comprehensive it is in metrics. It also describes the number of users and the users used to collect the data.

Table 3. Messsage Size

Dataset	Tweets	Character Count	Characters/tweet	Word count	Words/tweet
2016	65,330,066	6,965,181,300	106.62	1,009,198,753	15.45

Table 3 describes the 2016 electoral database we analyzed, in terms of character count, characters per tweet, word count, and words per tweet. It describes through metrics the verbosity of users in the collected tweets.

Table 4 describes the 2016 electoral database we analyzed, in terms of mentions of each candidate, and gauges the popularity of each candidate as similar in scope.

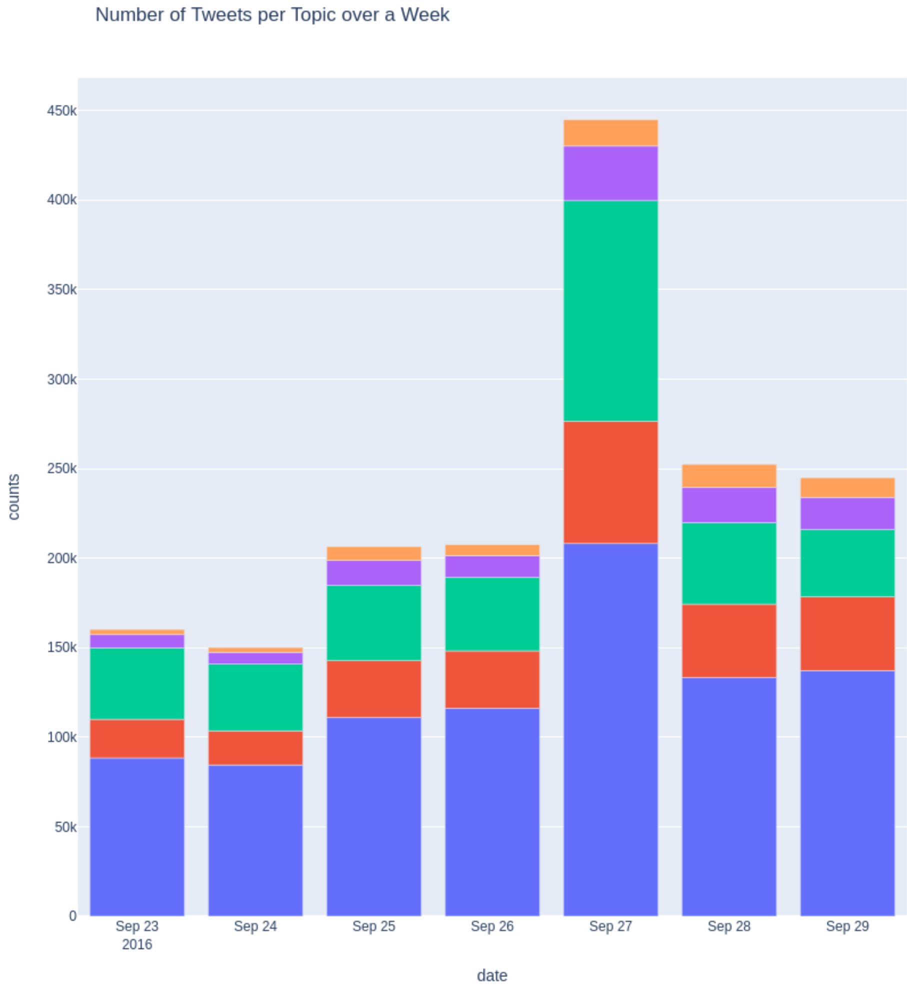
4. Results

We utilized the topic analysis tool BERTopic to calculate the predominant topics in the 2016 elections in each week of the first three debates in the 2016 United States elections in order to ascertain if any of the candidates managed to set the tune of the discussions through innacurate information.

Week 1: First Clinton-Trump debate

Table 4. Mentions

Year	Mentions Trump	Mentions Hillary
2016	2,445,500	2,477,708



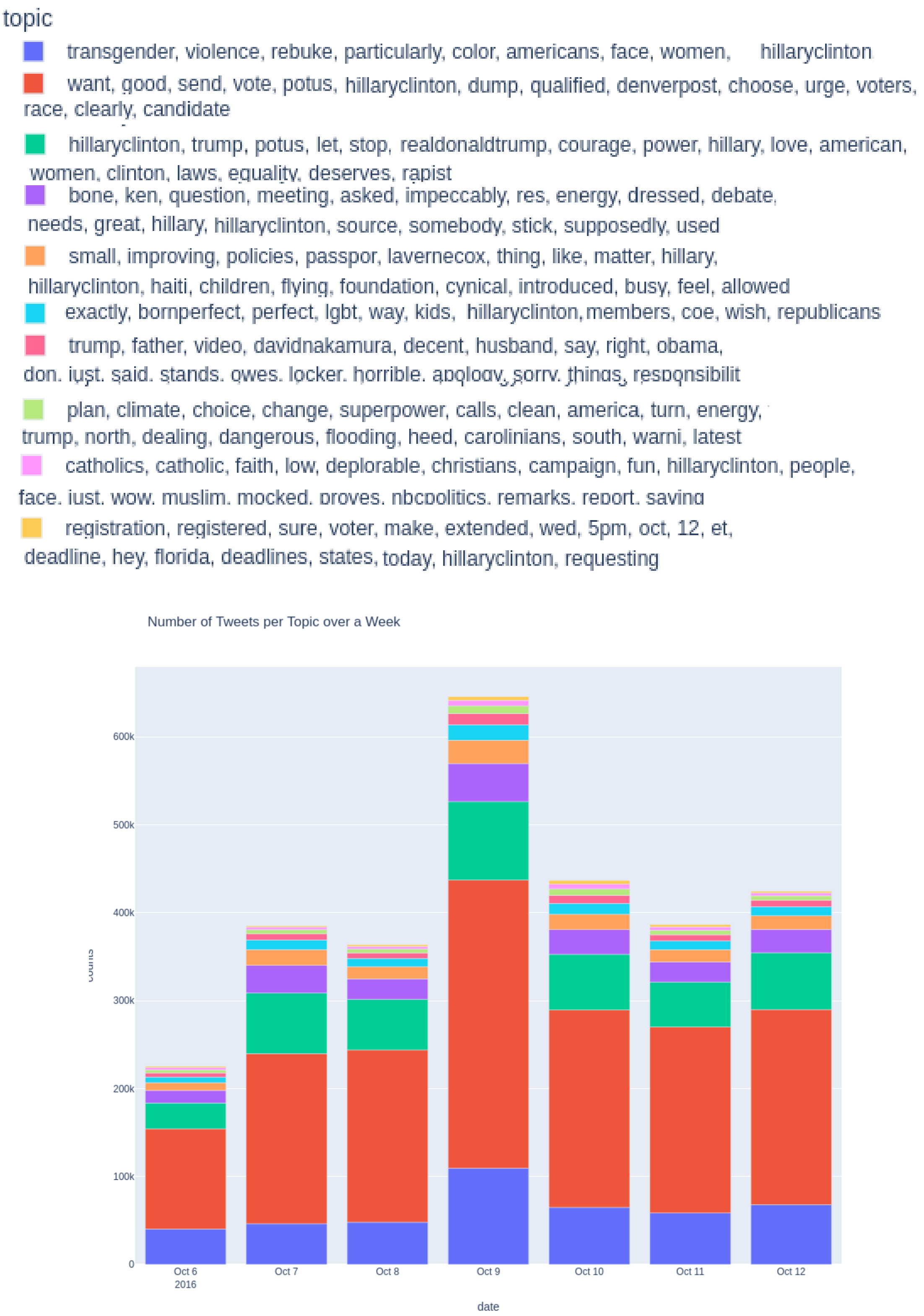
During the week of the first debate, from September 26th, the debate day, to September 27th, there was a significant 50% surge in discussions mentioning Obama as evidenced by the green bar in the graph, a term that President Trump frequently used during the debate. Similarly, topics involving Trump and Clinton saw a 50% increase as seen in the green, blue, red and purple bars in the graph, with those mentioning Trump rising more prominently, aligning with expectations. Alongside these political discussions, mainstream topics also surfaced, covering pop culture events of the week like Lana Parrilla, PlayStation, and CelebrityDimSum, as evidenced in the green and red bars in the graph. Additionally, key issues such as school shootings, women's rights, and the Cuban embargo were discussed, even though they were not frequently or directly addressed in the debate. The media [Bradner 2016a] at the time noted Trump interrupted Clinton during the debate many times, which could be construed as gender based aggression, and the topic of Trump's alleged racism was en vogue as well according to CNN, complete with Trump's record of claiming Obama wasn't born in the United states.

Table 5. Topic Counting with keybert for the first debate

Keyword	Count Trump	Count Hillary
obama	12	3
clinton + hillary	11 + 7	na
donald	na	7
president	9	6
isis	4	3
china	4	na
iraq	4	2
nuclear	4	3
tax	3	9
police	3	5
economy	1	7

During the first debate, candidate Trump repeated the keyword Obama more than any other term as shown in Table 5. Both debaters frequently brought up several political subjects like 'Iraq', 'nuclear', 'tax', and 'police'. After the debate, discussions about Obama saw a surge on Twitter. This, together with the persistent focus on this topic during the debate, might be seen as an effort to inundate the conversation with an irrelevant subject, as described in the firehose technique, possibly involving untruths, such as incorrect claims about Obama's place of birth, which was a factor in Trump's political rise and remained a factor of his popularity with his fans after he shifted positioning.

Week 2: Second Clinton-Trump debate



In the week encompassing the second debate, spanning October 8th, the date of the debate, through October 9th, there was a notable uptick in discussions related to transgender rights, as seen in the blue bar in the graph, climate change, as seen in the yellow bar, and LGBT rights, as seen in the light blue and blue bars. These subjects, typically central to political discourse in the United States, were not explicitly highlighted in the debate's primary topics. President Trump kept his focus on former President Obama, as previously observed in the first debate, achieving limited success in this endeavor as there were topics related to Obama but they were in the minority. Additionally, he aimed to engage with topics that had garnered significant attention on social media following the previous debate, like women's rights, by strategically reiterating certain keywords during the debate. The media at the time [Bradner 2016b] notes that trump was on the defensive regarding accusations of gender oriented abuse and directed personal attacks towards Clinton claiming he intended to arrest her, as well as using of falsehoods such as claiming he was against the then unpopular war on iraq when it was declared, whereas at the time he supported it. Trump's falsehoods and attacks saw no resonance on social media directly.

Table 6. Topic Counting with keybert for the second debate

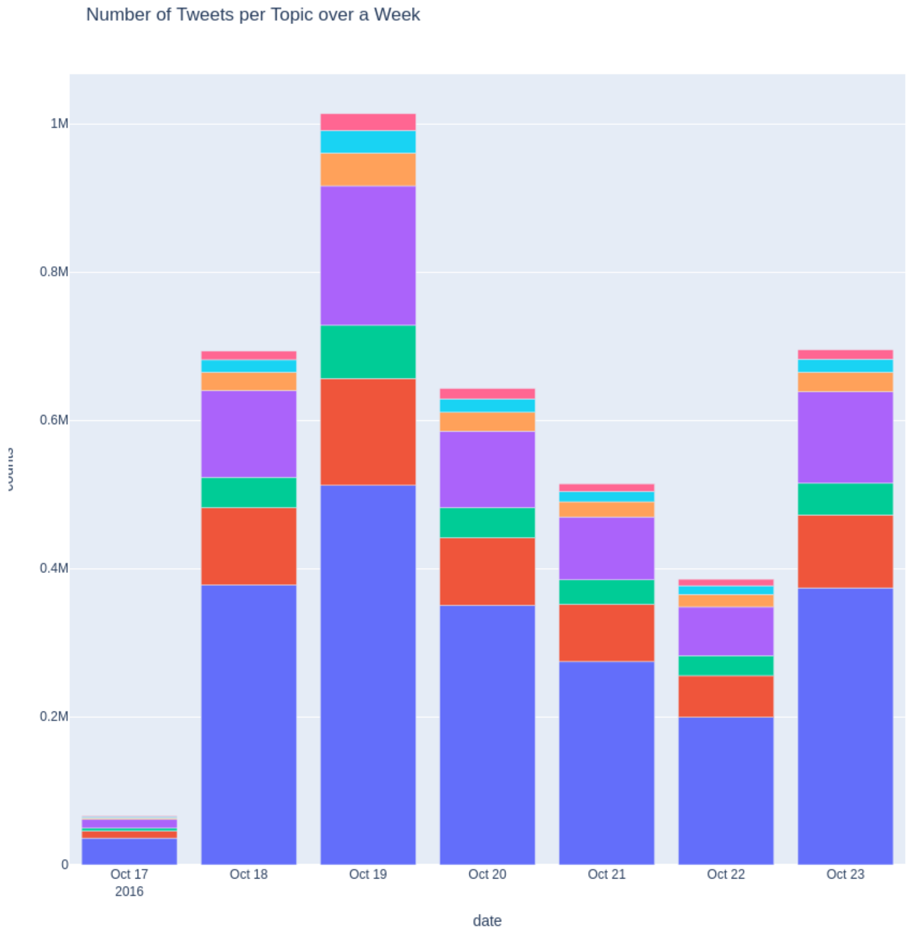
Keyword	Count Trump	Count Hillary
obama	8	3
clinton + hillary	3 + 14	3
donald + trump	na	9 + 4
president	4	9
isis	7	2
obamacare	4	1
iraq	3	2
women	2	na
taxes	4	1
disaster	4	na
russia	2	na

During the second debate trump continued attempting to shift the focus of the discussions towards former president Obama, as shown in Table 6, but there weren't correlated discussions on twitter. Obama was still one of the most repeated topics on the Trump Campaign, as well as 'isis' and there being an increase in the topic 'women', as an attempt to catch up to social media discussions involving women's rights.

Week 3: Third Clinton-Trump debate

topic

- trump, realdonaldtrump, clinton, hillaryclinton, hillary, like, president, donald, vote, women, agree, twitter, election, campaign, media, kellyannepolls, real
- thewalkingdead, hours, romanatwood, pop, thank, new, ebay, funko, quixoticthings, happy, like, taste, thetalkingdead, birthday, vlog, app, know
- know, proudwestindian, life, unfollowed, parents, future, told, don, yes, shannoncoulter, great, ppact, person, ooo, liveaction, liveactionnews, automatically, checked
- things, wikileaks, mushtaq30875787, people, love, don, like, station, rei, let, check, पर, near, death, savannahsnider _, school, new
- firefan, nfl, link, free, seavsaz, sign , nevspit, game, fan, drake, fuck, twd, cubs, good, twd7, seahawks, nflnetwork, titans
- mpadhyapak, ho, shivraj, ji, mp, melysapadilla, bhi, mere, se, thanx, yu, aur, mera, saare, que, jaaye, na, hable, duniya, lot
- brutal, twd, hype, ptsd, raccoons, seeing, realise, sonka_vanda, samurai, hacks, suspenseful, thrones, traumatizing, watching, nate_cohn, gorgeous, music



In the week of the third debate, specifically from October 18th to the 19th, the latter being the debate day, there was a noticeable increase in activity. Unlike the first two debates, where the surge in engagement occurred a day later, this spike was immediate or preceding the final debate. Key issues like WikiLeaks and women's rights gained traction, as seen in the blue and purple bars respectively. During the debate, candidate Hillary Clinton addressed the topic of women's rights, with consistent focus. In contrast, Trump did not mention these issues. The media at the time [Caldwell 2016] noted trump was on the defensive regarding accusations of collusion with russia, as well as of abuse by different women. This database saw no topics mentioning russia, but did see traction in topics mentioning women's rights, including the dominant one. The clinton campaign managed to dominate the discourse this week with allegations of Trump's abuse, and was very influential as noted on the previous week's topics. Whereas the accusations could be true, the tactic of utilizing them in an electoral context before they could be proved or disproved constitutes of the firehosing tactic as well.

Table 7. Topic Counting with keybert for the third debate

Keyword	Count Trump	Count Hillary
obama	8	5
hillary	10	
donald + trump	na	7 + 8
president	6	15
isis	5	na
wikileaks	0	1
iraq	2	3
women + woman	na	4 + 2
taxes + tax	2 + 0	3 + 3
russia + russians	4 + 0	0 + 1
putin	8	2

The topics of the third debate, as shown in Table 7, didn't mention topics that were highly contested on twitter during that week, like 'wikileaks', but did show an attempt by candidate Clinton to push for the discussion regarding women's rights, as she increased the usage of this topic during the debate and Trump didn't mention it.

In our analysis of the three debates, we noted that both parties effectively employed the 'firehose of lies' technique to varying degrees of success. For example, Trump strategically emphasized criticisms of former President Obama in the debates and on social media, as reflected in trending topics on both platforms. On the other hand, Clinton concentrated on Trump's gender-based aggressions during the debates and on social media. Clinton's topics appeared to dominate the discourse more successfully, possibly because her campaign's use of the firehosing tactic was better aligned with the political zeitgeist at that time.

5. Conclusion

In conclusion, our comprehensive analysis employing BERTopic has successfully delineated the utilization of the 'firehose of falsehoods' technique by both factions within the political arena. This was achieved through the automated extraction of pertinent topics from an extensive corpus of Twitter posts, harnessing sophisticated natural language processing methodologies. In this study, the researchers confronted the challenge of analyzing a notably extensive dataset, exceeding 100 gigabytes in size. The sheer volume of this data necessitated prolonged processing times, with several weeks dedicated to each phase of the analysis. To effectively manage and interpret this data, the research team employed BERTopic, an advanced tool in Natural Language Processing (NLP). BERTopic distinguishes itself by leveraging BERT embeddings for refined topic modeling, surpassing traditional methods through the use of a robust language model. This approach enabled a nuanced and contextually rich extraction of topics from the textual data, with a standout feature being its Contextual Topic Identification. This capability allowed for the discernment of context in word usage, leading to the identification of topics that were both meaningful and representative of the actual content.

BERTopic further demonstrated its versatility by dynamically adjusting the number of topics post-training, thus fine-tuning the granularity of the extracted topics for more precise analysis. The tool also facilitated the creation of hierarchical topic structures, allowing for a layered understanding of the text data through the categorization of broader topics into specific subcategories.

In addition to its analytical capabilities, BERTopic offered practical user-friendly visualization tools, such as topic hierarchies and similarity heatmaps, greatly aiding in the interpretation of complex results. Its ability to classify each document within a dataset into identified topics streamlined the process of organizing and categorizing large text corpora.

The flexibility of BERTopic was further highlighted by its compatibility with other transformer-based models, such as GPT-4, making it adaptable to various types of text data and languages. This adaptability was particularly valuable in handling large datasets, a frequent challenge in real-world applications. The efficiency and efficacy of BERTopic were evident in its application to the extensive Twitter dataset, showcasing its utility in scenarios requiring deep contextual understanding. These scenarios ranged from analyzing customer feedback and academic research to monitoring social media trends and providing insights into political campaign actions on Twitter, underscoring the tool's significant contribution to the field of topic modeling and text analysis.

The depth and breadth of the data provided a rich foundation for extracting profound insights, effectively bridging modern technological advancements with sociopolitical analysis and underscoring the significance of the findings derived from this comprehensive examination, notably highlighting that political figures, despite their influential capabilities, do not singularly dictate the trajectory of political discourse. It becomes apparent that the general populace retains a measure of autonomous thought, thereby fostering a heterogeneous narrative landscape.

Notably, political candidates appear to strategically align their rhetoric with preexisting dialogues, aiming to amplify existing biases rather than forge new narratives. This tactic implies a preference for capitalizing on established trends rather than crafting them, underscoring a symbiotic relationship between political leadership and public discourse—a milieu where influence is reciprocally exerted rather than unilaterally imposed.

Furthermore, our findings invite a critical reassessment of the perceived omnipotence of authoritarian tactics like the 'firehose of falsehoods'. By overstating their efficacy, there is a risk of inadvertently neglecting effective countermeasures, thereby inadvertently bolstering the very agents of such misinformation. This revelation calls for a nuanced understanding of the dynamics at play, advocating for more strategic and informed approaches in navigating and mitigating the impact of such practices in our increasingly interconnected and politically charged digital landscape.

6. Future Works

Further scrutiny of electoral campaigns and their underhanded tactics is necessary. Analyzing firehosing through topic analysis in different electoral contexts could benefit civil society globally and empower citizens of the world, utilizing other models such as GPT integrated with BERTopic.
Analyzing the impact of gender, age, sex, race and other identification-based rhetoric on the public perhaps through sentiment analysis could contribute to prejudice-based rhetoric being diminished, and for a more equitable society.
Analyzing authoritharian countries' propaganda campaigns on social networks like twitter, tiktok, or, more importantly, reddit, could contribute to shedding light between cooperation between radicals in the west and in those countries, linking known propaganda accounts and famous influencers and reddit moderators through data analysis.

References

Bradner, E. (2016a). 6 takeaways from the first presidential debate. https://edition.cnn.com/2016/09/27/politics/winner-presidential-debate-takeaways/index.html.

Bradner, E. (2016b). 7 takeaways from the second presidential debate. https://edition.cnn.com/2016/10/10/politics/presidential-debate-takeaways/index.html.

Caetano, J. A., Almeida, J., and Marques-Neto, H. T. (2018a). Characterizing politically engaged users' behavior during the 2016 us presidential campaign. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.

Caetano, J. A., Lima, H. S., Santos, M. F., and Marques-Neto, H. T. (2017). Utilizando análise de sentimentos para definição da homofilia política dos usuários do twitter durante a eleição presidencial americana de 2016. *Congresso da Sociedade Brasileira de Computação-CSBC*.

Caetano, J. A., Lima, H. S., Santos, M. F., and Marques-Neto, H. T. (2018b). Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 american presidential election. *Journal of Internet Services and Applications*.

Caldwell, L. A. (2016). 5 major takeaways from the third and final 2016 presidential debate. https://www.nbcnews.com/storyline/2016-presidential-debates/5-major-takeaways-third-final-presidential-debate-n669831.

Grootendorst, M. (2020). Keybert: Minimal keyword extraction with bert.

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Kennedy, C., Blumenthal, M., Clement, S., Clinton, J. D., Durand, C., Franklin, C., McGeeney, K., Miringoff, L., Olson, K. M., Rivers, D., Saad, L., Witt, G. E., and Wlezien, C. (2020). An evaluation of the 2016 election polls in the united states. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1553context=sociologyfacpub.

Nawabdin, F. (2021). Perspectives on gender stereotypes: How did gender-based perceptions put hillary clinton at an electoral disadvantage in the 2016 election? https://www.degruyter.com/document/doi/10.1515/spp-2021-0014/html.

Oliveira, J. F. d., Karsai, M., and Marques-Neto, H. T. (2020). Information adoption via repeated or diversified social influence on twitter. *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 237–241.

Oliveira, J. F. d., Karsai, M., and Marques-Neto, H. T. (2022). Measuring the effects of repeated and diversified influence mechanism for information adoption on twitter. *Social Network Analysis and Mining*, 12(1):16.

Paul, C. and Matthews, M. (2016). *The Russian Firehose of Falsehood Propaganda Model: Why It Might Work and Options to Counter It*. RAND Corporation, Santa Monica, CA.

Pew Research Center (2018). For most trump voters, 'very warm' feelings for him endured. https://www.pewresearch.org/politics/2018/08/09/an-examination-of-the-2016-electorate-based-on-validated-voters. [Online; accessed 21-april-2023].

Savoy, J. (2018). Trump's and clinton's style and rhetoric during the 2016 presidential election. *Journal of Quantitative Linguistics*, 25(2):168–189.

Sorrentino, J. E. A. (2021). Deal me in: Hillary clinton and gender in the 2016 us presidential election. *Feminism Psychology*, 32(1):23–43.

Souza-Almeida, H. (2023). Trabalho de conclusão de curso. https://github.com/dezugin/tcc/blob/main/elusa/topic_bar.py.