TwiSP: A framework for exploring polarized issues in Twitter

Gabriela Andrea Díaz Carlos Iván Chesñevar Elsa Estevez Ana Maguitman
Departamento de Ciencias e Ingeniería de la Computación,
Universidad Nacional del Sur
Instituto de Cs. e Ing. de la Computación (ICIC UNS–CONICET)
Bahía Blanca, Buenos Aires, Argentina
{gabriela.diaz,cic,ece,agm}@cs.uns.edu.ar

ABSTRACT

Social and political polarization has become a dramatically intensifying force that is having a huge impact on political discourse, public policies and electoral outcomes in the 21st century. Twitter is a social media platform that mirrors to a large extent the sociological notion of public opinion, and has notably fueled these polarization dynamics worldwide. A proper understanding of how different issues become polarized in Twitter and their interrelationship is therefore crucial for the development of effective policies and governance strategies in our democracies. This paper introduces TwiSP, a framework for analyzing polarization on controversial topics in Twitter. TwiSP utilizes a combination of two cutting-edge machine learning techniques: stance detection for identifying attitudes and perspectives and BERTopic for topic modeling. The outcome of TwiSP is a visual tree-like representation of all tweets related to conflicting topics (rooted in a particular topic), contrasting their relationship using different colors to denote the degree of polarization. As a case study, we show how the TwiSP framework can be used for analyzing polarized issues in the context of the COVID-19 vaccine, exploring the resulting degree of polarization and the key topics driving it. The results reveal the diversity of opinions and the presence of highly polarized clusters in social media discussions. We contend that the TwiSP framework provides a novel and valuable tool for decision makers, helping them to recognize contentious issues behind the dynamics of polarization and ultimately identifying potential opportunities for bridging divides.

CCS CONCEPTS

• **Information systems** → **Data mining**; • **Applied computing** → **E-government**; • **Computing methodologies** → *Supervised learning by classification.*

KEYWORDS

Artificial Intelligence, E-Participation, Stance Detection, Social Media, Argumentation, Topic Modeling, Political Polarization

ICEGOV 2023, September 26–29, 2023, Belo Horizonte, Brazil

ACM ISBN 979-8-4007-0742-1/23/09…$15.00
https://doi.org/10.1145/3614321.3614324

ACM Reference Format:
Gabriela Andrea Díaz Carlos Iván Chesñevar Elsa Estevez Ana Maguitman. 2023. TwiSP: A framework for exploring polarized issues in Twitter. In *16th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2023), September 26–29, 2023, Belo Horizonte, Brazil.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3614321.3614324

1 INTRODUCTION AND MOTIVATION

The social phenomenon of polarization has significant implications for the present and future of governance and decision-making. In a highly polarized environment, it can be challenging to reach consensus and find common ground, which can hinder effective governance and policy-making, leading to a lack of trust in institutions and undermining the legitimacy of democratic processes [13, 32]. Moreover, the rise of digital platforms and social media (notably Twitter) has encouraged the creation of echo chambers [25], where users reinforce their own viewpoints and discredit the view points they do not agree with. As discussed in [16], this can potentially lead to a downward spiral of ever increasing political polarization, which, in turn, makes it harder to have a fact-based debate and to reach a consensus on controversial issues

Thus, it is clear that the ability to successfully navigate and understand polarization in social media will be critical for shaping the future of governance and ensuring the continued success of democratic societies. In this context, analyzing large collections of tweets using advanced machine learning techniques can provide valuable insights into the nature and extent of polarization within specific domains, helping policymakers to identify the most effective strategies for promoting healthy and productive public discourse. In the last years there have been different research efforts towards analyzing polarization and sentiment orientation in Twitter (e.g. [9], [5]). However, such approaches fall short in providing a global picture of polarization scenarios, identifying those trends and topics that are particularly sensible and introduce 'sentiment shifts' in public opinion.

The purpose of this ongoing research paper is to tackle the computational problem of processing millions of tweets associated with a particular topic which polarizes public opinion, identifying stances (attitudes or perspectives towards a particular topic or issue) and their interrelationships. The outcome of such a process is a *stance tree*, a visual representation which helps decision makers to better assess a polarization scenario, so that by zooming in this global picture different sentiment shifts (and their associated subtopics) in public opinion can be identified.

To achieve this, we propose TwiSP (Twitter Stance Processing), a framework that builds a hierarchical representation of conflicting positions in Twitter. This is accomplished by combining two cutting-edge machine learning techniques: stance detection for identifying attitudes and perspectives and BERTopic for topic modeling. Following the same spirit as in the saying "A picture is worth a thousand words", we contend that the proposed TwiSP framework provides the proper conceptualization needed for understanding polarizing issues in Twitter, advocating that "a stance tree is worth a million tweets".

This paper is inspired in preliminary findings presented in a previous publication [citation omitted for blind review], and extends significantly that research work (where a first approach for building stance trees from a collection of tweets was introduced). In the current article, we delve into a more evolved conceptualization, extending the notion of stance tree into a full-fledged framework. We discuss the different components associated with the architecture of the framework. As a case study, we present empirical results obtained from a massive dataset of tweets which account for public opinion concerning the sentiments about "COVID-19 vaccine". The tweets analyzed in this study were obtained from the datasets made available by Hayawi et al. [22] and Banda et al. [6]. The final dataset consists of more than 400,000 tweets, providing a rich and diverse source of opinions and perspectives.

The rest of this paper is structured as follows: in Section 2 we summarize the main concepts related to BERTopic, stance detection in social media and argumentation, which provide the background theory for our proposal. Then, in Section 3 we discuss related work in the area, pointing out some of the salient features of our proposal with respect to other alternative approaches. Section 4 introduces TwiSP, the proposed framework for analyzing Twitter data using machine learning techniques, organizing politically polarized content according to topic and stance. Section 5 presents a full-fledged case study related to the COVID-19 vaccine issue. Finally, Section 6 presents the main conclusions that have been obtained, discussing some research lines for future work.

2 BACKGROUND

As discussed in the introduction, our proposal is based on the combination of machine learning techniques, along with a particular tree-like representation (called *stance tree*) which accounts for possibly conflicting relationships among topics when considering polarizing issues. Next we will discuss the main aspects associated with topic modeling, stance detection and argumentation, providing as well a brief analysis of how these three elements are intertwined in the formalization of the TwiSP framework presented in Section 4.

In recent years, the explosion of user-generated content on social media platforms has led to a growing interest in methods for automatically analyzing and categorizing text data. BERTopic [18] is a state-of-the-art technique for topic modeling that uses pre-trained language models to identify clusters of similar documents based on semantic similarity. This method has shown promising results in a variety of applications, from sentiment analysis to content recommendation. BERTopic uses the Bidirectional Encoder Representations from Transformers (BERT) model [14] to cluster similar documents together and extract representative keywords for each cluster. It is a powerful unsupervised machine learning method that can be used to automatically group large collections of textual data into topics or themes, without the need for prior knowledge of the number or nature of topics present in the data. The BERTopic algorithm is based on the idea of projecting the embeddings of the documents into a lower-dimensional space, where clustering can be performed efficiently using a technique called hierarchical density-based spatial clustering of applications with noise (HDBSCAN) [29]. BERTopic is particularly useful for exploratory data analysis, document clustering, and topic modeling in domains such as social media, news, and academic literature.

Stance detection [4, 31] refers to the task of automatically identifying the attitudes or perspectives that a speaker or writer has towards a particular topic, issue, or entity. This involves analyzing natural language texts, such as tweets, news articles, or online comments, to determine whether the author is in favor of, against, or neutral towards the subject at hand. Stance detection is a key component of sentiment analysis and can be used in various applications, including political opinion tracking, market research, and brand monitoring. The ability to automatically identify the stance of a given text as pro, anti, or neutral towards a particular topic can provide valuable insights into public opinion on contentious issues.

Argumentation is an essential aspect of communication. Individuals rely on arguments to develop opinions, make decisions, or persuade others to adopt a particular stance. In recent years, there has been a growing interest in the computational analysis and synthesis of natural language argumentation from text [7, 21, 27]. The idea behind organizing arguments to build a tree-like structure, also known as a *dialectical tree*, has a long tradition in argumentation theory. A dialectic tree represents the different arguments related to a particular topic and how they are related to each other [8]. Dialectic trees are rooted on the main issue or question that is being discussed. Those arguments that are relevant to that issue can be organized in a hierarchical structure, where each argument is linked to its parent argument based on the type of relationship between them. For example, an argument can be used to support or attack another argument. The resulting tree-like structure can be used to represent the different perspectives and positions related to the particular topic, as well a to visualize the relationships between them.

As pointed out in [15], the last decade has witnessed a return to the early AI goal of understanding and building human-like intelligent systems that operate in a cognitively-compatible and synergistic way with humans. The growing market demand for AI systems of different kinds has encouraged the development of the so-called *Human-Centric AI*, which aims to deliver services within the realm of natural or commonsense intelligence to support and enhance the users' natural capabilities. In this context, machine learning algorithms for topic modelling and stance detection are extremely efficient and can hardly be outperformed by other alternative human-centered approaches. Nevertheless, their efficiency is achieved at a very high price, since they behave like *black boxes*, being unable to explain their outcomes, and having consequently very low accountability and transparency when used in intelligent decision making.

As clearly explained in [15], the efficiency of black-box approaches falls short when defining human-centric systems, as *[...] building such human-centric systems necessitates a foundational shift in the problem-solving paradigm that moves away from the strictness and absolute guarantees of optimal solutions that are typically adopted* *for conventional computing, which are often brittle and break down completely when new information is acquired. Instead, human-centric systems would benefit by adopting satisficing solutions that strike an acceptable balance between a variety of criteria, are tolerant to uncertainty and the presence of incompatible alternatives, are robust across a wide range of problem cases, and are elastic in being gracefully adapted when they are found to have become inappropriate or erroneous in the face of new information.*

We agree with the above reasoning, and contend that analyzing polarized issues in Twitter cannot be handled by machine learning algorithms alone, as policymakers need to be provided with an accountable and clear picture of the different elements involved (e.g. conflicting topics, sentiment shifts, etc.) The TwiSP framework provides thus a combination of machine learning and computational argumentation, where the resulting stance tree accounts for a structured provision of explanations or alternatives, rather than focusing on achieving "the best" possible answer. We deploy the full capabilities of machine learning to process efficiently millions of tweets (identifying relevant topics and their conflicts), while keeping a stance tree as an explanation-based representation which is intended to be cognitively compatible with the ultimate human users.

3 RELATED WORK

Political polarization has been on the rise in recent years, and social media has been identified as a contributing factor [26]. Social media platforms have enabled the proliferation of echo chambers and filter bubbles [25], where individuals are more likely to be exposed to information that confirms their existing beliefs and less likely to be exposed to information that challenges them. This has led to an increase in political polarization, with individuals becoming more entrenched in their beliefs and less willing to engage with opposing viewpoints.

The study of social media (notably Twitter as a microblogging platform) has become an increasingly important area of research in recent years, particularly in the context of social and political polarization. One approach to studying polarization in social media involves analyzing *network topologies* and the spread of content through retweets [3, 11, 38]. By examining the structure of networks of users and the patterns of information sharing, researchers can gain insights into the dynamics of polarization and how it manifests in online communities. On the other hand, misinformation has become a significant concern on social media platforms, particularly during crises, where it can spread quickly and widely [30]. False or misleading information can fuel polarization and hinder efforts to address critical issues such as public health crises or political events. Moreover, the proliferation of misinformation can create confusion, fear, and distrust among the public, potentially leading to negative social and economic consequences. Thus, understanding the role of misinformation in polarized issues on social media is crucial for identifying potential sources of conflict and developing effective strategies to combat its effects [33, 34].

Network-based approaches have also played a role for *echo chamber detection* in social networks, as discussed in [25]. Echo chamber detection refer to identifying and analyzing the formation and reinforcement of homogeneous groups in online communities, where individuals with similar beliefs and attitudes tend to congregate and reinforce each other's views. This phenomenon is a concern as it can lead to the spread of misinformation and the reinforcement of harmful stereotypes and biases. The authors investigate whether purely structural analysis of communication and interaction patterns can be used to identify echo chambers, and present a network-based, macro-scale, polarity-based approach for detecting echo chambers in Twitter.

The COVID-19 pandemic and the subsequent vaccination campaigns [37] have received significant attention on social media platforms, making them an ideal topic for stance detection analysis. These and other significant events have brought to the forefront the importance of effective communication during crises [24, 30, 36]. In the digital age, social media platforms have emerged as key tools for disseminating information and coordinating emergency response efforts. However, the widespread use of social media has also led to the rapid spread of misinformation and the polarization of public opinion on sensitive issues. Thus, analyzing social media data to understand the polarization of stances on such events has become increasingly important.

Topic modeling is a powerful technique that has been widely used in social media analysis to understand what concerns people about a specific issue. By applying topic modeling to social media data, we can identify the most discussed topics and the key themes within those discussions. These topics can then be analyzed further to gain a better understanding of people's opinions, beliefs, and attitudes towards the issue in question. In particular, topic modeling has proven useful in understanding the public's concerns and sentiments during crises, such as the COVID-19 pandemic as done in [2]. By analyzing social media data related to the pandemic, researchers have been able to identify the most pressing concerns and topics, such as vaccine efficacy, mask mandates, and government responses, and use this information to inform public health policies and communication strategies.

Recent studies on detecting stance in Twitter have a statistical foundation and offer various metrics to address misinformation and polarized opinions (e.g. [28]). Our proposal differs from these approaches in that it not only provides a classification for removing tweets that are considered "uninformative" for downstream epidemiological analyses, but also offers a more structured view or taxonomy of the tweets.

The work presented in [20] represents collective arguments through sets of tweets that are syntactically linked to a query Q. In contrast, our approach generates arguments by employing a soft clustering algorithm to calculate a coherent set of tweets. Therefore, our method represents an evolution of the concept of opinion tree, as it employs a supervised model to identify polarization rather than performing a syntactic search on the Twitter database.

Chesñevar et al. [10] propose the use of computational argumentation as a foundation for intelligent decision-making in cognitive cities, advocating for the usefulness of "opinion trees" [19, 20] in handling conflicting opinions on social media. They argue that Twitter provides valuable information for extracting computational arguments, referred to as "opinions", which emerge bottom-up from social interactions surrounding messages. By incrementally generating queries, opinions can be mined from Twitter, resulting in an opinion tree rooted in the original query. Unlike Chesñevar et al.'s approach, our method utilizes a soft clustering algorithm to generate cohesive sets of tweets and a supervised model to identify polarization rather than relying solely on Twitter queries for argument extraction.

In [17], the authors propose a two-phase classification system for stance detection in tweets, exploiting topic modeling features. Explanation of stance labels is provided through the most relevant terms within topics over the tweets. In contrast with our approach, this research work does not consider how to deal with conflicts between topics as done in our paper. A combination of machine learning techniques is also applied in [35], where the authors consider an ensemble-based model for stance detection on social networks. Random Forest and Support Vector Machines are used as base learners, which are combined into three alternative ensemble models. Experimental results show that the proposed ensembles outperform the state-of-the-art models in the overall test score. In contrast with our approach, classification efforts are aimed at stance detection without considering conflicts among topics, nor the relationship between such conflicts, as done in our paper.

4 PROPOSED FRAMEWORK

In this section, we describe TwiSP, a novel framework for analyzing twitter data that leverages machine learning techniques to organize politically polarized content according to topic and stance. The TwiSP framework is composed of four main components as outlined in Figure 1.

The **data collection** component is in charge of collecting tweets relevant to an issue under analysis. One way to collect tweets is by using the Twitter Search API. This allows to search for tweets that contain specific keywords or hashtags related to the issue being analyzed. In particular, the API can be used to retrieve tweets from a specific time range or location. Also, it is possible to retrieve a real-time stream of tweets based on various parameters, such as keywords, hashtags, locations, languages, etc., by using the Twitter Streaming API. Alternatively, it is possible to access datasets with Twitter data related to specific topics that have already been collected and curated for research purposes. This component is also in charge of removing duplicates, which is an important step to ensure that the analysis is based on unique opinions and not skewed by a small group of highly active users who may have tweeted the same message multiple times. By ensuring that each tweet in the dataset is unique, it is possible to more accurately capture the diversity of public opinion and produce a more reliable analysis.

The **stance detection** component is in charge of training a model capable of predicting the stance of individual tweets. To achieve this, the proposed framework uses the Hugging Face Transformers library to load a pre-trained BERT model into TensorFlow. Then, a classification layer is added on top of the BERT model to fine-tune it for stance prediction. Fine-tuning is achieved by training the model using a dataset with tweets labeled as anti, pro or neutral with respect to the issue under analysis. The fine-tuned stance prediction model is applied to the collection of unlabeled tweets to build a dataset with predicted labels. Since the goal is to focus the analysis on highly polarized opinions, tweets predicted as neutral are excluded from the resulting dataset. While these tweets may still contain valuable information, including them in the analysis could dilute the degree of polarization within the dataset. By only including tweets that are predicted to be either anti or pro with respect to the issue under analysis, it is possible to create a more focused dataset that better represents the extremes of public opinion on the analyzed topic. The stance associated with each tweet represents a valuable signal for further analysis. Hence, the text of each tweet was extended with a tag indicating the predicted stance, either *Stance_Anti* or *Stance_Pro*.

The **topic modeling** component is in charge of generating a hierarchical visualization of the topics. The BERTopic library is used to perform topic modeling and clustering analysis. Specifically, it uses the agglomerative clustering method, which is a bottom-up hierarchical clustering approach that starts by assigning each tweet to its own topic and iteratively merges them into larger clusters based on their similarity. The result is a dendrogram, which is a hierarchical diagram representing the arrangement of topics by showing their relationships in a tree-like structure. In this hierarchy, the more general topics are located at the top, while the more specific ones are at the bottom. Through this method, it becomes possible to hierarchically visualize the relationships among various topics and subtopics, with more closely related topics being grouped together at lower levels of the tree.

Finally, the **stance tree visualization** component is in charge of assessing the stance of each topic with the purpose of building a stance tree. To this end, the nodes representing each topic are color-coded according to their level of polarization as shown in Figure 2. Nodes with a high proportion of tweets that take an anti stance towards the analyzed issue are colored in red, while nodes that express a pro position are colored in green. Nodes colored in yellow represent topics with high levels of polarization. The resulting colored dendrogram represents a stance tree. This approach offers an intuitive and informative visualization of the topics aligned with different stances and their relationships, making it possible to identify the most polarized topics and gain a deeper understanding of the underlying patterns in public opinion on the issue of interest.

Figure 2: Color palette used to represent the degree of polarization for each topic node in the tree. The color of the node depends on the number of tweets with anti or pro stances for that particular topic. Red nodes represent topics with a high proportion of an anti stance, green nodes represent topics with a high proportion of a pro stance, and yellow nodes represent topics with polarized positions.

5 CASE STUDY

In this case study, we used a variety of datasets to train, validate and test the stance prediction model and generate a stance tree. Specifically, we started by selecting a random sample of 1050 tweets from the ANTi-Vax dataset [23] and manually labeled them with stance information (anti-vaccine, pro-vaccine, or neutral) as it was described in [citation omitted for blind review]. In addition to the ANTi-Vax dataset, we also used 1944 labeled tweets from the dataset

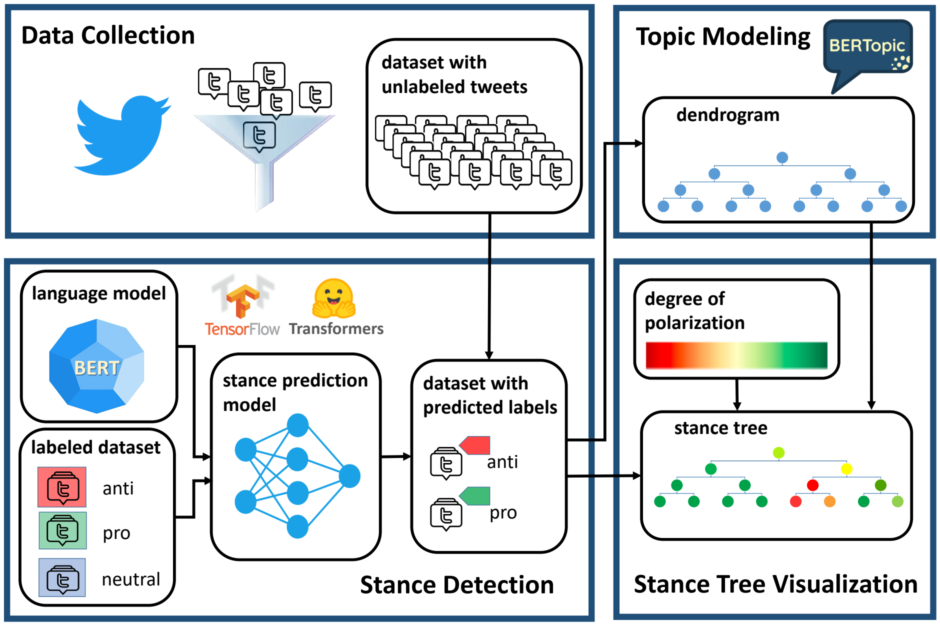

Figure 1: Overview of the TwiSP framework, consisting of four main components: (1) Data Collection, (2) Stance Detection, (3) Topic Modeling, and (4) Stance Tree Visualization. The framework enables the analysis of social media data by identifying relevant topics, assessing attitudes and beliefs towards those topics, and visualizing the relationships between them.

made available by Cotfas et al. [12] and 2054 labeled tweets from the VADET dataset [39].

The resulting labeled dataset was composed of 5048 tweets with the class distribution shown in Figure 3. To develop and assess the performance of the stance prediction model, we split the labeled dataset into three parts: 80% for training, 10% for validation, and 10% for testing. This allowed us to train and validate the model on a large enough dataset to capture the nuances and complexities of public opinion on vaccination, while also testing the model's performance on new, unseen data. The accuracy achieved on the test dataset was higher than 85% (details on the evaluations will be presented in a future paper). This approach allowed us to develop a highly accurate and robust model that could effectively predict the stance of new tweets on the topic of vaccination.

To further expand the scope of our study and gain a deeper understanding of public opinion on vaccination, we utilized an unlabeled dataset of 1,440,705 tweets from VADET [39]. Using our trained stance prediction model, we were able to predict the stance of each of these tweets, allowing us to analyze a much larger and more diverse set of public opinions on vaccination.

After applying the stance prediction model, filtering out neutral tweets and removing duplicates, we obtained a dataset of 462,254 unique tweets that were predicted to be either pro- or anti-vaccination.

After applying topic modeling on the dataset with predicted labels, the TwiSP framework generated a total of 4219 distinct topics. However, some of the topics were considered outliers and were consolidated into a single "noisy" topic, which was later discarded. Once the topics were created, it was possible to calculate the degree to which each topic was associated with a pro or anti-vaccine stance,

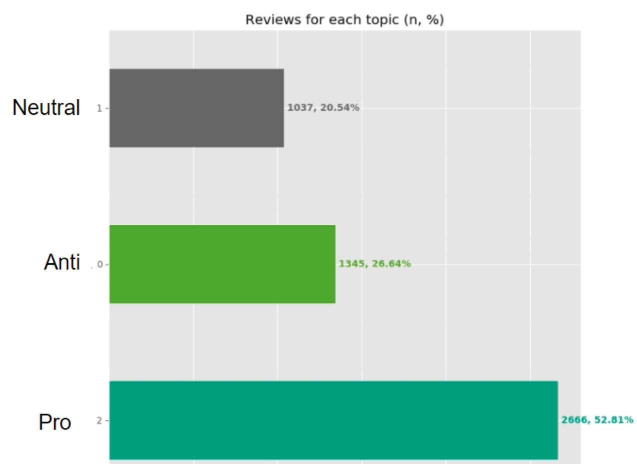

Figure 3: Class distribution of the dataset used to train, validate and test the stance prediction model. The dataset contains 5048 tweets labeled with stance information.

which allowed us to identify different levels of polarization. Finally, the TwiSP framework offered a visualization of the resulting stance tree. This visualization provided a comprehensive view that enabled us to gain valuable insights into the attitudes and perspectives of Twitter users on the topic of the "COVID-19 vaccine." Figure 4 displays a portion of the generated stance tree, as the complete tree is too large to be shown in the paper.

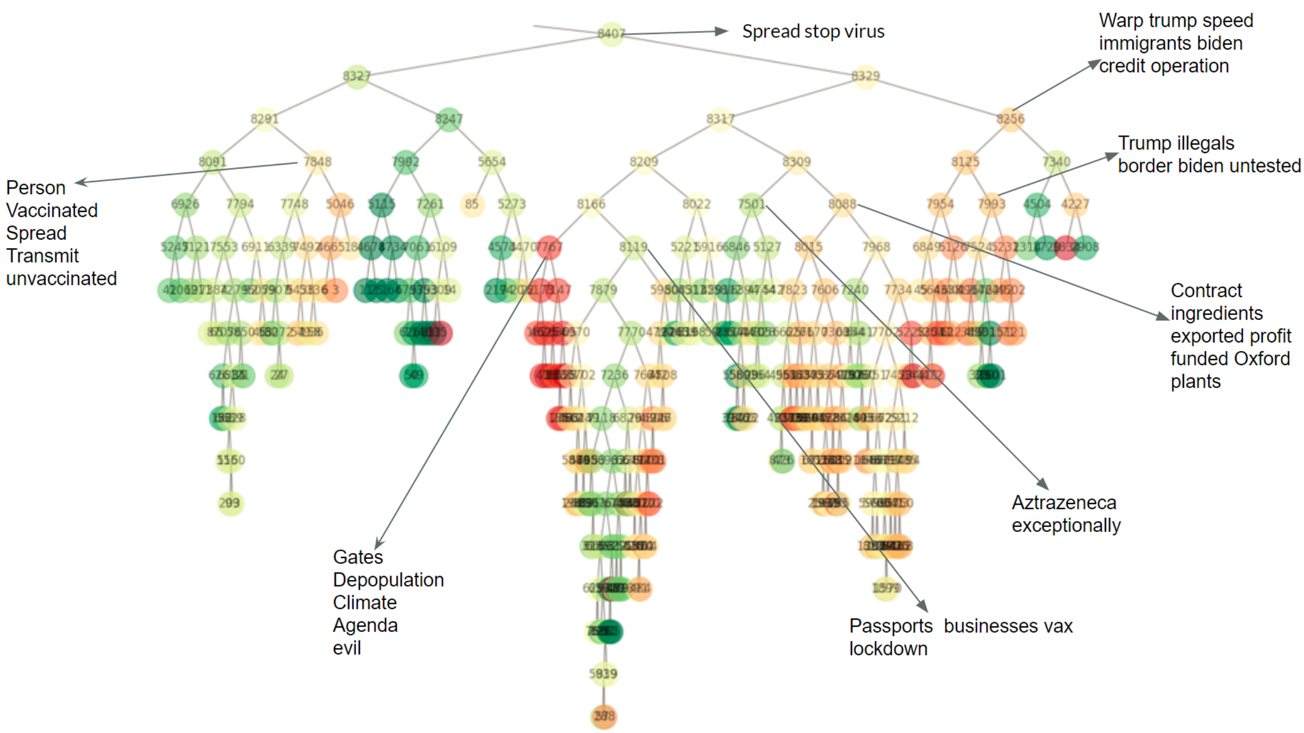

Figure 4: Sub-tree of the "COVID-19 vaccine" issue. The image depicts a section of the tree that was generated for the topic of "COVID-19 vaccine". The nodes in the image represent subtopics related to the main topic, and their colors indicate the degree of polarization. Some of the nodes in the image show a high degree of polarization, indicating a marked difference in public opinion on those subtopics. By analyzing the tree, we were able to gain insight into the most contentious issues related to the COVID-19 vaccine and the degree of polarization surrounding them. This information can be valuable in understanding public opinion on the vaccine and informing public health campaigns and policies.

The results of our analysis showed that there are several topics with high polarization, represented by yellow nodes, indicating divergent viewpoints within the analyzed issue. It is often possible to identify subtopics with either a positive or negative stance by examining the subtrees that stem from polarized nodes, which are denoted by green and red nodes, respectively. For instance, some highly polarized topics can be seen in the cluster that discusses illegal immigrants and untested individuals. This suggests that this topic is a highly controversial issue within the domain being analyzed, with strong disagreements in opinion. Other examples of polarized topics associated with COVID-19 vaccination include vaccine passports, vaccine mandates for business, and restrictions on activities such as travel and work through lockdowns. Another polarized topic is evoked by the fact that the University of Oxford, which developed the AstraZeneca vaccine, is engaging in profit-making activities from selling the vaccine, despite its claim to be a non-profit institution [1]. Additionally, some topics are associated with the concerns raised about the ingredients in some COVID-19 vaccines, such as the use of fetal cell lines in their development. By identifying these highly polarized topics, it is possible to gain insights into the discourse surrounding the issue and better understand the different perspectives and positions held by various groups within the domain.

Some topics that express a negative stance mention notable individuals or organizations in connection with the idea of depopulation. This is usually due to conspiracy theories circulating on social media that falsely claim the existence of a plot to use the vaccine as a tool for population control or even implant microchips in people. These claims have been debunked by numerous fact-checking organizations and public health authorities, but they continue to spread on social media platforms. TwiSP has the potential to assist authorities in identifying this kind of misleading information, which can lead to implementing appropriate measures to enhance public awareness, such as targeted information campaigns.

6 CONCLUSION

In this paper, we presented TwiSP, a framework for analyzing polarization in public opinion by training a stance prediction model on a labeled dataset of tweets and using it to predict stances on a large, unlabeled dataset. The framework also applies the BERTopic algorithm to identify the main topics in the tweets and generate a hierarchical visualization that allowed us to analyze different levels of polarization within each topic. We applied this framework the "COVID-19 vaccine" issue and found that many of the topics were highly polarized, reflecting the contentious nature of this issue.

By analyzing the polarization within each topic, we were able to gain insights into the main arguments and concerns expressed by individuals on different sides of the issue. Our results demonstrate the potential of the TwiSP framework for analyzing polarization in public opinion and identifying the underlying reasons for this polarization. This analysis may prove valuable for policymakers and other stakeholders who aim to implement suitable measures for improving public knowledge and understanding by information campaigns aimed at specific groups or demographics.

In our upcoming work, we intend to apply the TwiSP framework on other subjects that have polarized views such as "climate change" or "gun control". The primary limitation to applying a comparable analysis to other topics is the need for a labeled dataset to train our stance detection models. In line with the methodology employed in this study, our future research will involve the utilization of transfer learning, where a pre-trained large language model is fine-tuned on a small amount of labeled data specific to the task of stance detection. We anticipate that the requirement for labeled data may decrease in the future, as the accessibility and prevalence of large language models increase. Furthermore, as a part of our forthcoming research, we intend to devise alternative algorithmic techniques to construct stance trees, with a specific emphasis on identifying collective arguments against or in support of a particular topic.

ACKNOWLEDGMENTS

This work was enabled by support provided by CONICET (PUE Code 229 201801 00020 CO), Universidad Nacional del Sur (PGI-UNS 24/N051) and ANPCyT (PICT 2018-03627 and PICT 2019-03944). We thank anonymous reviewers for their constructive comments.

REFERENCES

[1] [n.d.]. AstraZeneca revenues soar with nearly $4bn in Covid-19 vaccine sales. https://www.pharmaceutical-technology.com/news/astrazeneca-revenue-soars-nearly-4bn-covid-19-vaccine-sale/. Last accessed: 5th May, 2023.

[2] Alaa Abd-Alrazaq, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, and Zubair Shah. 2020. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *Journal of Medical Internet Research* (2020). https://doi.org/10.2196/19016

[3] Lada A Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*. 36–43.

[4] Abeer AlDayel and Walid Magdy. 2020. Stance Detection on Social Media: State of the Art and Trends. *CoRR* abs/2006.03644 (2020). arXiv:2006.03644 https://arxiv.org/abs/2006.03644

[5] Mohd Zeeshan Ansari, M.B. Aziz, M.O. Siddiqui, H. Mehra, and K.P. Singh. 2020. Analysis of Political Sentiment Orientations on Twitter. *Procedia Computer Science* 167 (2020), 1821–1828. https://doi.org/10.1016/j.procs.2020.03.201 International Conference on Computational Intelligence and Data Science.

[6] Juan M. Banda, Ramya Tekumalla, Guanyu Wang, Jingyuan Yu, Tuo Liu, Yuning Ding, Ekaterina Artemova, Elena Tutubalina, and Gerardo Chowell. 2021. A Large-Scale COVID-19 Twitter Chatter Dataset for Open Scientific Research—An International Collaboration. *Epidemiologia* 2, 3 (2021), 315–324. https://doi.org/10.3390/epidemiologia2030024

[7] Roy Bar-Haim, Lilach Eden, Roni Friedman, Yoav Kantor, Dan Lahav, and Noam Slonim. 2020. From arguments to key points: Towards automatic argument summarization. *arXiv preprint arXiv:2005.01619* (2020).

[8] Pietro Baroni, Dov Gabbay, Massimilino Giacomin, and Leendert van der Torre. 2018. *Handbook of Formal Argumentation*. London, England: College Publications.

[9] Staphord Bengesi, Timothy Oladunni, Ruth Olusegun, and Halima Audu. 2023. A Machine Learning-Sentiment Analysis on Monkeypox Outbreak: An Extensive Dataset to Show the Polarity of Public Opinion From Twitter Tweets. *IEEE Access* 11 (2023), 11811–11826. https://doi.org/10.1109/ACCESS.2023.3242290

[10] Carlos Iván Chesñevar, María Paula González, Ana Gabriela Maguitman, and Elsa Estevez. 2020. A first approach towards integrating computational argumentation in cognitive cities. In *ICEGOV 2020: 13th International Conference on Theory and Practice of Electronic Governance, Athens, Greece, 23-25 September, 2020*, Yannis Charalabidis, Maria Alexandra Cunha, and Demetrios Sarantis (Eds.). ACM, 25–32. https://doi.org/10.1145/3428502.3428506

[11] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, Vol. 5. 89–96.

[12] Liviu-Adrian Cotfas, Camelia Delcea, Ioan Roxin, Corina Ioanăş, Dana Simona Gherai, and Federico Tajariol. 2021. The Longest Month: Analyzing COVID-19 Vaccination Opinions Dynamics From Tweets in the Month Following the First Vaccine Announcement. *IEEE Access* 9 (2021), 33203–33223. https://doi.org/10.1109/ACCESS.2021.3059821

[13] Philipp Darius. 2022. Who polarizes Twitter? Ideological polarization, partisan groups and strategic networked campaigning on Twitter during the 2017 and 2021 German Federal elections 'Bundestagswahlen'. *Soc. Netw. Anal. Min.* 12, 1 (2022), 151. https://doi.org/10.1007/s13278-022-00958-w

[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[15] Emmanuelle Dietz, Antonis C. Kakas, and Loizos Michael. 2022. Argumentation: A calculus for Human-Centric AI. *Frontiers Artif. Intell.* 5 (2022). https://doi.org/10.3389/frai.2022.955579

[16] Venkata Rama Kiran Garimella and Ingmar Weber. 2017. A long-term analysis of polarization on Twitter. In *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*. AAAI PRESS, United States, 528–531. https://www.icwsm.org/2017/ International AAAI Conference on Web and Social Media, ICWSM ; Conference date: 15-05-2017 Through 18-05-2017.

[17] Manuela Gómez-Suta, Julián David Echeverry Correa, and José A. Soto Mejía. 2023. Stance detection in tweets: A topic modeling approach supporting explainability. *Expert Syst. Appl.* 214 (2023), 119046. https://doi.org/10.1016/j.eswa.2022.119046

[18] Maarten Grootendorst. [n.d.]. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794 [cs] http://arxiv.org/abs/2203.05794

[19] Kathrin Grosse, Carlos Iván Chesñevar, and Ana Gabriela Maguitman. 2012. An Argument-based Approach to Mining Opinions from Twitter. In *AT*. 408–422. http://ceur-ws.org/Vol-918/111110408.pdf

[20] Kathrin Grosse, Maria P Gonzalez, Carlos I Chesnevar, and Ana G Maguitman. 2015. Integrating argumentation and sentiment analysis for mining opinions from Twitter. *AI Communications* 28, 3 (2015), 387–401.

[21] Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics* 43, 1 (2017), 125–179.

[22] K. Hayawi, S. Shahriar, M. A. Serhani, I. Taleb, and S. S. Mathew. [n.d.]. ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection. 203 ([n. d.]), 23–30. https://doi.org/10.1016/j.puhe.2021.11.022

[23] Kadhim Hayawi, Sakib Shahriar, Mohamed Adel Serhani, Ikbal Taleb, and Sujith Samuel Mathew. 2022. ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection. *Public health* 203 (2022), 23–30.

[24] Sameera Horawalavithana, Kin Wai Ng, and Adriana Iamnitchi. 2021. Drivers of polarized discussions on twitter during venezuela political crisis. In *13th ACM Web Science Conference 2021*. 205–214.

[25] Nane Kratzke. 2023. How to Find Orchestrated Trolls? A Case Study on Identifying Polarized Twitter Echo Chambers. *Computers* 12, 3 (2023). https://doi.org/10.3390/computers12030057

[26] Emily Kubin and Christian von Sikorski. 2021. The role of (social) media in political polarization: a systematic review. *Annals of the International Communication Association* 45, 3 (2021), 188–206. https://doi.org/10.1080/23808985.2021.1976070 arXiv:https://doi.org/10.1080/23808985.2021.1976070

[27] John Lawrence and Chris Reed. 2020. Argument mining: A survey. *Computational Linguistics* 45, 4 (2020), 765–818.

[28] Christian E. López, Malolan Vasu, and Caleb Gallemore. 2020. Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset. *CoRR* abs/2003.10359 (2020). arXiv:2003.10359 https://arxiv.org/abs/2003.10359

[29] Leland McInnes, John Healy, and Steve Astels. 2017. HDBSCAN: Hierarchical density based clustering. *J. Open Source Softw.* 2, 11 (2017), 205.

[30] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter under crisis: can we trust what we RT? *null* (2010). https://doi.org/10.1145/1964858.1964869

[31] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. [n.d.]. SemEval-2016 Task 6: Detecting Stance in Tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (San Diego, California, 2016). Association for Computational Linguistics, 31–41. https://doi.org/10.18653/v1/S16-1003

[32] Jacqueline M. Otala, Gillian Kurtic, Isabella Grasso, Yu Liu, Jeanna N. Matthews, and Golshan Madraki. 2021. Political Polarization and Platform Migration: : A Study of Parler and Twitter Usage by United States of America Congress Members. In *Companion of The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Eds.). ACM / IW3C2, 224–231. https://doi.org/10.1145/3442442.3452305

[33] Francesco Pierri, Matthew R DeVerna, Kai-Cheng Yang, David Axelrod, John Bryden, and Filippo Menczer. 2023. One Year of COVID-19 Vaccine Misinformation

on Twitter: Longitudinal Study. *Journal of Medical Internet Research* 25 (2023), e42227.

[34] Francesco Pierri, Brea L Perry, Matthew R DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer, and John Bryden. 2022. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. *Scientific reports* 12, 1 (2022), 5966.

[35] Sara S. Sherif, Doaa M. Shawky, and Hatem A. Fayed. 2023. An Ensemble Model for Stance Detection in Social Media Texts. *Int. J. Inf. Technol. Decis. Mak.* 22, 2 (2023), 737–775. https://doi.org/10.1142/S0219622022500481

[36] Simone Splendiani and Antonella Capriello. 2022. Crisis communication, social media and natural disasters–the use of Twitter by local governments during the 2016 Italian earthquake. *Corporate Communications: An International Journal* (2022).

[37] Maxwell A. Weinzierl and Sanda M. Harabagiu. 2022. Identifying the Adoption or Rejection of Misinformation Targeting COVID-19 Vaccines in Twitter Discourse. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, Frédérique Laforest, Raphaël Troncy, Elena Simperl, Deepak Agarwal, Aristides Gionis, Ivan Herman, and Lionel Médini (Eds.). ACM, 3196–3205. https://doi.org/10.1145/3485447.3512039

[38] Sarita Yardi and Danah Boyd. 2010. Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of science, technology & society* 30, 5 (2010), 316–327.

[39] Lixing Zhu, Zheng Fang, Gabriele Pergola, Rob Procter, and Yulan He. 2022. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. arXiv:2205.03296 [cs.CL]