Temporal Analysis of Brazilian Presidential Election on Twitter Based on Formal Concept Analysis

Daniel Pereira[a], Julio Neves[b], Wladmir Brandão[c] and Mark Song[d]
Universidade Católica de Minas Gerais, Computer Science Department, Minas Gerais, Brazil
{danielmop, juliocesar.neves}@gmail.com, {wladmir, song}@pucminas.br

Keywords: Topic Evolution, Twitter, Formal Concept Analysis, Social Network Analysis.

Abstract: Social networks have become an environment where users express their feelings and share news in real-time. However, analyzing the content produced by users is not a simple task, given the volume of posts. It is important to comprehend the expressions made by users to gain insights into politicians, public figures, and news. The state-of-the-art lacks studies that propose how the topics discussed by social network users change over time. In this context, this work measures how topics discussed on Twitter vary over time. Formal Concept Analysis was used to measure how these topics were varying, considering the support and confidence metrics. Our solution was tested on tweets related to the Brazilian presidential election. The results confirm that it is possible to comprehend what Twitter users were discussing and how these topics changed over time. Our work is beneficial for politicians seeking to analyze the discussions about them among users. Our analysis of 3,634 tweets revealed several significant patterns, such as the association between political figures and topics like fake news and election fraud. These findings demonstrate how social media discussions evolve during key political events, providing insights that can assist political campaigns in real-time.

1 INTRODUCTION

The Internet is no longer just a repository for documents to be shared, it is now a hybrid space for different media and applications that reach a large audience (Zhang et al., 2012). Some of these applications are social networks, which allow their users to generate a large amount of content that exemplifies their impressions and experiences. A specific social network that stands out for forcing its users to express themselves concisely is Twitter. On Twitter, users express themselves through tweets, which consist of text content with a maximum length of 280 characters.

The fact that the Tweet is a short textual model allows users to quickly report what they are experiencing at the time the post is posted, unlike a journalist, for example, who, to generate a story, needs to ensure its excellence. Since Twitter users report their experiences without worrying about their writing or who will read their text, Twitter is probably the fastest means of disseminating information in the world (Cataldi et al., 2010).

With this large amount of information provided, it is hard to extract knowledge from a group of tweets. This task is relevant for politicians, for example, to check the opinions that users are expressing about them. Therefore, it would be relevant to develop a tool that is capable of analyzing and extracting knowledge from a group of tweets.

An alternative to solve this challenge is through Natural Language Processing (NLP) and Formal Concept Analysis (FCA). The objective of our work is to use NLP to find recurrent groups of words from tweets and then analyze how these groups of words relate to each other. The relation between these terms is measured using FCA, using the metrics of support and confidence. Also, how these terms change over time is another metric analyzed.

While social media platforms like Twitter offer rich datasets for sentiment analysis, existing methods often fall short in tracking topic evolution over time. Traditional NLP techniques primarily focus on static analysis, which limits their ability to capture the dynamic nature of online discussions. This work aims to address these gaps by employing Formal Concept Analysis FCA to observe how political topics evolve during critical events, such as elections.

[a] https://orcid.org/0009-0005-2276-2334
[b] https://orcid.org/0000-0002-0520-9976
[c] https://orcid.org/0000-0002-1523-1616
[d] https://orcid.org/0000-0001-5053-5490

Pereira, D., Neves, J., Brandão, W. and Song, M.
Temporal Analysis of Brazilian Presidential Election on Twitter Based on Formal Concept Analysis.
DOI: 10.5220/0012869300003825

In *Proceedings of the 20th International Conference on Web Information Systems and Technologies (WEBIST 2024)*, pages 167-174
ISBN: 978-989-758-718-4; ISSN: 2184-3252

A case study to solve the problem was used. The case study consists of analyzing tweets that discuss the Brazilian presidential election, checking which terms are related to each candidate and how they evolve over time.

The remainder of the paper is organized as follows: the background is outlined in Section 2. The Literature Review is described in Section 3. Section 4 presents the defined Methodology. Results are discussed in Section 5. The conclusion and further research are in Section 6.

2 BACKGROUND

2.1 Formal Concept Analysis

FCA is a technique based on formalizing the notion of concept and structuring concepts in a conceptual hierarchy. FCA relies on lattice theory to structure formal concepts and enable data analysis. The capability to hierarchize concepts extracted from data makes FCA an interesting tool for dependency analysis. With the increase of social networks and due to the large amount of data generated by users, the study and improvement of techniques to extract knowledge are becoming increasingly justified. Also, it permits the data analysis through associations and dependencies attributes, and objects, formally described, from a dataset.

Formally, a formal context is formed by a triple (G,M,I), where G is a set of objects (rows), M is a set of attributes (columns) and I is defined as the binary relationship (incidence relation) between objects and their attributes where $I \subseteq G \times M$.

Table 1 exemplifies a formal context. In this example, objects correspond to tweets, attributes are the characteristics (terms), and the relationship of incidence represents whether or not the tweet has that characteristic. An $'X'$ is present in the table if the tweet possesses the corresponding characteristic.

Table 1: Formal Context Example.

	Lula	Bolsonaro	Fake News	Elections
Tweet 1	**X**			
Tweet 2		**X**		**X**
Tweet 3	**X**	**X**		**X**
Tweet 4			**X**	

2.2 Formal Concepts

Let (G,M,I) be a formal context, $A \subseteq G$ a subset of objects and $B \subseteq M$ a subset of attributes. Formal concepts are defined by a pair (A,B) where $A \subseteq G$ is called extension and $B \subseteq M$ is called intention. This pair must follow the conditions where $A = B'$ and $B = A'$ (Ganter and Wille, 1999). The relation is defined by the derivation operator ($'$):

$$A' = \{ m \in M | \forall g \in A, (g,m) \in I \}$$
$$B' = \{ g \in G | \forall m \in B, (g,m) \in I \}$$

If $A \subseteq G$, then A' is a set of attributes common to the objects of A. The derivation operator ($'$) can be reapplied in A' resulting in a set of objects again (A''). Intuitively, A'' returns the set of all objects that have in common the attributes of A'; note that $A \subseteq A''$. The operator is similarly defined for the attribute set. If $B \subseteq M$, then B' returns the set of objects that have the attributes of B in common. Thus, B'' returns the set of attributes common to all objects that have the attributes of B in common; consequently, $B \subseteq B''$.

As an example, using Table 1, objects A = $\{Tweet2, Tweet3\}$, when submitted to the operator described above, will result in $A' = \{Bolsonaro, Elections\}$. So $\{\{Tweet2, Tweet3\}, \{Bolsonaro, Elections\}\}$ is a concept. All concepts found from Table 1 are displayed in Table 2.

Table 2: Existing concepts in the formal context of Table 1.

Objects	Attributes
{Tweet 1, Tweet 2, Tweet 3, Tweet 4}	{}
{Tweet 4}	{Fake News}
{Tweet 1, Tweet 3}	{Lula}
{Tweet 2, Tweet 3}	{Bolsonaro, Elections}
{}	{Lula, Bolsonaro, Fake News, Elections}

In Table 2 there is a concept with an empty attribute set and a concept with an empty object set. They are called *infimum* and *supremum*, respectively.

2.3 Triadic Concept Analysis

Initially, Triadic Concept Analysis (TCA) was defined by Lehmann and Wille (Lehmann and Wille, 1995) which extends FCA, but a new dimension was added (Wille, 1995).

Formally, a triadic context is given by the quadruple (K_1,K_2,K_3,Y), where K_1, K_2 and K_3 is defined by the sets and Y the relation of the K_1, K_2 and K_3, i.e., $Y \subseteq K_1$ x K_2 x K_3, the elements of K_1, K_2, and K_3 are called (formal) objects, attributes, and conditions, respectively, and $(g,m,b) \in Y$ is read: the object g has the attribute m under the condition b. An example of a triadic context is represented in Table 3. This example shows the dataset with 3 dimensions: Users, Months and Terms. We have the Users/ID $\{1,2,3,4\}$ as objects, Months $\{July, August\}$ as attributes and Terms $\{Lula, Bolsonaro, Fake News, Elections\}$ as conditions.

Implications are dependencies between elements of a set obtained from a formal context. Given the context (G,M,I) the rules of implication are of the form $B \rightarrow C$ if and only if $B,C \subset M$ and $B' \subset C'$ (Ganter et al., 2005). An implication rule $B \rightarrow C$ is considered valid if and only if every object that has the attributes of B will also have the attributes of C.

We can define rules, as follows: $r : A \rightarrow B(s,c)$, where $A,B \subseteq M$ and $A \cap B = \varnothing$. We can also define the support of the rules, which is defined by $s = supp(r) = \frac{|A' \cap B'|}{|G|}$ and the confidence of the rules, which is defined by $c = conf(r) = \frac{|A' \cap B'|}{|A'|}$(Agrawal and Srikant, 1994).

Table 4 shows two existing rules in the context of Table 1. The rule Bolsonaro $\rightarrow$ Elections has 50% support because this rule happens in 2 tweets, out of a total of 4 tweets. The confidence is 100%, since whenever a tweet has Bolsonaro it also has Elections. When a rule has 100% confidence, such as the rule Bolsonaro $\rightarrow$ Elections, it is called an implication.

2.4 Database Processing

Textual databases need to be pre-processed before being analyzed. The steps performed in this work are the following: N-Gram, stop word removal, and Regular Expression.

- N-Gram: is a contiguous sequence of n items from a given sample of text. The items can be letters or words that are in sequence on a text sample;
- Stop Word Removal: consists of removing words such as articles and prepositions, as these words are not significant for textual analysis;
- Regular Expression: a technique to determine a pattern in a text sample. It is used to find a group of words that need to be replaced or deleted.

The steps described above were applied through the Python package Natural Language Toolkit (NLTK). The NLTK package has a list of stop words, such as “the”, “a”, “an”, and “in”, so those words in the list are removed from the database being preprocessed, as these words are not meaningful to the analysis. This allows the database after pre-processing to have a reduced size and also reduces the analysis time (Contreras et al., 2018).

An n-gram is a contiguous sequence of n items from a given sample of text. The items can be letters or words that are in sequence in a text sample. An n-gram of size 1 is referred to as a unigram and does not consider other words that are in sequence. Size 2 is a bigram, and size 3 is a trigram, meaning that a group of three words are in sequence in a text sample (Roark et al., 2007). Table 5 shows an example of bigrams and trigrams found in a text sample.

3 LITERATURE REVIEW

Several works are relevant to the context of this study. These include works related to topic detection in social networks, topic evolution, and the classification of textual content. These works will be described in the following paragraphs.

Zhang et al. (Zhang et al., 2012) detail how the detection of topics on the Internet is a challenge because the information produced on the Internet is succinct and does not adequately describe the real context being addressed. To solve this characteristic of the information produced on the Internet, the authors used the technique of pseudo-relevance feedback, which consists of adding information to the data being analyzed.

With this strategy, the authors were able to enhance the quality of information available on the Internet, refining the context with which this information is associated. Consequently, they identified the trends within this information that are likely to become more prevalent on the Internet in the future. This research also aims to detect topics within the content produced on Twitter. However, the pseudo-relevance feedback technique was not applied, as the authors specifically analyzed tweets related to one context—the Brazilian presidential election.

Cataldi et al. (Cataldi et al., 2010) used the topic detection technique to identify emerging topics in the Twitter community. The authors were able to carry out the identification considering that if the topic occurs frequently in the present and was rare in the past, and thus characterized them as emerging. To enhance the strategy addressed, an analysis of the authors of these emerging topics was carried out through the Page Rank algorithm, to ensure that the emerging topic is not present only in some bubble of the Twitter community. Finally, a graph was created that connects the emerging topic with other topics that are related to it, and that therefore have a greater chance of becoming emerging topics as well. Unlike the work described above, this research aims to use topic detection to analyze how these topics change over time.

Our results align with findings from previous studies, such as Cataldi et al. (Cataldi et al., 2010) on topic detection in social media, but provide a unique contribution by focusing on temporal evolution. In contrast to studies like Zhang et al. (Zhang et al., 2012),

Table 3: Triadic formal context sample.

	July				August			
ID	Lula	Bolsonaro	Fake News	Elections	Lula	Bolsonaro	Fake News	Elections
1	X	X					X	
2			X	X	X			
3	X						X	
4	X		X			X		X

Table 4: Example of supported and trusted rules.

Rule	Support	Confidence
Bolsonaro → Elections	50%	100%
Lula → Bolsonaro, Elections	25%	50%

which used pseudo-relevance feedback, our approach focused on the context of the Brazilian presidential election, providing more targeted insights.

Dragoş et al. (SM. et al., 2017) present an approach that investigates the behavior of users of a learning platform using FCA. The log generated by the platform contains information about the actions that each student performs on the platform. Thus, the log allows the identification of the profiles of students.

The use of FCA by Dragoş et al. occurs to consider the instant of time that the actions are performed by the students. It is relevant to profile students to understand whether they are performing actions late, early or on time. Therefore, FCA can be considered as an alternative to study temporal events.

Cigarrán et al. (Cigarrán et al., 2016) used FCA to group tweets according to the topics found. By using FCA, the work still manages to obtain a conceptual grid of the topics found, obtaining a hierarchical view of the topics, which is a differential to other techniques. The proposal was among the best results of the RepLab 2013 forum, proving the effectiveness of FCA for the topic detection challenge.

Arca et al. (Arca et al., 2020) propose an approach to suggest tags (meaningful human-friendly words) for videos that consider hot-trend subjects, ensuring the video receives more visibility by being related to a trending subject. The original tags are inserted manually, and these tags serve as input for the algorithm, which matches them with hot trend subjects. Our proposed method also identifies meaningful words, the difference is that our input is tweets, and then analyzes how these words vary over time.

4 METHODOLOGY

This section presents our methodology to achieve the proposed objectives. For this, the steps presented in the sections below were performed.

4.1 Creation of the Dataset of Tweets About the Presidential Election

To create the dataset, a Python script was used to run daily and collect the most relevant tweets of the day, using the filter provided by the Twitter API. This filter ensures that only tweets with significant reach within the social network are returned. The script collected 3,634 tweets during the period from July 23, 2022, to September 8, 2022. The extended collection period was a differentiator in achieving better results.

The dataset was created using the Twitter API with filters applied to retrieve tweets in Portuguese, specifically from Brazilian users. Only tweets containing keywords Lula, Bolsonaro, and Elections were considered. Additionally, the collection focused on tweets with significant reach, such as those with a minimum of 100 retweets or likes.

The challenges encountered during this process include limitations imposed by the Twitter API, restricting the number of requests that can be made by each user. Currently, these limitations are even greater, as it is not possible to use the Twitter API for free, which hinders the reproducibility of this work.

After that the preprocessing of the database was necessary, transforming the textual content of tweets into a list of words that will be analyzed. To perform the task, the techniques described in the Database Processing section were used. Therefore, the final result consists of a database that includes a tweet identifier, the topic extracted from the tweet, and the timestamp indicating when the tweet was published.

4.2 Reducing Formal Context

After the preprocessing stage, it is necessary to reduce the formal context to eliminate topics that are not significant because they are associated with only a few tweets. Thus, in addition to removing attributes from the formal context, this stage helps reduce the number of concepts found, facilitating the final analysis of the work, which is the validation of the concepts found.

To select topics that are present in the largest possible number of tweets, a Python script was utilized. This script counted how many tweets each term was present in and sorted these topics, listing first those

Table 5: Example of bigram and trigram.

Text Sample	Bigram	Trigram
Topics change over time	{Topics change} {change over} {over time}	{Topics change over} {change over time}

with the highest number of appearances. The topics with the highest number of appearances in tweets were then chosen to compose the formal context.

4.3 Lattice Miner

With the topics already defined, it is necessary to prepare the files that will be used as input for the Lattice Miner tool. This tool which was developed at the University of Québec (Missaoui and Kwuida, 2011) will be used for constructing, visualizing, and manipulating contexts. The tool reads files in JSON format to generate the formal context. To accomplish this, a Python script was created to generate these files in the expected format for the tool.

Once the formal context is constructed, it becomes possible to extract implication rules for analysis, thus producing the expected results. The implication rules are provided in XML files and were analyzed manually, as few rules were generated due to the low support they showed. In practical terms, a search was conducted for rules that exemplify events or sentiments of users towards a politician.

5 RESULTS

To analyze the presidential election, the tweets were collected between the period of 07/23/2022 and 09/08/2022. The tweets were obtained through the Twitter API using the keywords Lula, Bolsonaro, and Elections. The Twitter API filter was applied to retrieve only popular tweets, which reduced the total number of tweets. However, these tweets generated significant engagement on the social media platform.

After applying NLP techniques and selecting N-Grams with meaning, the following attributes were obtained for the formal concept: Alexandre de Moraes, good versus evil, elections, Bolsonaro, Lula, democracy, electoral research, Senator Rogério Carvalho, secret budget, Bolsonaro no Flow, armed forces, fake news, first round, Guilherme de Pádua, genocidal, former convict, Jornal Nacional, corruption, vote, president, interview, cash, purchased properties, nursing salary floor, Fachin, and suspends arms decree.

The condition of the formal context created is the period in which the tweet was published. The period used was 4 days, resulting in a total of 12 conditions. As an example, Table 6 displays a sample of the generated formal context.

With the obtained formal context, it was possible to generate implication rules and analyze how the generated rules reflect the events of the presidential election in Brazil. Table 7 provides details on the generated rules.

The first implication rule to be discussed in this study is composed of the terms “Elections” and “Fake news”. This rule was observed during a crucial period that coincided with the announcement of potential penalties for candidates spreading fake news during the electoral period, between 08/08 and 08/15. During this time, there was also a significant increase in the scrutiny of the effectiveness of social media in detecting and removing fake news, aiming to prevent users from being misinformed. It is believed that these events were reflected in Twitter discussions, justifying their identification in our study.

Furthermore, it is important to highlight that the spread of fake news can have serious consequences for democracy, such as interference in the electoral process and manipulation of public opinion. For this reason, the fight against fake news has become a global concern, and the analysis of such data can contribute to the understanding and combat of this growing phenomenon.

Analyzing the confidence metric among rules resulting in fake news with the antecedents “Elections” and “Bolsonaro”, it is possible to observe an interesting trend. It is noted that the rule with the antecedent “Bolsonaro” has higher confidence compared to the rule with the antecedent “Elections”. This difference in confidence suggests that the name of the presidential candidate, Jair Bolsonaro, was more strongly associated with fake news than the electoral context itself. This highlights that social media users had a distinctive perception that candidate Bolsonaro was involved in the dissemination of misleading information. The association between the name of a political candidate and the spread of fake news may have influenced public perception and sparked heated debates and discussions on social media during the election period.

In Table 7, the second identified implication rule

Table 6: Formal context sample.

	07-23-2022 - 07-26-2022				**07-27-2022 - 07-30-2022**			
ID	Lula	Bolsonaro	Genocidal	Former convict	Lula	Bolsonaro	Genocidal	Former convict
1	**X**			**X**				
2		**X**	**X**					
3					**X**			**X**
4						**X**	**X**	

Table 7: Implication rules that varied over time.

Time Period	Antecedent	Consequence	Support	Confidence
08-08-2022 to 08-11-2022	Elections	Fake news	4%	13%
08-12-2022 to 08-15-2022	Elections	Fake news	2%	16%
08-08-2022 to 08-11-2022	Lula	First round	2%	16%
08-16-2022 to 08-19-2022	Lula	First round	8%	26%
08-16-2022 to 08-19-2022	Bolsonaro	Fake news	4%	33%
09-01-2022 to 09-04-2022	Bolsonaro	Fake news	3%	22%
08-20-2022 to 08-23-2022	Lula	Former convict	4%	25%
08-28-2022 to 08-31-2022	Lula	Former convict	2%	14%
08-24-2022 to 08-27-2022	Lula	Interview	2%	25%
08-28-2022 to 08-31-2022	Lula	Interview	2%	7%
08-28-2022 to 08-31-2022	Bolsonaro	Cash, Purchased properties	2%	10%
09-01-2022 to 09-04-2022	Bolsonaro	Cash, Purchased properties	3%	22%

is composed of the terms “Lula” and “First round”. Temporal analysis reveals that this rule occurred in two distinct phases: the first between 08/08 and 08/11 and the second between 08/16 and 08/19. At the time, the possibility of Lula winning the election in the first round was a recurring topic in the media, as electoral polls indicated that he was close to reaching 50% of valid votes.

It is interesting to note how the support for this rule increased significantly, going from 2% to 8%, reflecting the growing support for Lula’s candidacy and the public’s interest in this electoral scenario. It’s worth highlighting that the increase in support is directly related to the evolution of electoral polls, which showed the increasing preference of voters for the former president. On August 18th, a poll indicated that Lula reached 51% of valid votes, further fueling the discussion about the possibility of his victory in the first round.

Therefore, it is possible to observe how the support metric can be useful for evaluating the evolution of discussions on the social network, and tracking the rise of topics that become increasingly relevant to users. This analysis is crucial for understanding the impact of public opinion on elections and the construction of candidates’ images.

The third identified implication rule in Table 7 is composed of the terms “Bolsonaro” and “Fake news”. It is interesting to note that this rule appeared in dispersed periods, first between 08/16 and 08/19 and then between 09/01 and 09/04. The fluctuation in the occurrence frequency of this rule can be explained by two relevant events related to Bolsonaro and fake news.

The first event occurred in mid-August when news broke that a group of businessmen was allegedly orchestrating a coup in favor of President Bolsonaro. This event generated significant media and social media attention, with Bolsonaro dismissing the news as fake. This could have influenced the occurrence of this rule in the first analyzed phase.

The second event that may have driven the occurrence of this rule took place in early September when the Superior Electoral Court (TSE) fined President Bolsonaro for spreading fake news linking Lula to the First Capital Command (PCC). This false news circulated widely on social media, sparking debates and discussions about the spread of misinformation and its impact on elections. This event could have contributed to the occurrence of the rule in the second analyzed phase.

This implication rule illustrates how a particular theme can emerge on the social network, be discussed for a period, and then temporarily disappear, only to resurface at another time. This highlights the volatility of social media and the dynamic nature of discussions that take place within them. Understanding these dynamics is crucial for the analysis of public opinion and electoral campaign strategies.

The fourth identified implication rule in Table 7 is particularly interesting as it reveals the strong political polarization in Brazil. This rule is composed of the terms "Lula" and "Former convict" and was observed shortly after a televised debate among presidential candidates in which Jair Bolsonaro referred to candidate Lula as a "former convict".

This act sparked a heated discussion on social media about whether Lula could be called innocent after events that discredited Operation Car Wash, the largest corruption investigation in the country's history. While Bolsonaro's supporters applauded the provocation, Lula's supporters reacted with indignation, accusing the current president of attempting to tarnish the image of the former president and leader of the Workers' Party.

The implication rule "Lula" and "Former convict" illustrates how social media can be used as tools for the propagation of political discourse and how political disagreements can lead to polarized and heated discussions on the internet.

It is interesting to note that the fourth identified implication rule in Table 7 showed a confidence of 25% between 20 and 23 August, making it the second-highest confidence found in the analysis. This highlights the strong association between the name of the candidate Lula and the term "former convict", suggesting that Jair Bolsonaro's strategy was effective in linking Lula to his past as a convicted individual in the eyes of the law.

This result indicates how political rhetoric can influence discussions on social media and how polarization can lead to the widespread dissemination of biased information. Furthermore, this implication rule emphasizes the importance of sentiment and opinion analysis on social media to understand how political rhetoric can influence public opinion and shape political discourse around certain topics and public figures.

The fifth identified implication rule in Table 7 is composed of the terms "Lula" and "Interview". The latter term refers to the interview conducted by Jornal Nacional of TV Globo with all presidential candidates. It is interesting to note that this interview took place on August 25, and yet the subject continued to be discussed on Twitter until August 31. This fact demonstrates the relevance of the interview for platform users, who engaged in discussions about the candidates' performance in that event.

In the specific case of the term "Lula", it can be inferred that the candidate's participation in the interview generated even greater interest among users who discussed his performance and the ideas presented. This highlights the importance of traditional media in shaping public opinion and how social media platforms can amplify and prolong the impact of these events in society. Additionally, the persistence of the discussion about the interview on Twitter also demonstrates how social media can be used to monitor and analyze public engagement regarding significant political events.

It is notable how the confidence metric of this specific rule significantly decreased in the second analyzed period. This drop in confidence indicates that, after a few days from the mentioned interview, new terms and topics related to candidate Lula started to gain relevance in the discussion. This highlights the volatility of the social network, where the discussed themes and topics can change rapidly within a short period.

This agile dynamic of the social network reflects the ephemeral nature of conversations and the rapid dissemination of information. In a matter of days, other events, statements, or more recent occurrences can capture the users' attention and shift the focus to more recent topics. This observation reinforces the importance of continuously monitoring discussions on social media to gain an updated understanding of the landscape of opinions and topics of interest.

The sixth identified implication rule in Table 7 is very relevant for understanding the Brazilian political context. It is composed of the terms "Bolsonaro", "Cash", and "Purchased properties" and is related to news articles published by the Brazilian media that highlighted many of the properties purchased by the Bolsonaro family were acquired in cash.

This fact raised questions from Jair Bolsonaro's opponents, who sought to understand the reason behind the cash purchases and the origin of this money. The implication of these terms shows how issues related to ethics in politics are relevant to Twitter users, who aim to discuss and comprehend the implications of such actions by politicians.

It is noticeable how the analysis of topics discussed on Twitter can provide valuable information for presidential campaigns. By understanding which topics are generating more interest and discussion among social media users, campaigns can adjust their communication strategies and act more effectively to win over the electorate. Additionally, the analysis can also help campaigns identify weaknesses in their strategies and improve them.

In summary, the analysis of topics on Twitter can be an important tool for electoral campaigns to connect with the electorate and better understand their concerns and interests.

6 CONCLUSIONS

This paper presents an approach to analyze topics discussed on the Twitter social network. To identify these topics, the Twitter API is used to extract tweets, and NLP is applied to find topics within the tweets. Finally, FCA was utilized to generate implication rules between these topics, providing metrics such as support and confidence.

With the generated implication rules and metrics, it was possible to compare how Twitter users were affected by events occurring at the time, such as the Brazilian presidential election.

The Brazilian presidential election was evaluated with a focus on the two most relevant candidates, Lula and Bolsonaro. The analysis was conducted over a month and a half, allowing topics that were being discussed to be forgotten and then remembered by users of the social network, as was the case with the topics "Bolsonaro" and "Fake news".

The results demonstrate that it is possible to assess how users are reacting to events related to the election, information that is relevant for presidential campaigns. Campaigns need metrics to evaluate the impact of their actions.

As future work, there is a plan to automate the entire process so that data can be extracted from the social network and analyzed automatically, generating metrics on the topics discussed at the same moment. This is relevant for analyzing this information at the same time it is being discussed on the social network. Also, could extend this analysis by incorporating sentiment analysis to better understand the emotional tone of discussions. While our current study focuses on topic evolution, adding a sentiment dimension could reveal more nuanced insights into public opinion. Future work could focus on automating the rule analysis process using machine learning techniques, which would allow for a more extensive and unbiased evaluation of the generated rules and improve the overall efficiency of the method.

ACKNOWLEDGEMENTS

The authors thank the Pontifícia Universidade Católica de Minas Gerais – PUC-Minas and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — CAPES (CAPES – Grant PROAP 88887.842889/2023-00 – PUC/MG, Grant PDPG 88887.708960/2022-00 – PUC/MG - Informática, and Finance Code 001). The present work was also carried out with the support of Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) under grant number APQ-01929-22.

REFERENCES

Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB '94, pages 487–499, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Arca, A., Carta, S., Giuliani, A., Stanciu, M., and Reforgiato Recupero, D. (2020). Automated tag enrichment by semantically related trends. In *Automated Tag Enrichment by Semantically Related Trends*, pages 183–193.

Cataldi, M., Di Caro, L., and Schifanella, C. (2010). Emerging topic detection on twitter based on temporal and social terms evaluation. In *Proceedings of the Tenth International Workshop on Multimedia Data Mining*, MDMKDD '10, New York, NY, USA. Association for Computing Machinery.

Cigarrán, J., Ángel Castellanos, and García-Serrano, A. (2016). A step forward for topic detection in twitter: An fca-based approach. *Expert Systems with Applications*, 57:21–36.

Contreras, J. O., Hilles, S., and Abubakar, Z. B. (2018). Automated essay scoring with ontology based on text mining and nltk tools. In *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, pages 1–6.

Ganter, B., Stumme, G., and Wille, R., editors (2005). *Formal Concept Analysis: Foundations and Applications*. Springer.

Ganter, B. and Wille, R. (1999). *Formal concept analysis: mathematical foundations*. Springer, Berlin; New York.

Lehmann, F. and Wille, R. (1995). A triadic approach to formal concept analysis. *Conceptual structures: applications, implementation and theory*, pages 32–43.

Missaoui, R. and Kwuida, L. (2011). Mining triadic association rules from ternary relations. *Formal Concept Analysis*, pages 204–218.

Roark, B., Saraclar, M., and Collins, M. (2007). Discriminative n-gram language modeling. *Computer Speech Language*, 21(2):373–392.

SM., D., C., S., and Şotropa DF. (2017). An investigation of user behavior in educational platforms using temporal concept analysis. In *ICFCA 2017*.

Wille, R. (1995). The basic theorem of triadic concept analysis. *Order*, pages 149–158.

Zhang, J., Liu, D., Ong, K.-L., Li, Z., and Li, M. (2012). Detecting topic labels for tweets by matching features from pseudo-relevance feedback. In *Proceedings of the Tenth Australasian Data Mining Conference - Volume 134*, AusDM '12, page 9–19, AUS. Australian Computer Society, Inc.