NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Extraction and Exploration of Morality on Social Networks

Analyzing Twitter discussions with the Moral Foundations Theory

Naomi Ferreras Custódio

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

EXTRACTION AND EXPLORATION OF MORALITY ON SOCIAL NETWORKS

Analysing Twitter discussions with the Moral Foundations Theory

by

Naomi Ferreras Custódio

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Data Science

Supervised by

Professor Doutor Flávio Luís Portas Pinheiro, NOVA IMS

February, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Venda do Pinheiro, February 21st, 2023

DEDICATION

In loving memory of Maria José Valentim, my grandmother and best friend, who passed during my journey of completing this master's thesis. Without her, this would never have been possible.

ACKNOWLEDGEMENTS

I extend my heartfelt gratitude to my family, especially my mother, Laura, Gonçalo, Marta and Luís, for being my pillars of support throughout my entire journey, and particularly in these last crucial months. A major thanks to Patrik Pop, Ana Catarina Ribeiro, and Pedro Ribeiro for their amazing help with reviewing, troubleshooting, and keeping me motivated.

To my grandmother, one of my biggest supporters since the beginning of times. Thank you for always making sure I kept on track and fulfilled my promises and goals. You were the boost I always needed. Wherever you are, I am sure you are proud of me.

To my best friends, Ângela, Marta and Mónica who were always so supportive of my journey.

A sincere note of appreciation goes to Microsoft, my team specifically. Thank you for your motivation, flexibility, interest, and always being ready to assist if needed. Sometimes, a helping hand is all we need.

To IMS, an institution that I can now call home, a simple thank you does not suffice. I am grateful for the lessons, memories, accomplishments, experiences, and the personal, academic, and professional growth that I have gained. May IMS continue to nurture creative and motivated students each with a dream - much like myself.

Last but certainly not least, I want to express my profound gratitude to my supervisor and professor, Flávio Pinheiro. Thank you for the opportunity to delve into such a fascinating and cutting-edge topic, and for your consistent availability to guide me throughout the master's program.

"Não foi por acaso que o meu sangue veio do sul
se cruzou com o meu sangue que veio do norte
não foi por acaso que o meu sangue que veio do oriente
encontrou o meu sangue que estava no ocidente
não foi por acaso nada do que hoje sou
desde há muitos séculos se sabia
que eu havia de ser aquele onde se juntariam todos os sangues da terra
e por isso me estimaram através da História
ansiosos por este meu resultado que até hoje foi sempre futuro."

- Almada Negreiros, in Rosa dos Ventos

(Artist, Poet)

ABSTRACT

In the last decade, social media platforms, particularly Twitter, have played a major role in shaping public discourse and influencing real-world events, ranging from political debates to social movements. The insights of the data generated on these platforms provides a unique lens through which we can analyze and understand society's trends, making it crucial to process this information effectively. Twitter has not only enhanced people's voices, thoughts, and sentiments but has also facilitated public discussions and information dissemination through features like tweeting, replying, and retweeting. This research aims to leverage social media data, specifically Twitter data, to uncover the underlying morality embedded in these discussions. The Moral Foundations Theory serves as an explanatory factor for variations in moral behavior, shedding light on how individuals express opinions and categorize thoughts using both individualizing and binding foundations. Understanding the interplay between morality, viral topics and user groups is of extreme importance, especially in explaining political ideologies and cross-cultural differences in moral behavior. By employing advanced natural language processing for extracting moral insights from Tweets related to viral topics, this study seeks to unravel the intricate dynamics that shape topics' lifetimes and uncover the role of morality in justifying various perspectives. The investigation draws on data obtained from the Twitter Search API, specifically focusing on the year 2020, a year in which humanity faced many challenges. Anticipated outcomes of this study include the development of a robust approach/methodology capable of providing insights and predictions on diverse challenges, such as discerning the presence of morality in specific topics. This research contributes to the evolving landscape of social science studies by addressing ambiguous questions arising from the intersection of social media, morality, and topic virality.

KEYWORDS

Moral Foundations Theory; Twitter; Morality; Social Media; Political Annotation; Topic

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

API	Application Programming Interface
COVID-19	Coronavirus disease of 2019
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
EDA	Exploratory Data Analysis
LDA	Latent Dirichlet Allocation
MFD	Moral Foundations Dictionary
MFT	Moral Foundations Theory
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NMI	Normalized Mutual Information
URL	Uniform Resource Locator
US	United States
USA	United States of America

1. INTRODUCTION

Over the past decade, social networks have evolved beyond mere virtual connections, becoming influential platforms that shape public discourse, political narratives, and societal attitudes. In early 2021, 23% of US adults were active on Twitter (Dinesh, S. et al, 2023). Platforms such as Facebook, Snapchat, Instagram, and Twitter itself, are nowadays in the root of topic emergence and decay, as the majority of humanity is virtually present in these networks. These platforms can, therefore, have a tremendous impact on real-life events such as political discussions like the US Presential Elections, social movements as Black Live Matters, humanitarian and health crisis like the covid-19 pandemic, environmental catastrophes, among others.

The Capitol riot, which happened on the 6th of January 2021, in Washington DC and in which 5 people lost their lives and many were injured, is an accurate representation of the impact of social media on real-life events ("Capitol Riots Timeline," 2021). It started with a tweet from former President Donald Trump stating a vote fraud during the elections period as well as sounding people for a protest. These tweets were considered the beginning of the movement "Save America March", which had a tremendous impact in US politics, one of the largest political forces worldwide. Many hashtags referred to this sequence of events such as #CapitolRiot, #January6, #CapitolAttack, #StopTheSteal, #SaveAmerica, and #DCProtests.

When considering the impact that social networks can have on society, beyond virtually connecting humans and acting as a commercial platform, understanding better user's speech becomes crucial. Individuals' voices, thoughts, and sentiments are amplified and collected daily by social platforms, and therefore they become very useful when targeting user types or ideas on certain subjects. As mentioned in the work of Geller et al. (2023), it is relevant to understand the language used within a topic, their efficiency in carrying information, and the resulting perception actors have of speech. To have a greater explanation of user perspectives on certain topics, it is essential capturing the sentiment associated to each comment/discussion. Performing Sentiment Analysis, resides on a simplistic positivity and negativity categorization, and therefore lacks a broader understanding of the discussion. Other speech knowledge dimensions are required to understand better the nuances of user messages, enriching our comprehension of the overall meaning of a text.

In this study, Morality is considered as a mechanism for characterizing speech and helping enhance the sentiment associated to each discussion. Morality relates to the standards of good or bad behavior, such as fairness and honesty, that each person believes in and follows as universal laws. This includes behaving in ways considered by most people to be correct and honest, whilst being able to identify morally righteous actions in society. Delineating morality is a daunting task in ethics, as such it is not trivial to extract the morality associated to a particular text. In an effort to accomplish this, within this project morality embedded in text

will be assessed through the Moral Foundations Theory as a classification technique of Twitter data.

In the 2023 study from Zúquete et al., they state that human morality stands atop five different foundational dimensions: Care (it makes us sensitive to signs of suffering and need); Fairness (it makes us sensitive to indications that another person is likely to be a good (or bad) partner for collaboration and reciprocal altruism); Loyalty (it makes us sensitive to signs that another person is (or is not) a team player); Authority (it makes us sensitive to signs of rank or status, and to signs that other people are (or are not) behaving properly, given their position); and Sanctity (it makes it possible for people to invest in objects with irrational and extreme positive and negative values, which are essential for binding groups together).

The MFT itself has been the subject of both praise and criticism. It is the most well-established theory in psychology and broadly adopted in the computational social science field since it defines a clear taxonomy of values together with a term dictionary, the Moral Foundations Dictionary (Kevin B. et al., 2016). The creators of the MFD, highlight the difficulty of creating such a resource since linguistic, cultural, and historical context reflect on language usage (Araque et al., 2020). The studies related to Natural Language Processing and MFT pave the way for further advancements in moral text analysis, and in this work, they are considered under two dimensions:

- Linguistics- exploring how to characterize speech and enhance morality on speech from textual data.
- Social Sciences- detecting the morals narrative which may significantly improve our understanding of the peoples' dispositions in, for instance, controversial social phenomena, as well as the evolution of opinions over time.

While it has provided a valuable framework for research into the cultural and psychological underpinnings of morality, it also faces several challenges. The theory has potential for cultural bias moral values as the emphasis on morality assessment might not be replicable or generalized across all cultures and ethnicities (Davis, D. E., et al, 2016). Oversimplification of morality, as it is assumed in this theory, reduces complex moral reasoning to a small set of foundational principles. Thus, potentially oversimplifying the complexity of moral decision-making, and which might not fully capture the richness and nuance of moral thought. Furthermore, political polarization is possible, since the application of the theory brings concerns for it being used as a weapon to reinforce political divisions, or unfairly stigmatize certain moral foundations. There is a possibility for reduced predictive power when ascertaining individual behavior, as it may not consistently account for variations in moral judgments and actions across different contexts. Also, the theory can have limited scope since it cannot encompass the full range of human moral experience. It focuses on a specific set of foundations (Care, Fairness, Loyalty, Authority, and Sanctity), potentially neglecting other important moral dimensions. Other limitations, such as the calculation of the scores associated to each word/dimension, or the evolutionary aspect on the speech/moral

association, also concern scientists that seek to use this approach for describing moral behavior and speech.

When looking into the possible results and applications of this study, a significant example could be to use this methodology in a political campaign. In such a context, understanding user reactions, analyzing morality and general sentiments, generates valuable insights for political parties since a political campaign can have an immediate impact on each person's opinion and thought. E.g. after a debate if a party's team needs to understand how people felt about the discussions, they can check the moral association to the topics approached, mapping to relevant political keywords, and thus decide on the next move being taken to increase their vote probability.

In order to conduct this study, iterative exploratory data analysis and preprocessing tasks were conducted, as well as polar or neutral sentiment extraction and multiple applications of unsupervised learning. This work uses as case study political discussions to assess the impact of the investigation and verifies the possibility of reproducing this analysis on a different language set. These techniques guide the research in gathering conclusions regarding topic virality and the morality associated to them in the context of social media.

In the following sections of this paper, we will be reviewing the implementation of a topic modelling algorithm, and a moral score of each data entry. This includes a tentative solution for preparing the data prior to extracting the topics and computing morality, together with an in-depth analysis of the insights generated from each cluster and topic. The obtained results are utilized for a discussion regarding the potential future directions for this research.

2. LITERATURE REVIEW

2.1. TWITTER API DATA

In recent years, the proliferation of social media platforms has transformed the landscape of information exchange and communication. Twitter has emerged as a significant source of real-time data, offering a wealth of information about diverse topics, user interactions, and global events. Researchers have recognized the potential of Twitter data in gaining valuable insights into public opinions, trends, and sentiments. This literature review discriminates the nuances involved in conducting exploratory data analysis on Twitter data.

Kumar, S. et al. (2014), approaches an in-depth construction of a crawler that collects Twitter data in real time and explains how to analyze this data to extract important time periods, users, and topic information. In addition to that, diverse visual analysis schemes, encompassing both plots and basemaps, are proposed for conducting exploratory data analysis on Twitter Data. One of the challenges highlighted by their study is regarding Tweets geo-location (Kumar, S. et al., 2014). This attribute is generally absent on overall tweets, meaning that researchers have to use profile information to determine tweet's location as a workaround for any accurate geospatial analysis. Another challenge mentioned with geolocation is the margin percentage of false positives in the veracity of the information, e.g., "CandyCastle" as a geo-filter in user location data. Even though we have a value for the geolocation attribute of this tweet, this entry becomes irrelevant and therefore, further exploration and treatment of the data is required.

There is an emerging importance of encompassing diverse domains and languages for comprehensive Twitter data analysis. Researchers highlight the challenge and importance in addressing distinct language families and ensuring broad coverage of the most spoken languages (Graff, M. et al., 2020), as globalization is also very present in social networks regardless of the user location.

Existing studies have compared the performance of samples obtained from different Twitter APIs, highlighting issues such as potential bias and over-representation of certain user groups. Researchers mention the four main avenues for accessing tweets via APIs (Buskirk, T.D. et al., 2022): Streaming API; Search API; Decahose API; Firehose API. These APIs have different sampling methods, costs, and limitations, which can bring challenges to accurate data analysis, mainly due to lack of random sampling methods, potential biases in existing approaches and coverage error due to differences between Twitter users and non-users. In a 2022 study focused on the distribution of total tweets and keyword incidence related to covid-19 (Buskirk, T.D. et al., 2022), the researchers propose creating a sampling method called Velocity-Based Estimation for Sampling Tweets (VBEST). This algorithm provides a representative random sample of Tweets for estimating frequencies or totals of keywords of interest. This resulted in generally lower biases, than existing in the Twitter Search API methods and an efficient

representation of the full Twitter corpus for estimating frequency or total volume of tweets containing keywords.

2.2. MORAL FOUNDATIONS THEORY

The MFT besides being highly controversial among moral psychologists and other professionals, for placing too much emphasis on the innate nature of moral foundations and not adequately accounting for the role of social learning, cultural transmission, and individual experiences in shaping moral intuitions and values, brings many limitations when using it to categorize text.

The low number of lemmas and stem of words severely impacts the NLP techniques, since prior to working with the data we need to standardize the text structure for consistent formatting. Araque et al. (2020), propose on their study the expansion of their solution to a lexicon of approximately 1000 lemmas obtained as an extension of the MFD based on WordNet synsets, which allowed overcoming this constraint. Other studies also mention an extreme importance, of the expansion of the MFD to make better moral foundations analysis in documents. In the official website of the MFT (*Moral Foundations Theory | Moralfoundations.Org*, 2024), the available dictionary file is very limited, which does not allow yielding proper conclusions from large datasets with a variety of sources and topics, such as the case of Twitter discussions. Therefore, using an expansion version of this dictionary as suggested by previous studies, is essential for the most accurate scoring of morality.

Researchers also highlighted the difficulties in determining the strength and impact of some words since they are rarely used in everyday language and are associated with a moral binary scale, vice and virtue terms (Araque et al., 2020). Words in virtue categories are described as morally righteous actions, whereas words in vice categories typically are associated with moral violations, as observed by figure 2.1 (Hopp et al., 2020).

Figure 2.1. MFD words represented in separate wordclouds. Word size represents the score association to the foundation. Red wordclouds represents vice foundations and blue represents virtue foundations. (Source: Hopp et al., 2020).

To tackle this binary representation of words within the MFT, researchers propose a continuous scoring approach that captures the overall valence of the context in which the word appeared, rather than assigning a word to a binary vice or virtue category. The Valence Aware Dictionary and sentiment Reasoner approach help calculating the valence of each word. For each annotation, a valence score - ranging from -1 (most negative) to +1 (most positive) - was computed, denoting the overall sentiment of the annotation. For each word in the dictionary, they obtained the average sentiment score of the annotations in which the word appeared in a foundation-specific fashion. This process resulted in a vector of five sentiment values per word (one per moral foundation). As such, each entry in the sentiment vector denoted the average sentiment of the annotations in which the word appeared for that foundation (Hopp et al., 2020).

2.3. User Segmentation

The work of Götz et al. (2020) investigates the association between geographic location and personality traits in the United States. The authors explore whether certain personality traits are more prevalent in mountainous regions compared to flatter areas. The study focuses on the 5 distinct personality traits: agreeableness, conscientiousness, extraversion, neuroticism, and openness to experience. This work could justify considering user characterization per region, when assessing morality from social media namely due to the demonstration that personality traits exhibit geographical clustering, with certain traits being more prevalent in specific areas. Researchers also address how sociocultural influences, including local

traditions, customs, and lifestyles, can shape specific social norms in different regions. These norms, in turn, may influence people's moral values and decision-making processes.

The research of Adnan et al. (2014) investigates the social dynamics of Twitter usage in London, Paris, and New York City, focusing on ethnic diversity and gender distribution among users. It analyzes geospatial patterns of tweeting activity, hourly and daily Twitter activity, and the correlation between names and ethnicity. Using data from the Twitter Streaming API, the study reveals insights into the ethnic composition and gender distribution of Twitter users in each city. The findings suggest that Twitter usage varies across cities and among different ethnic groups.

2.4. TOPIC MODELLING

Identifying morality in text can be quite challenging, especially in two main properties: context of text and length of text.

Topic modelling brings many benefits in analyzing discussions and viewpoints on cross societal topics such as Abortion, Homosexuality, Immigration, Religion, and Immorality, as well as by reporting critical social issues like Hurricane Sandy, Baltimore Protest, Black Lives Matter movement and US Presidential Elections.

When trying to understand morality inherent to a text, especially when it comes to Twitter, which can be considered one of the trendiest discussion/debate social platforms due to the nature of the application itself, we facilitate trend and pattern identification when we segment topics, creating context in the text.

Researchers conclude that natural disaster management, such as earthquakes, could be leveraged with social media data (Graff, M. et al., 2020). However, the overwhelming amount of information generated daily on Twitter, as observed in figure 2.2, can hide the relevant events. It is then crucial to determine a strategy for Twitter's data retrieval on major social events or phenomenon.

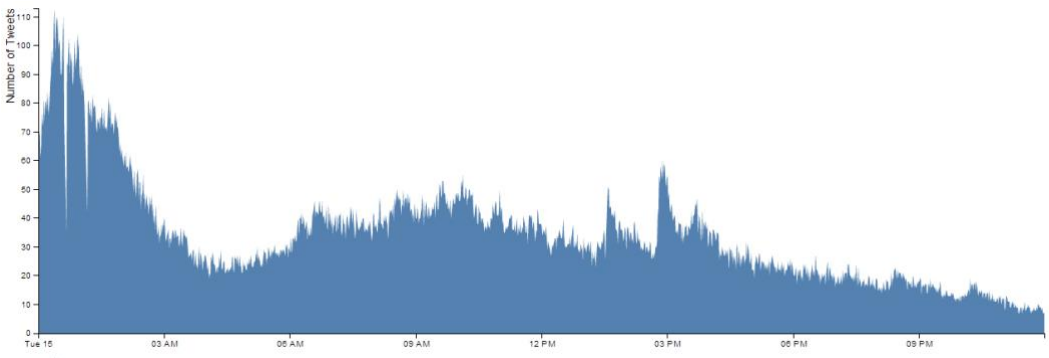

Figure 2.2. Number of Tweets generated per minute on November 15, 2011 (Source: Kumar, S. et al., 2014).

Previous studies propose a method based on tracking users' comments about earthquakes to generate a variant of the Mercalli scale computed with Twitter data (Mendoza et al., 2019).

Authors use counties (municipalities) as the unit of analysis through several features acquired at that level, such as number of tweets, average words, question marks, exclamation marks, number of hashtags, the occurrence of the earthquake word, municipality population, among others.

Twitter text is usually short and often less coherent than other text documents, which makes it challenging to apply NLP algorithms to their API data streams efficiently.

Albanese & Feuerstein (2021) proposes a pooling method to perform topic modelling in Twitter data without changing the underlying structure of a topic decomposition model. This approach uses the LDA algorithm and compares the results of this model with a community pooling technique with other conventional methods, such as Tweet-pooling and Hashtag pooling. The goal of pooling in text mining is to condense information, reduce dimensionality, and capture essential features that can be used for subsequent analysis or modeling. The technique this study proposes addresses the challenge of Twitter document size by aggregating tweets into longer documents based on user communities in the retweet network. This study presents better results on community pooling when compared to other techniques, except from event retrieval, in which it falls behind hashtag pooling results, as seen in figure 2.3.

Scheme	Purity		NMI		Classification		Retrieval		Running time	
	generic	event	generic	event	generic	event	generic	event	generic	event
Unpooled	0.664	0.733	0.436	0.110	0.814	0.843	0.837	0.893	137	388
Author	0.696	0.736	0.374	0.149	0.798	0.859	0.839	0.900	429	926
Hashtag	0.724	0.719	0.383	0.066	0.779	0.762	0.839	0.869	1,737	17,758
Conversation	0.658	0.733	0.436	0.110	0.814	0.843	0.835	0.908	738	1,569
Network-based	0.695	0.736	0.372	0.149	0.798	0.859	0.840	**0.910**	1131	2,841
Community	**0.780**	**0.779**	**0.439**	**0.310**	**0.827**	**0.889**	**0.843**	0.868	141	340

Figure 2.3. Results for different pooling schemes and datasets. Purity measures the fraction of tweets in a cluster having the assigned cluster query label, in which higher scores reflect better clusters representation and better LDA decomposition. NMI measures the cluster quality using information theory in which minimum and maximum values are both 0 when labels and clusters are independent sets, and 1 when cluster results exactly match all labels. Classification represents a supervised machine earning classifying task with naïve bayes classifier and the measurement is in F1 score. Retrieval represents improved topic modeling in Twitter through community pooling 5 document retrieval task and is also represented in F1 score. Running time is represented in seconds and represents the time measured of tweet pooling and LDA topic modelling. (Source: Albanese & Feuerstein, 2021).

2.5. POLITICAL ANNOTATIONS IN DATA

Often when considering morality, we immediately think about controversial topics such as politics, that involve intensive discussions around people's values and repercussions in real life events, such as the aforementioned "Save America March" riot.

The political ideology subject is a topic that historically has been splitting people and classifying them in terms of mindset and beliefs. Often this is a consequence from the place they were born, the education received, and the propaganda consumed. Individuals are often

binarily categorized along the spectrum of political ideologies, from conservative to liberal, left to right and libertarian to authoritarian, shaping not only their perspectives on governance but also influencing their societal roles. As the digital age advances, the intersection of morality and politics has become increasingly prominent, with vast amounts of textual data available for analysis, such as Twitter data itself.

One of the main challenges surrounding political annotation in all types of data in previous studies, was defining the concept of political viewpoints identification. Although liberals preferred individualizing-promoting resolutions (care and fairness) and conservatives preferred binding-promoting resolutions (authority, loyalty, sanctity) there was stronger agreement across political identity in the importance of individualizing concerns, as suggested by figure 2.4 (Gehman et al., 2021).

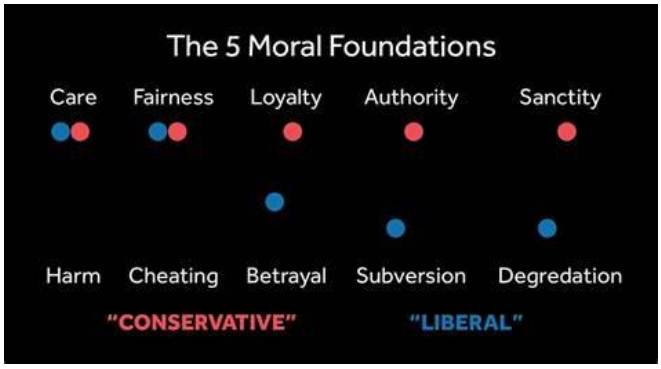

Figure 2.4.The political polar viewpoint association on moral dimensions of vice-virtue spectrum. Blue dots represent the Liberal political ideology association to the foundations, while red represents the Conservative political ideology association to the foundations. (Source: Lambert, 2017).

The MFT suggests and helps define the moral political viewpoint by proving that individuals on the political left draw upon moral intuitions relating primarily to care and fairness, whereas conservatives are more motivated than liberals by authority, ingroup, and purity concerns, even though a mutual speech exists on the individualizing foundations.

The study of Lewis P.G. (2019) concluded that individuals on the political left focus disproportionately on fairness concerns. This study proposes a separation between Trump and other republican candidates (non-trump republicans) since Donald Trump deviated from his republican candidate competitors (and from the Democrats) in making the least use overall of moral-foundations vocabulary. They conclude that the Democratic candidate's speech tended to use more Fairness-Vice words than the non-Trump Republicans (GOP), as observed by figure 2.5. The study also concludes that Democrats' have fewer intensive references to the Authority-Vice foundation. Regarding Trump's speech morality assessment, a high usage of words related to Fairness-Vice rendered him more similar to the Democrats than to his fellow Republicans.

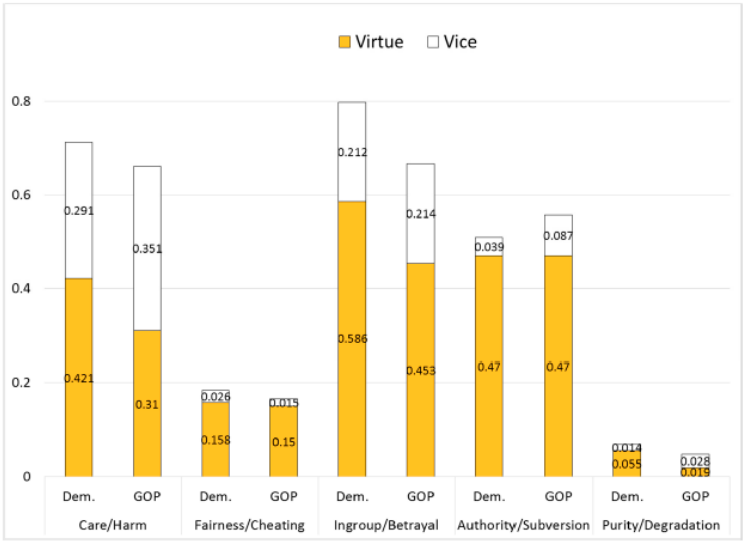

Figure 2.5.Partisan differences in word use (in %) across the five moral foundations. Note: These are simple cross-tabulations without statistical controls. The vertical axis represents the average percentage of words in a segment representing a particular moral (Source: Lewis P.G., 2019).

With content analysis for political viewpoints, Tu My Doan et al. (2022), determines to what extend the data expresses political views, linguistic clues analysis and network analysis for patterns and relationships among common phrases and parties' association. However, their study used the mentioned techniques for topic similarity among their data to segregate on already pre identified political text, from social media platforms, discussion forums, and political speeches.

2.6. Application of the Moral Foundations Theory in Different Language Sets

One of the main limitations from the application of MFT to assess user speech in social networks is the lack of development of the dictionary for other languages, impeding its broader application in cross-cultural studies. Developed primarily for English, its utility has prompted the need for translated versions to facilitate cross-cultural examinations of morality.

As societies become increasingly interconnected, understanding morality in a global context becomes imperative. Matsuo et al. (2019) addresses this gap by proposing the development of a Japanese version of the MFD. These types of developments set the groundwork for future translations of the dictionary into other languages. The study approaches a semi-automated method in developing the Japanese MFD, by providing insights into the challenges and strategies involved in the translation process. The introduction of the Japanese MFD aids on the research of culture-specific moral variations and, enables multicultural comparisons in moral behavior. The significance of this cross-cultural adaptation extends beyond the Japanese context, emphasizing the need for a broader, culturally diverse approach in understanding the universality and cultural specificity of moral foundations.

3. DATA

This research uses four distinct data sources to extract morality from tweets, analyzing in more detail viral topics such as North America politics. In the following pages we describe the source, purpose, properties, and insights of each dataset with the support of visuals from the Exploratory Data Analysis conducted.

3.1. TWITTER API DATA

For the purpose of this research, Twitter was elected as the social media network in study. Twitter data was used to extract the morality from social media discussions since Twitter stands out as a platform often used for debating, getting information, and sharing thoughts.

The data collected from the Twitter API consists of tweets from the state of New York in the year of 2020. The dataset starts on the 17th of October and ends on the 3rd of November, thus comprising 17 days of data. This is represented below on figure 3.1. The day of 31st of October is not present in the dataset, as the file was corrupted. The dataset used is balanced throughout the sample days, despite two days that have a bigger tweet count, 23rd of October and 3rd of November. Standing out due to the release of results from the 2020 United States presidential election.

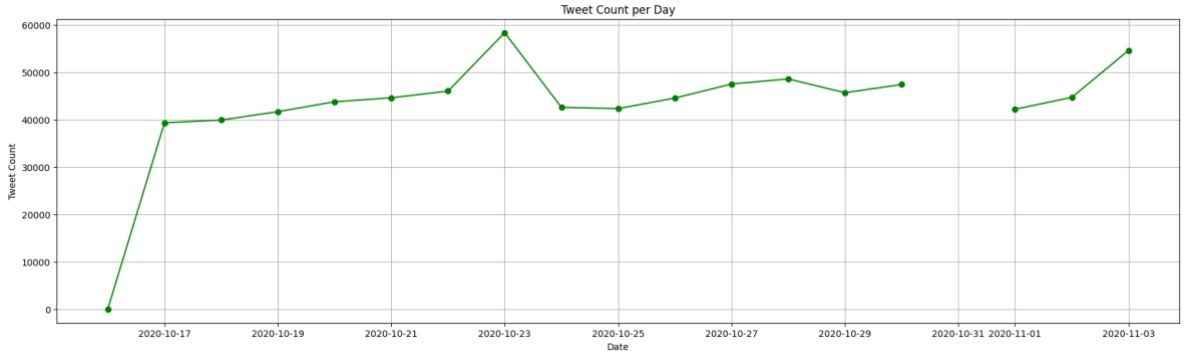

Figure 3.1. Tweet Total Count per Day between 17/10/2020 and 03/11/2020 in the state of New York (excluding 31/10/2020).

In 2020, Twitter had 347.6 million active users, which accounted to approximately 4.46% of the world's population.

The Twitter API can be used to analyze, learn from, and interact with Tweets. It allows interactions with direct messages, users, and other Twitter resources. Twitter's API also gives developers access to all kinds of user profile information, like user searches, block lists, real-time tweets, and more.

The data in the study consists of originally 35 columns, and 960765 entries per column (null and non-null). Each data entry/row consists of a tweet that was posted by a user. In table 3.1

only, the relevant columns used for the research are described, as well as additional columns created throughout the work development.

Table 3.1. Column attributes from the dataset that were relevant for the study.

Column	Meaning
created_at	Timestamp of the post release, in the format of datetime
id	Unique identifier of the tweet entry
user	User information of the twitter account that posted the tweet
geo	Longitude and latitude points, defining a bounding box
lang	Specifies the language the Tweet was written in, as identified by Twitter's machine language detection algorithms
extended_tweet	Includes full_text (which may be more than 140 characters), display_text_range (the offsets in full_text of the body of the tweet), and entities / extended_entities
hash_word	Represents hashtags which have been parsed out of the Tweet text
sentiment_scores	Normalized, weighted composite score calculated by the SentimentIntensityAnalyzer library. It is a single value that represents the overall sentiment of the text
cluster	Clusters that resulted from user segmentation. Integer values that identify to which cluster the row belongs to.
cleaned_tweet	Tweets text after preprocessed, such as removing stop words, lemmatization, removing URLs, punctuation
topic	Categorizes tweets within each of the topics/clusters identified by the LDA model
political_tweet	Indicates if a tweet has political annotations or not.
care.vice	Score of Care foundation with vice connotation.
care.virtue	Score of Care foundation with virtue connotation
fairness.vice	Score of Fairness foundation with vice connotation
fairness.virtue	Score of Fairness foundation with virtue connotation
loyalty.vice	Score of Loyalty foundation with vice connotation
loyalty.virtue	Score of Loyalty foundation with virtue connotation
authority.vice	Score of Authority foundation with vice connotation
authority.virtue	Score of Authority foundation with virtue connotation
sanctity.vice	Score of Sanctity foundation with vice connotation
sanctity.virtue	Score of Sanctity foundation with virtue connotation

Some columns were modified during the preprocessing stages of this research such as grouping the “extended_tweet” column with the “text” column, as they represented the same information, with the exception that the “extended_tweet” held the longer Tweet messages and complete entity metadata. In the preprocessing steps, other columns were also created such as “hash_word”, which contains hashtags extracted from tweets. Table 3.2 illustrates the result from extracting the hashtags to the new column, representing part of the hashtags, with respective relative frequency, across the entire dataset.

Table 3.2. Top 10 Hashtags in the Twitter Dataset descending from relative frequency of occurrence after preprocessing.

Hashtags	Relative Frequency
vote	1
nyc	0,65564
Debates2020	0,522498
BidenHarris2020	0,418328
NewYork	0,360592
Election2020	0,3045
Vote2020	0,254777
VoteEarly	0,253544
traffic	0,240189
COVID19	0,202794

Pre-existent columns suffered some changes, namely the “geo” column that was split by latitude and longitude later in the research for plotting clusters with a basemap.

Regarding the language column, as New York was on the top 10 most diverse states of United States of America, according to Bureau (2024) is home to 3.1 million immigrants which comprise nearly 38% of the city population and 45% of its workforce (*Annual Report - MOIA*, 2024), it is expected that multiple languages are identified in New York Twitter data. English is the language with more tweet counts, 774805 tweets, 80% of the dataset, as well as Spanish with 31354, which corresponds to 3%, as observed on figure 3.2. As for undefined language type (und), they corresponded to 11% of the dataset, however these were not considered for the analysis since no language could be detected. From reviewing these entries, we can observe that they consist of random characters, emojis, tweets containing only URLs, hashtags, or user mentions, and other terms that are not present in the dictionary. For the analysis of morality identification on tweets, the English and Spanish tweets were considered, separately, due to the nature of the selected morality dictionaries.

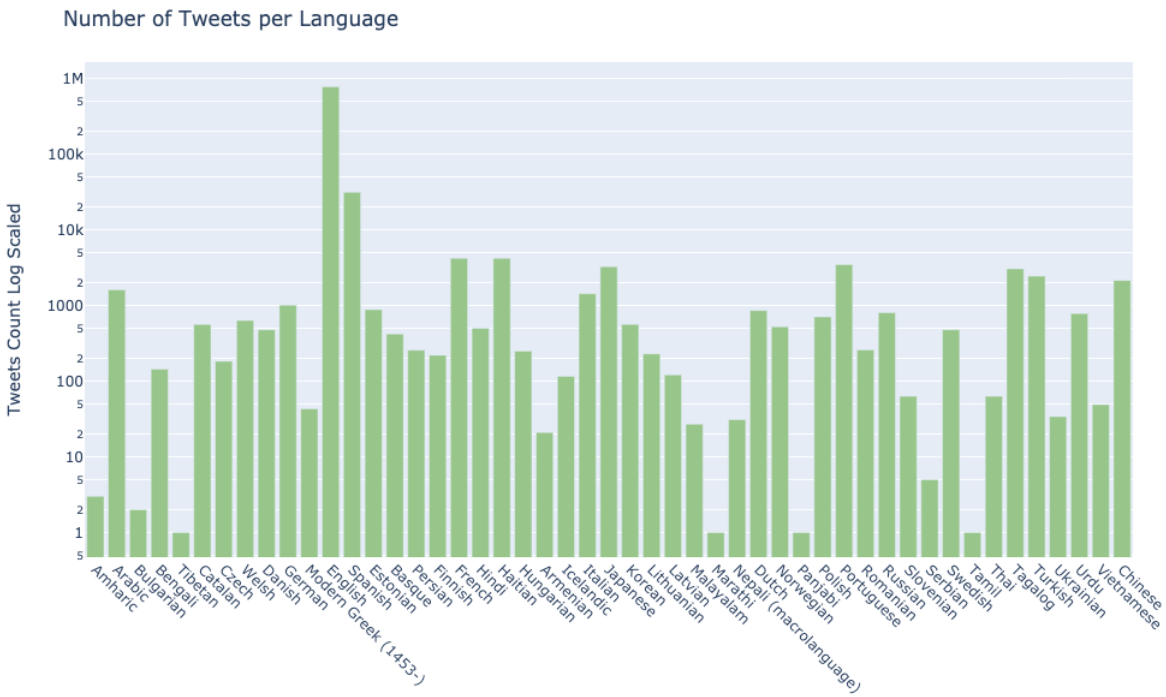

Figure 3.2. Tweet Count per Language Pair (Log Scaled normalization).

Only approximately 10% of the sample Tweets had geolocation. To conduct a spatial analysis, the missing Tweet locations were replaced with the median of available geographic details (latitude and longitude) using the SimpleImputer library.

After our analysis, other important insights could be found in the data namely, the moral foundation scores associated to each tweet, political annotations on tweets and the sentiment of the tweet.

3.2. EXTENDED MORAL FOUNDATIONS DICTIONARY

There are several versions of the MFD available online. For the purpose of this work the dictionary of Hopp et al. (2020), which is available on their GitHub repository, was used. This dictionary contains 3270 words with a score in each moral dimension, on the vice-virtue scale. As observed on table 3.3, words that have a vice score in one of the moral foundations, do not have a virtue score and vice-versa. E.g. the word *outsiders* has a score associated with loyalty and authority vice, however it is not associated with loyalty and authority virtue connotation on these moral dimensions.

Table 3.3. Last 5 rows of the English Moral Foundations Dictionary mapping the dictionary words with the respective moral score in the vice-virtue scale.

	Vice					Virtue				
	Care	**Fairness**	**Loyalty**	**Authority**	**Sanctity**	**Care**	**Fairness**	**Loyalty**	**Authority**	**Sanctity**
outsiders	0,15	0	0,444444	0,375	0	0	0	0	0	0
border	0,123377	0,085	0,063457	0,093023	0,081911	0	0	0	0	0
implement	0,121951	0	0	0	0	0	0,163636	0,190476	0,166667	0,056604
recruit	0	0	0	0	0	0	0	0,066667	0,208333	0,105263
information	0,110312	0	0,087179	0	0,075145	0	0,135	0	0,115839	0

As mentioned previously, there are other versions of the MFD which have a smaller length, such as the one available on the official Moral Foundations website. The decision of using the extended MFD resulted from the complexity and diversity of the corpus used on social media, in which using a broader dictionary may help improve the results of the study.

Sanctity.virtue is the foundation that contains a lower number of words, with stronger affinity to this foundation, having a representation of less than 5% of the total. In contrast, Care.vice is the foundation that contains a greater number of words representing almost 20% of the dictionary. This is expected since the care and fairness foundations are often considered primary and more universally recognized in speech.

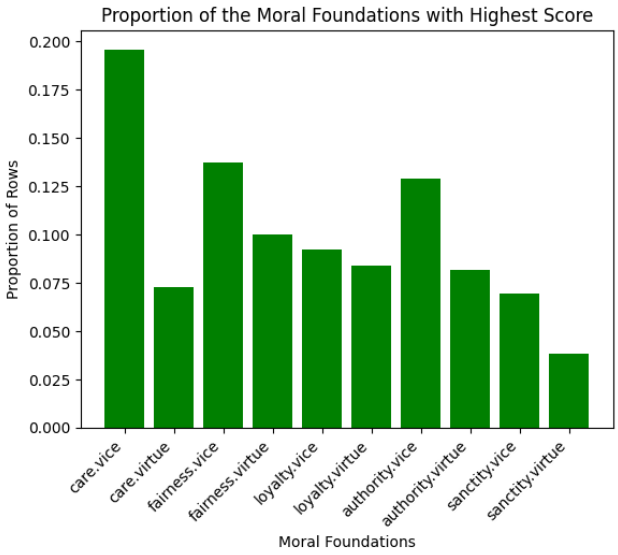

Figure 3.3. Representation of Moral Foundations in the MFD. The y scale represents the foundation up to the value 1 which is the absolute value.

3.3. POLITICAL DATASET

To perform a deeper analysis on tweets morality with political annotation, we require a political word base to help track these comments. In previous works, researchers took political texts to recognize the writer's opinions towards a political matter, however, these texts were already identified as political, and strategically extracted from relevant newspapers, articles, social media posts and debates (Tu My Doan et al., 2022). Initially, by researching several articles, newspapers, and online sources on politics, more specifically US politics/government, we manually created a 171-row dataset that referred to important figures, agents, concepts, events, or any other related terms in politics. E.g. the top 10 words present in the political vocabulary file created: politics; political; politician; governance; government; public; affairs; liberal; conservative; socialist.

After applying the Wordnet library, including lemmas and stem words, as well as synonyms and antonyms of each of the entries manually added, and lowercasing terms, we were able to extend the political dictionary to 1122 words.

When checking Google's search trends in the United States, 2020, the top trends were associated to politics, both on top searches, news, and people, which helped constructing the political vocabulary dataset.

3.4. SPANISH MORAL FOUNDATIONS DICTIONARY

To yield conclusions on the application of this research in other language sets, a Spanish adaptation of the MFD, available online, was used. The dictionary consists of 1373 terms, after removing the legend from the top rows. This dictionary employed a different structure from the selected English MFD. Instead of containing a score per moral foundation, in both vice-virtue scales, it directly categorized each word with the moral foundation of the biggest association, as observed in table 3.4.

Table 3.4. Labelling of Moral Foundations on the Spanish MFD varying from 1 to 10.

Word	ID
Care (virtue)	1
Care (vice)	2
Fairness (virtue)	3
Fairness (vice)	4
Loyalty (virtue)	5
Loyalty (vice)	6
Authority (virtue)	7
Authority (vice)	8
Sanctity (virtue)	9
Sanctity (vice)	10

Table 3.5. Spanish MFD first 5 rows, after label removal, in which MF_1 represents the moral label associated to each Word.

Word	MF_1
áreas de un sitio	5
íntegro	3
íntegro	9
órdenes	7
Corán	9

Other aspects of this dictionary were considered, such as the punctuation in place, duplicated values, and multiple words in the same cell, which brings challenges to the data integrity and accurate analysis of the morality in Spanish tweets. E.g. the word *íntegro* is associated with the moral foundation fairness-virtue.

4. METHODOLOGY

This study employs a comprehensive set of NLP techniques and data visualizations to extract and analyse moral content from textual data. The primary focus of this analysis is to investigate the results of these data-driven methodologies, with a specific emphasis on understanding social phenomena, as exemplified by our case study: USA 2020 Presidential elections. The experimental framework was implemented using Python, a versatile programming language known for its efficiency in handling and analyzing large datasets.

The core objective is to characterize each tweet through topic modelling and associated hashtags as well as investigating possible correlations with each moral foundation. A particular emphasis is placed on the political annotation of tweets, contributing to a nuanced understanding of moral dimensions within political discourse. Furthermore, a user segmentation is applied to group users, enabling a visual exploration of variations in moral foundations among different user groups. An important aspect of the study is the incorporation of temporal information. Leveraging timestamps, we explore the temporal evolution of moral foundations, seeking to discern patterns and changes over time. This temporal analysis is crucial for understanding whether there were significant changes in the moral foundations' presence during specific periods, particularly those coinciding with events related to the elections. Additionally, this study addresses the broader question of reproducibility and generalizability. We extend our analysis to assess whether the observed patterns and results can be replicated for other languages. Given the global nature of Twitter data and the platform's role in facilitating cross-cultural conversations, this evaluation is essential for establishing the robustness and applicability of our findings beyond the specific context of the English-speaking users.

Considering the application of the MFT to Twitter Data throughout the work, our research goals include:

1. Characterizing each Moral Foundation per topic, specifically:
    a. Obtaining a score of each Moral Foundation per tweet.
    b. Identifying topic virality on the dataset and retrieving the moral score of each topic.
    c. Deep diving on the insights generated with the moral foundation's theory on one of the identified viral topics, which in this study is Politics.
    d. Segmenting user groups and analyzing the morality distribution among them. The goal here namely inserts on checking if there are morality trends of specific topics of discussion in certain locations. Can this lead to indications of economic or societal values according to the user location? In this case we analyze the New York state, which varies greatly in terms of quality of life.

2. Comparative temporal analysis on the distribution of morality among Tweets. Is the temporal evolution correlating to the real-world events? This will allow us to prove the assertiveness of social network reflection of reality.
3. Feasibility and challenges of the theory application when considering other language sets.

4.1. EXPLORATORY DATA ANALYSIS: UNVEILING THE DATA LANDSCAPE

To ensure robustness and efficiency in the computational applications, a subset spanning 17 distinct days within a two-week timeframe was meticulously selected for performing the research.

The preliminary phase of this study involved the Exploratory Data Analysis (EDA), a critical component in comprehending the intricacies of the datasets. The EDA encompassed two iterative stages: an initial exploration conducted before implementing NLP and other data cleaning steps, followed by a subsequent analysis.

During this research, the Exploratory Data Analysis (EDA) encompassed a thorough examination of the dataset's content, delving into various columns, formats, and dimensions of each dataset. The analysis included an exploration of the languages identified in the tweets; a feature provided by the Twitter API. A detailed examination of word frequencies shed light on the dataset's most prominent words, offering insights into prevalent themes, potential viral topics, and the identification of common stop words. The utilization of wordclouds and relative frequencies, provided a visual representation of the most recurrent terms, elucidating their significance. In table 4.1, we can observe that 6 out of the 10 most frequent hashtags, are related to the US presidential elections.

Table 4.1. Top 10 Hashtags in the Twitter dataset during the 1st Iteration of the EDA, descending, in relative frequency of occurrence.

Hashtags	Relative Frequency
vote	1
usa	0,590687
job	0,549942
nyc	0,531781
Debates2020	0,52922
BidenHarris2020	0,403725
NewYork	0,2922
Election2020	0,254715
Vote2020	0,24447
VoteEarly	0,237718

To gain an initial understanding of topic frequency, the research employed the Latent Dirichlet Allocation (LDA) topic modeling algorithm on raw data. Geolocation and hashtags frequency were considered crucial, given their optional nature for users and their importance in spatially

distributing discussions and assessing topics derived from hashtags. The SentimentIntensityAnalyzer from the Natural Language Toolkit (NLTK) library was employed to provide a nuanced understanding of user perspectives on the raw data, allowing the analysis of sentiment in tweets. Furthermore, an examination of the MFD involved determining the prevalence of each moral foundation across the dataset. This comprehensive assessment included identifying the dominant moral foundation, indicated by the highest score, and pinpointing the least represented foundation. These analyses collectively contributed to a nuanced understanding of the dataset's composition and balance within the MFD, as observed by figure 3.3 of the Extended moral foundations dictionary section.

In essence, the Exploratory Data Analysis not only provided a comprehensive overview of the dataset's characteristics but also laid the groundwork for subsequent work, ensuring an informed approach to the research questions. These analytical steps were pivotal in shaping the direction of the study, namely the NLP required applications, and establishing a solid foundation for subsequent methodologies.

4.2. DATA CLEANING AND NATURAL LANGUAGE PROCESSING

In order to conduct the primary research, which involved extracting morality from text and analyzing it across various contexts, a rigorous process of dataset cleaning and preprocessing was conducted.

Initially, the dataset was filtered to exclusively include English tweets. The decision to filter by English language consisted in the fact that these represent 80% of the dataset, since their origin is New York, which despite being a multicultural state, is predominant of English speakers. As previously outlined in the Data section of this report, the 'text' column of the dataset did not encompass the complete information for tweets exceeding 140 characters. To address this, the 'extended_tweets' column was transformed, to contain each tweet's complete text, in scenarios in which they are both under and above 140 characters. Although the original format included various attributes, such as 'full_text', 'display_text_range', and 'entities', for the purpose of this analysis, only the 'text' attribute was deemed relevant.

As part of the preprocessing steps, emoticons were removed. On the Sentiment Analysis conducted by Kumar, S. et al. (2014), tweets containing emoticons were considered labeled data, however, on this scenario the removal of emoticons was essential, as they were not conducive when calculating moral foundation scores and performing topic modeling. The Vader package is particularly effective for analyzing social media text due to its awareness of sentiment intensity and polarity, including nuances such as emoticons and, therefore, would also work without performing this step. To specifically focus on hashtags, these were extracted to a new column. Any rows without hashtags were treated as NaN values.

For this study, some attributes were dropped such as ‘id_str', 'coordinates', 'truncated', 'display_text_range', 'timestamp_ms', 'possibly_sensitive', 'contributors', 'quote_count', 'reply_count', 'retweet_count', 'favorite_count', 'extended_entities',

'quoted_status_permalink', 'quoted_status', 'in_reply_to_status_id_str', 'in_reply_to_status_id', 'in_reply_to_user_id', 'in_reply_to_user_id_str', 'in_reply_to_screen_name', 'source', 'place' as they were not considered relevant upon identifying morality from Twitter posts.

Duplicated values across rows were systematically eliminated, employing a meticulous evaluation process to ensure that the decision to drop values was based on genuine duplications across all columns. This cautious approach was crucial, considering that certain values might be repeated legitimately without qualifying as duplicates, due to the architecture of social networks. For instance, a user posting two tweets might result in multiple rows with identical attribute values, yet these entries are not inherently duplicated.

To prepare text data for advanced analysis, NLP techniques helped refining the textual information for input on the subsequent steps. By leveraging the Gensim library along with NLTK, multiple operations were performed in a specific order, namely:

1. Lowercasing and URL Removal- Standardizing the text to lowercase and eliminating URLs to create a uniform text structure.
2. Elimination of User Mentions and Hashtags- Removing words starting with '@' to exclude user mentions and scrubbing hashtags for a more streamlined text.
3. Stopword Removal- Utilizing Gensim's stopword removal function to exclude common, non-informative words.
4. Numerical Data and Punctuation Removal- Clearing the text of numerical data and punctuation marks to maintain focus on the linguistic content.
5. Lemmatization- Applying lemmatization techniques for further text simplification, using a vocabulary and morphological analysis of words, thus removing sometimes the inflectional endings, and returning the base form of a word, lemma.

To avoid incorrect processing of the dataset, these operations were embedded in a custom function, as the order was extremely relevant for a concise work. For example, by changing the order between point 3 and 4, words such as “I’m”, would become “I” and “m” and would therefore, not be considered stopwords. This comprehensive preprocessing pipeline enhances the clarity of our text data, setting the stage for subsequent analysis and ensuring that our insights are derived from meaningful linguistic content.

4.3. User Segmentation and the Association with Moral Foundations

Applying unsupervised learning to discover user groups in the data is the first step taken after preprocessing the dataset.

For clustering the tweet’s data, SimpleInputer was used to complete missing values on the “geo” column, as only a small subset of the dataframe contained the geolocation of each tweet. The median was used to compute this value, as the missing values corresponded to location coordinates.

The first elected clustering algorithm was DBSCAN, which finds arbitrarily shaped clusters based on the density of data points in different regions and separates those regions by areas of low-density so that it can detect outliers between the high-density clusters. To apply this solution, the data selected, namely the “created_at”, and “geo” columns, were standardized. The hyperparameters that determined cluster definition were obtained with a Grid Search with 5 folds. The combination that yielded the best parameters was 'eps': 1.0, 'metric': 'euclidean', 'min_samples': 6, with a score of 0.8920207200199533. This final score resulted from the average of the scores obtained in each of the 5 cross-validation folds, thus ensuring the robustness of the model evaluation by using different train-test splits. The following combination was considered:

- 'eps': [0.1, 1.0, 10]
- ‘min_samples': [2, 10]
- 'metric': ['euclidean', 'manhattan']

Even though, DBSCAN is often better than K-means when it comes to working with oddly shaped data, K-means was elected for clustering. This algorithm brings multiple benefits namely with regards to the speed of the computation of clusters, even on larger datasets. K-means’ goal is to make the data within each cluster as similar as possible while making each cluster as different from the other clusters. In technical terms, it minimizes the intra-cluster distances while maximizing the inter-cluster differences (Garza, A. 2021). This work proposes a user segmentation, which splits users per location and timestamp, thus comprising a low-dimensional space. By exploring the cluster centers, we can understand the typical characteristics of each cluster which is very insightful to understand the average user. Prior to performing clustering and assigning cluster labels to the data points, K-means has a particularity which involves deciding on the appropriate number of clusters (k) for the data, however, when using only two attributes it can become more challenging. To tackle this challenge, the usage of the elbow method was proposed to identify the optimal number of clusters. To assess the optimal number of clusters, the inertia is analyzed with a line plot, which points to different numbers of clusters and looks for the "elbow" point where the inertia starts to level off. In figure 4.1, three clusters are indicated as the optimal value for the segmentation, when the inertia starts to decrease at a slower rate.

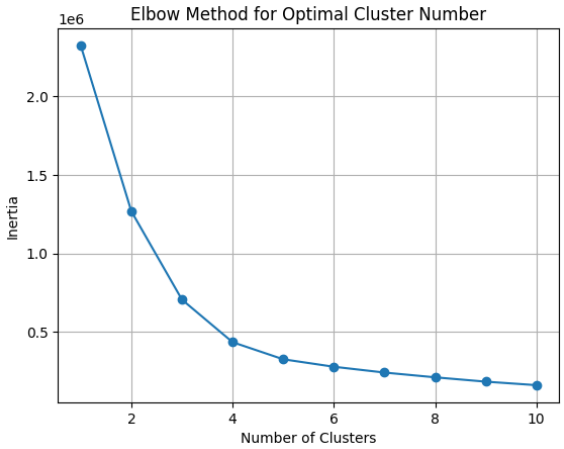

Figure 4.1. Elbow Method for Optimal Cluster Number of User Clustering.

To gain insights into the spatial distribution and characteristics of user clusters, three distinct visualization approaches were employed, each providing a unique perspective on the clustered data:

- 2-Dimensional space- In this representation, the geographical distribution of clusters is visualized in a 2-dimensional space. The plot utilizes a basemap of New York, overlaid with markers indicating the locations of users belonging to each cluster. Cluster distinctions are denoted by the red, green, and blue colors, allowing for a clear differentiation of user groups.
- Interactive 3-Dimensional space- An interactive 3-dimensional space was created to offer a more dynamic exploration of the clusters. The visualization allows for user interaction, enabling a multidimensional understanding of the spatial arrangement of clusters. This approach enhances the ability to discern patterns and relationships between variables, providing a richer interpretation of the clustered users.
- Interactive Basemap- Utilizing geospatial visualization, an interactive basemap, as seen in figure 4.2, was generated to present a comprehensive overview of user clusters. The map is centered on the median coordinates of the dataset, and clusters are represented by distinctive markers. It allows for an interactive exploration of the geographical distribution of clusters, gaining a deeper understanding of their prevalence in different and specific regions.

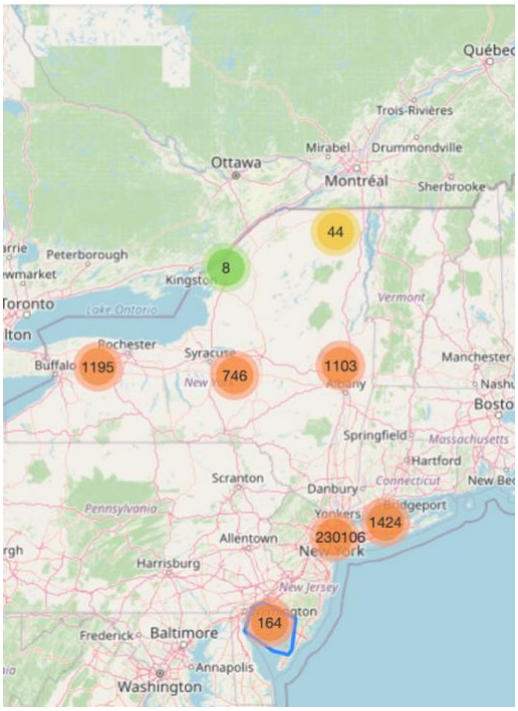

Figure 4.2. Snippet of the Interactive Basemap of user clusters.

These visualization methods collectively contribute to a nuanced interpretation of the segmentation results, facilitating a comprehensive analysis of user spatial distribution and cluster characteristics.

4.4. Topic Modelling with Latent Dirichlet Allocation

In order to identify viral topics within the Twitter dataset, the Latent Dirichlet Allocation, a widely employed topic modeling algorithm, was applied.

To identify the optimal parameters for the model, namely the number of topics, a grid search approach was employed. The primary aim was to explore different combinations of num_topics (number of latent topics) and passes (number of passes over the dataset) while calculating coherence scores to evaluate the quality of the generated topics. As this research comprised only 2 weeks of data, a lower stack of topics was considered for grid search:

- num_topics: [3, 4, 5]
- passes: [10, 20, 30]

The coherence score was computed for each model using the coherence measure, providing a quantitative assessment of the interpretability and coherence of the topics. The model with the highest coherence score was considered the optimal LDA model. The selected parameters were 5 topics and 10 passes with a coherence score of 0.5607554492082579. This grid search process aimed to find the most suitable hyperparemeters for the LDA model, ensuring coherent topic modeling results.

LDA is a probabilistic generative model designed for revealing latent topics within a corpus of documents. LDA operates by assigning probabilities of topics to each document and probabilities of words to each topic. These probabilities serve as indicators of the relevance or significance of different topics within each document. One of the key advantages of LDA is its ability to uncover latent topics without the need for pre-labeling each tweet as 'viral' or 'non-viral.' This characteristic aligns with the exploratory nature of the study, allowing for the discovery of emerging topics without prior assumptions about their nature or categorization. With pyLDAvis, we are able to see the top words for each topic which represent the most important terms for each discovered topic, visually and interactively. Additionally, each topic is associated with a dataframe entry in a new column named “Topic”. The application of the LDA on the preprocessed dataset requires initially the creation of a Gensim Dictionary ‘id2word’ to map unique words into numerical IDs. This step is essential for representing the vocabulary of the dataset in a numerical format, and afterwards the tokenized data is converted into a bag-of-words representation. The corpus variable stores this representation, where each document is represented as a list of tuples, point to the ID and the frequency of its presence. Finally, LDA is used to discover topics.

4.5. Calculation Scores of Moral Foundations per Tweet

To calculate the Moral Foundation score for each tweet, a simplistic approach was conducted by associating each word in the 'cleaned_tweet' column with scores from the MFD. The dictionary contains, for each word, moral scores for the five dimensions and vice/virtue categorization. Initially, the tweets are transformed into lists of space-separated words. Subsequently, a dictionary is initialized to accumulate the sum of scores for each moral dimension. A loop is then executed through each row in the dataset where for every word in the tweet list its presence in the MFD is verified. If the word is found, the corresponding row in the MFD is retrieved, and scores for each dimension are added to the evolving scores dictionary. Finally, the accumulated scores for each moral dimension are appended to the original dataset as new columns named after each moral foundation dimension. This methodological approach allows for a quantitative examination of the association of each tweet with each moral foundation, contributing to a greater understanding of the moral landscape present in the dataset. The initial approach considered for the calculation of the moral foundations’ score was the eMFDscore library developed from the work of Hopp et al. (2021). This would allow a concise extraction of the moral information from text data with the use of multiple versions of the MFD. However, due to the recent nature of this library, during our research process several errors were encountered in the scoring script which did not allow proceeding with this solution for the computation of moral valences. The procedure used was instead, as mentioned, calculating the scores associated to each foundation/word present in the corpus.

4.6. Moral Foundations Association per Viral Topic and User Clusters

For the analysis of the association of moral foundations with viral topics and user clusters, multiple data visualization techniques were applied, including grouped bar charts of moral foundation scores per topic in median and relative frequency measurements. The relative frequency measurement was used to compare tweets moral foundation association more efficiently among them, and later to compare these values with the Spanish tweet's analysis.

For the viral topics moral foundations analysis, we used the counts of occurrences for tweets that were previously grouped by topic. This visualization was normalized by dividing each topic moral foundation by the sum of counts for all moral foundations for that topic, resulting in the relatively frequency.

4.7. Temporal Analysis of Moral Foundations

With the temporal analysis we intended to explore the distribution of moral foundations on tweets over time. This analysis consisted mainly in the deployment of different visualizations over each moral foundation to potentially identify any patterns or trends, and investigate variations across the different clusters, throughout time.

For the plots, only a subset of columns was considered, as we did not need id, user, or location details for time analysis. The 'created_at' column was essential for assessing temporal distribution, as well as the moral foundation columns calculated previously. A line plot was used for plotting temporal distributions, also representing relative frequency.

4.8. Case Study: Political Discussions on Twitter

The decision to conduct the case study of political discussions on Twitter Data, consisted in the fact that US presidential elections were taking place at the time of the collected tweets, and this event was identified amongst the most common terms on the corpus from applying the topic modelling algorithm. A data source with annotations for deeper identification of these discussions was required. In the Data section of this report, the approach of the wordnet library is described. Araque et al. (2020), proposes a moral lexicon, adapted to their use case, which helps maximizing their dataset since it expanded to many more lemmas and less radical and more frequently used terms. This is also required for identifying political annotation data, since there are no available datasets that have this information. Applying this approach helped expanding the political dataset that was manually created by researching papers, news, and articles from the Internet on US politics. The developed library helped creating more words such as antonyms, synonyms, lemmas, and stems from the original terms present in the dataset, which consequently helped capturing political annotations in the Twitter dataset.

Some preprocessing steps were also applied to the political dataframe, namely removal of duplicates that were generated during the expansion process, and lowercasing of words to avoid case sensitive scenarios. To assess if a Tweet contained political content, we used hashtag pooling. In this scheme, the dataframe was composed of all tweets that mention a

given hashtag. It has been shown that this method outperforms the baseline scheme and user-polling as opposed to the other pooling system conducted, more specifically the default Tweet pooling (Albanese & Feuerstein, 2021). A custom function systematically examined the content of the 'hash_word' column. The function confirmed whether any of the words from the political dataset are present in the hashtag column. If the hashtag contained one of the words in the political dataset, the Boolean value *True* is assigned to a column, denoted as 'political_tweet'. This column contained Boolean values, with 'True' signifying the presence of political keywords within the respective tweet, and 'False' denoting their absence. This categorization helped in subsequent analyses, facilitating the isolation of tweets with political relevance to filter the dataset used for this case study.

4.9. MORAL FOUNDATIONS APPLICATIONS IN DIFFERENT LANGUAGE SETS

When working with other language sets, namely the Spanish tweets, which represented the 3rd biggest portion of the dataset, not all preprocessing techniques performed previously were applicable, and reproducible from the English sample.

Initially, we began by removing from the dataset the label mapping of the dataframe which was present on the continuation of the terms available for moral foundation categorization.

Table 4.2. Spanish MFD first 14 rows as a result from the Exploratory Data Analysis conducted prior to label removal. It includes label mapping in the same row/column and the words proceeding the labels.

Word	MF_1
%	
1 care.virtue	
2 care.vice	
3 fairness.virtue	
4 fairness.vice	
5 loyalty.virtue	
6 loyalty.vice	
7 authority.virtue	
8 authority.vice	
9 sanctity.virtue	
10 sanctity.vice	
%	
áreas de un sitio	5
íntegro	3

The steps related to the structure of the dataframe, that we could replicate from the previous work on the English dataset, were, namely the filtering of Spanish tweets with the lang attribute, the extended_tweet column transformation, the emoticons removal, hashtag extraction, duplicates removal, non-relevant columns drop and lastly, applying the custom function that lower cased text, removed words starting with "@", removed hashtags from the text column, applied lemmatization, and removed numerical punctuation. Stopwords were also eliminated from the sample with the NLTK library, by applying the Spanish corpus. The Spanish MFD also suffered natural language processing, with the application of the regex to remove punctuation, thus avoiding mismatches between both dataframes.

For the analysis of the Spanish moral foundations, we reproduced the steps conducted on the English subset. This included modelling topics and evaluating the moral foundations of each topic, analyzing moral foundations over time, and assessing the possible application of the political dictionary to assess the morality over the political discussions. Clustering was not applied for Spanish tweets, since we were reduced to a smaller portion of the dataset, and the English tweets' clustering had already low significance user distribution.

4.10 OTHER CONSIDERATIONS

Sentiment Analysis was conducted not only to explore the data during the Exploratory Data Analysis process, but also for assisting the analysis of the research conducted on clusters, viral topics, and political tweets in final stages of the research. The decision to undertake a sentiment analysis along with the morality extraction work was based on sustaining the virtue-vice dimensions of Moral Foundations, as these represent good and bad behaviors, and thus could bring additional conclusions when considering positive and negative sentiments.

As previously mentioned, the sentiment analysis was performed using the SentimentIntensityAnalyzer package from the NLTK library which helped assess the sentiment of the preprocessed text content. The sentiment scores were computed using the polarity_scores method from the package, which provided a comprehensive evaluation including a compound/polarity score indicative of the overall sentiment. The resulting sentiment scores were kept in an attribute denoted 'sentiment_scores'.

5. RESULTS AND DISCUSSION

As the data used for the study corresponded to users' comments, it is important to understand which type of users are under analysis. For the purpose of this work, a user segmentation was conducted, using clustering algorithms, and the results between them were compared. One of the main limitations of this segmentation, consisted in the fact that the dataset in question contained few attributes that could characterize the users, therefore only two columns were considered during clustering: location and timestamp of the tweet. Location was considered an ideal column among the available attributes. This can be justified by looking into the scope of social differences present within the state of New York (*New Yorkers in Need*, 2024). The inequality present in this state creates variations in education and wealth levels, even at a small-scale, such as different neighborhoods. When considering user characterization per region for assessing morality from Twitter data, it aligns with the idea that individuals' moral beliefs and behaviors can be influenced by geographical factors, environmental conditions, and sociocultural norms unique to each region, as the work of Götz et al. (2020) proves. In turn, this will introduce variability to the content of the tweets produced by the different locations. As for the timestamp measurement, this attribute was considered relevant for extracting user information, due to the assumption that people that work during night hours, are most likely tweeting during daytime. These jobs usually require shifts or working in night-life establishments which are considered, globally, low-income jobs. Adnan et al. (2014) also introduces results variability when analyzing tweeting times.

Using a reduced feature space can pose challenges, such as not categorizing users as efficiently into clusters. When visualizing this data and trying to interpret results, the reduced space could also be considered a benefit due to the simplification of the cluster and focus on key features. Another challenge worth mentioning was the lack of tweets geolocation as mentioned in the Data. To tackle this issue, we used the artificial samples thus making the results not completely accurate in terms of tweets spatial distribution. Kumar et al. (2014), suggests using profile location information to complement tweets' geolocation, however, this attribute was not originally extracted from the Twitter API on the sample used, and therefore, was not available for our analysis. To further extract and analyze morality, the artificial sample allowed us to see the results when comparing morality between user groups.

To perform clustering, two algorithms were selected for comparison, namely K-means and DBSCAN. In K-means we could observe significantly better results when clustering users. Even though K-means presents the advantage of being computationally efficient and easy to understand, it brought the main limitation of assuming that clusters were equally sized. With the algorithm split, the 2nd cluster contained 405222 tweets, and the 3rd contained only 8213 tweets, thus not necessarily yielding the desired results, as we can see on figure 5.1.

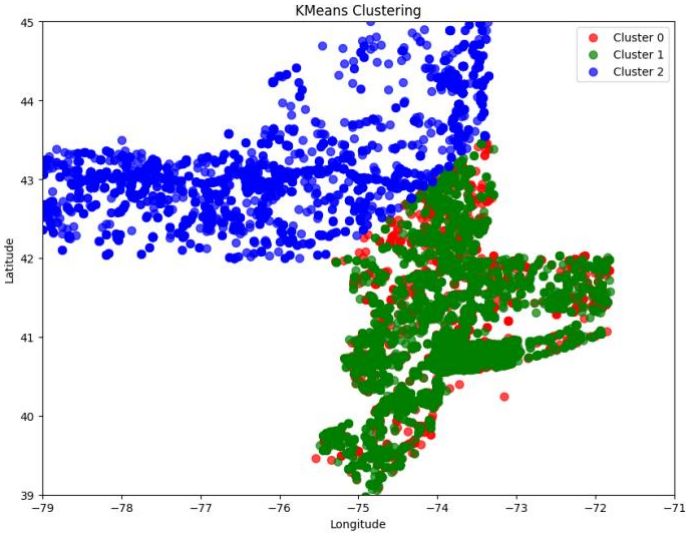

Figure 5.1. User Cluster Distribution on New York (latitude, in the y axis and longitude in the x axis) with K-means. Legend in the figure identifies each cluster by color dots, namely cluster 0 as red, cluster 1 as green, and cluster 2 as blue.

DBSCAN did not require specifying the number of clusters and even though it works better with arbitrary cluster shapes, after the eps and min_samples hyperparameters that control the density of the clusters, were iteratively adjusted, only one cluster was outputted by DBSCAN, as observed in figure 5.2. When comparing the results between both algorithms, we can conclude that according to DBSCAN's density-based criteria, the algorithm does not identify clear clusters in data, as opposite to K-means. For the morality analysis, K-means was therefore elected for user clustering.

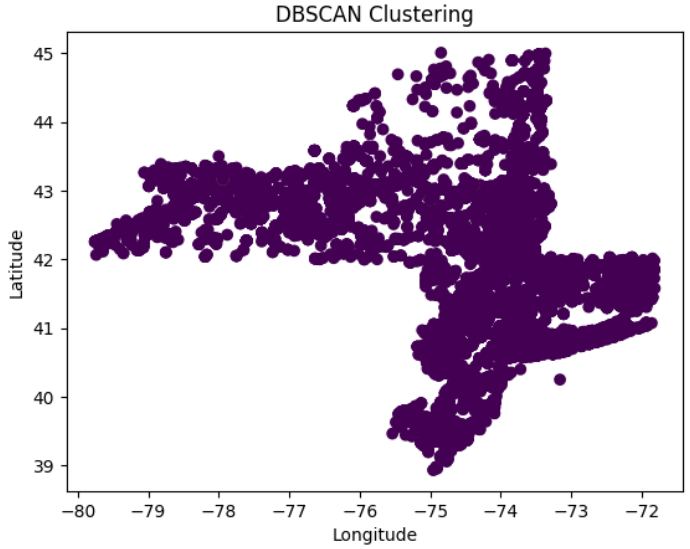

Figure 5.2. User Cluster Distribution (latitude, in the y axis and longitude in the x axis) with DBSCAN. The figure demonstrates only 1 cluster identifies with purple dots.

Even though cluster 0 and cluster 1 are geographically similarly distributed, these clusters represent different user groups, most likely due to the timestamp attribute of the tweets.

When considering cluster 2, we can see a different spatial distribution, which clearly distinguishes users' groups. By comparing to the statistics provided by Bureau (2024), Cluster 2 includes more zones which have higher rates of poverty, between 12% and 17.9%, as opposed to the other two clusters, which ranged between 0% to 11.9%. From interpreting the cluster, as mentioned, the tweeting times differentiates significantly from Cluster 0 and 1.

With the computation of the Moral Foundation' scores, the tweets contained the association strength of each moral dimension, and we could yield the highest associations among them. E.g. the tweet "The amount of support Im getting right now is making my heart melt" was mapped to loyalty-virtue, however other foundations such as care-virtue, fairness-virtue and authority-virtue, also gave high scores, thus indicating the positive connotation of the tweet. This can be confirmed by checking the sentiment analysis score, which resulted in a compound score of 0.4019, meaning that the tweet is leaning towards positivity. Other examples such as, "<namepii> And just like that, youre an insensitive jerk. Too bad we dont have a cure for that." which has a compound score of -0.6486, is effectively more associated to vice foundations, namely fairness and sanctity, which include purity foundations.

Not all scores, both for sentiment and moral foundations, reflect the reality of tweets. In this example "Still very relevant \n\nThe cults of personality that surround our faves whether theyre a political, entertainment, or otherwise public figureare a result of misguided worship of wealth and the status it affords...\n\ <link>" even though it shows a compound score of 0.802, indicating a very high positivity leaning, this tweet is more strongly associated with vice foundations, namely sanctity and loyalty, which from personal interpretation sounds more accurate.

The moral foundations score has shown, across the complete dataset, that most tweets were related to care foundations. This is a result from the higher frequency of care-vice foundations across the MFD. In contrast, when reviewing the reproducibility of the study in Spanish tweets we could conclude that Moral Foundations distribution differed from the English subset.

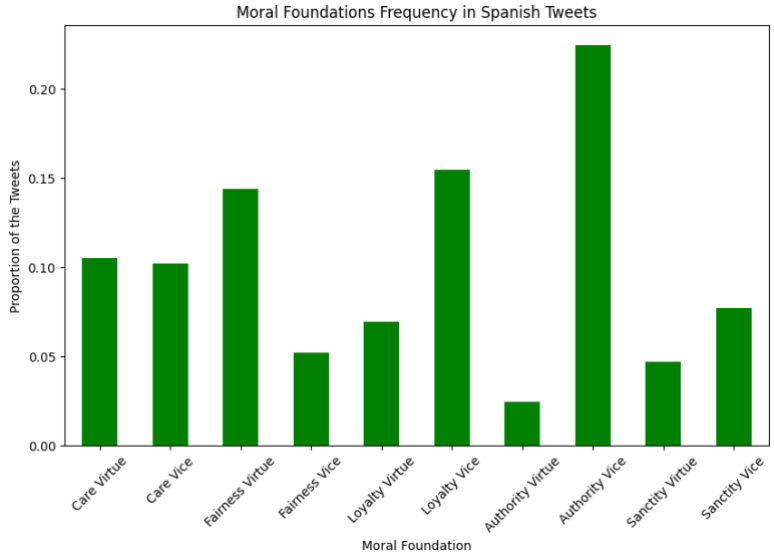

Figure 5.3. Distribution of Moral Foundations in the total Spanish Tweets, represented by frequency between 0 to 1.

The analysis of the distribution of tweets among foundations demonstrates that authority-vice is more often identified in the spanish tweets, opposite to authority-virtue, which is less present, as observed by figure 5.3. Care foundations are not so frequently represented in the overall corpus, as in the english subset.

Table 5.1. Spanish Tweets mapping to moral foundations, represented by the extended_tweet attribute of the dataset and the respective moral_foundation label.

extended_tweet	moral_foundation
Ojal pudiera grabar con mis ojos para que mi mam viera a detalle la ciudad tal como yo la veo	1
El jueves en el ltimo debate e// Trump & Biden, los candidatos tendrn silenciados sus micrfonos cuando NO sea su turno p/ hablar (En los segmentos d 2 despus quedarn abiertos) As lo dispuso la Comisin Debates Presidenciales p/evitar excesivas interrupciones #Election2020	4
Asqueada se estas basuras humanas llamados socialistas progresistas izquierdistas que son la misma mierda muy pocos de estos entes han producido algo menos an al desarrollado un Pas ellos solo saben destruir, Si yo fuera gobierno los quemo vivos as se acaban estas escorias. <link>	5
Sabrn los polticos ni siquiera escribir la palabra transparencia?	7
Yo no soy de ete planeta lo digo enserio puetaaa	2

When looking into specific examples, such as the tweet *"*<namepii> *Happy Birthday mi brother. Dios contine bendicindolo junto a los suyos, y por supuesto, que continen los xitos. ."* is mapped to sanctity-virtue, which is in fact related from analysing the content of the tweet. However, when looking into a tweet also mentioning sanctity figures, *"*<namepii> *Igual, ya fue y se gan gracias a Dios*", fairness-vice is identified, which might indicate lack of accuracy on the mapping taken, as it can lean both to fairness and sanctity. This can be justified by the labelling system used from the Spanish foundations dictionary, as well as for the low number of characters on the tweet itself, which represent smaller sample for analysis.

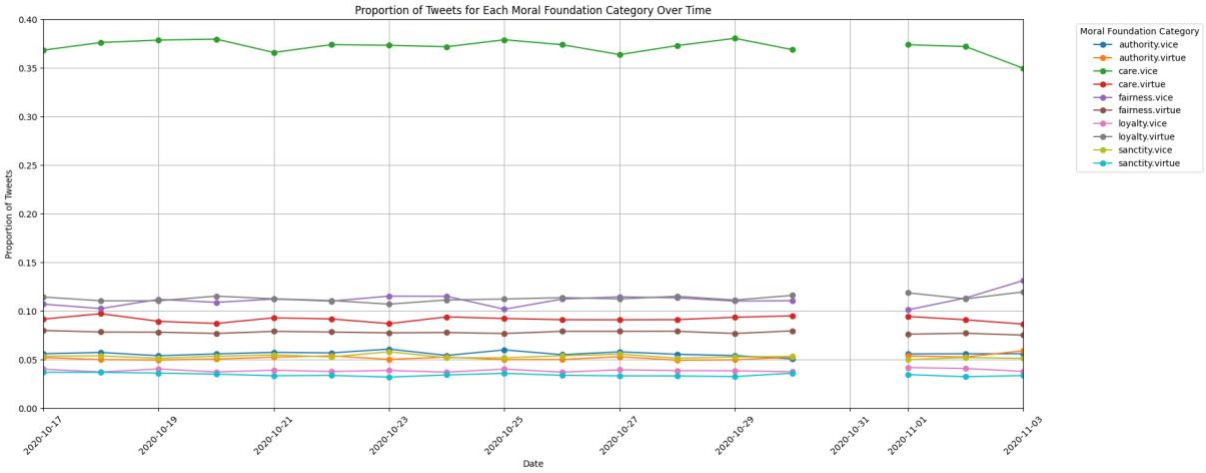

Figure 5.4. Temporal distribution of moral foundations in English Tweets, in proportion of total tweets of each day. Each foundation is represented by a distinct color identified on the legend. The day of 31st of October is marked as NaN value on the plot.

When analyzing the temporal distribution of morality, no significant changes happen across the sample, apart from the last days in the dataset that were Friday and Saturday, and in which president Joe Biden was elected the new president in the USA. Fairness-vice was the foundation that had the most significant increase.

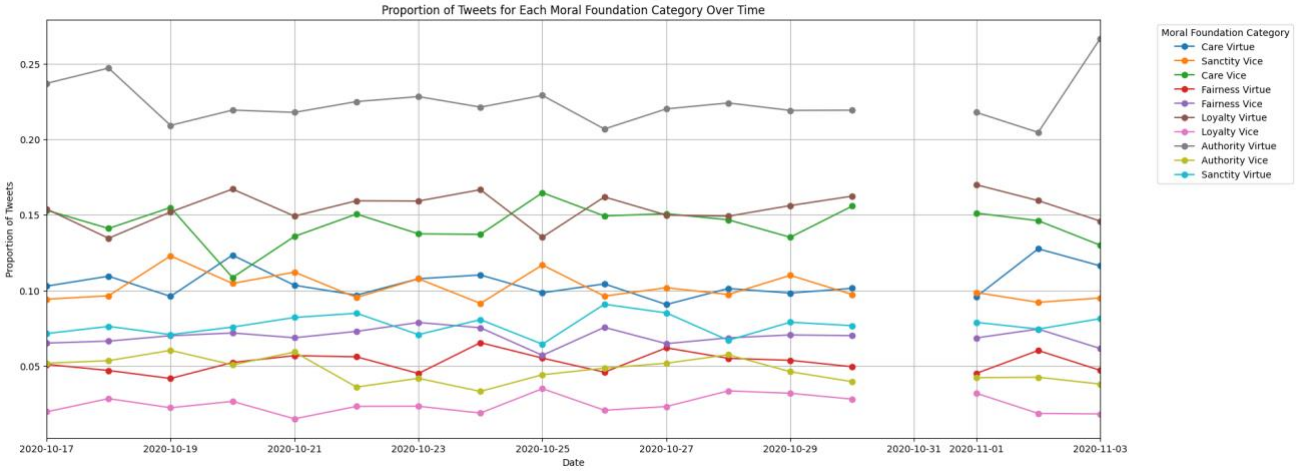

Figure 5.5. Temporal distribution of moral foundations in Spanish Tweets, in proportion of total tweets of each day. Each foundation is represented by a distinct color identified on the legend. The day of 31st of October is marked as NaN value on the plot.

In the Spanish tweets, the moral foundations evolution throughout time had no major differences, as well, however, on the last days of the corpus representation we can see increases in the authority and sanctity virtue foundations.

Regarding the identification of the main topics discussed on Twitter during this time period, the LDA showed results that were accurate with the real-life events happening during the sample time, the US presidential elections. Topic 1 included many words that could be considered stopwords, and words commonly used such as “like”, “get” “go”, “time”, one”, “lol”, “got”, “know”, “sh*t”, “cant”, “back”, “need”, “let”, “really”, “day”, “want”, “good”,

“going”, “year”, “see”, “yall”, “still”, “look”, “f*ck”, “make”, “never”, “feel”. The reason behind these words composing the main topic is likely due to the fact that they are very often used in speech. These terms were not removed in the data preprocessing steps as in some cases they may carry important semantic meaning and therefore are not considered stopwords. Within topic 2, that represented approximately 25% of the entire corpus, we could see words such as “trump”, “people”, “vote”, “biden”, “right”, “make”, “year”, “election”, “president”, “like”, “voting”, “american”, “joe”, “america”, “win” and “day” which are among the most relevant terms of this topic. “voting” and “american” were considered nouns when applying lemmatization, meaning that they were already on their base form. On the 3rd identified topic, an extension of common words used, like “love”, “thank”, “people”, “like”, “great”, “one” as well as trendy topics such as “covid”, are amongst the most present terms. Topic 4 modelled the discussions about New York life, pointing to many sites in the city: “newyork”, “posted”, “photo”, “video”, “city”, “Brooklyn”, “ny”, “pm”, “king”, “today”, “nyc”, “park”, “Manhattan”, “movie”, “queen”, “Bronx”, “jersey”, “yes”, “Sunday”, “rain”, “via”, “spade”, “awesome”, “nd”, “street”, “Halloween”, “Friday”, day”. Similar to this, topic 5 consisted of words related to more general daily city life such as “construction”, “exit”, “incident”, “love”, “cleared” “st” “ny”, ”direction”, ”thanks”, ”updated”, ”road”, ”gt”, ”side”, ”eb”, ”street” “nb”, “wb”, ”morning”, “sb” ”avenue”, “rd”, ”town”, “park”, “way”, “th”, ”line”, ”state”, ”av”, ”east”, ”county”, “route”.

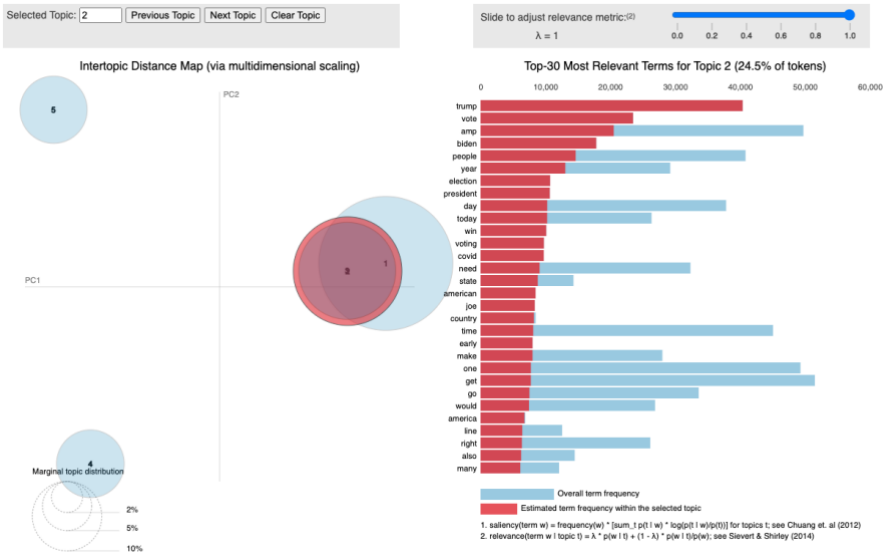

Figure 5.6. LDA visual and interactive representation of the most relevant terms and topics on the dataset after preprocessing.

The LDA parameters were defined with the coherence score calculation which resulted in a score of approximately 0.56 for the combination of 5 topics and 10 iterations, a moderate score meaning that there is some degree of semantic similarity.

When looking at the LDA results, prior to the cleaning and preprocessing steps applied to the corpus, only on the 4th topic, which has a representation of around 17% of the dataset, we can see words related to politics, namely “vote”, “party”, “ballot”, “voted” and “day”. The initial LDA considered words in other languages and with characters and punctuation, thus not clearly outlining the topics identified within the dataset. Additionally, the terms with biggest representation of the dataset, in topic 1, accounted for approximately 23%, as opposed to the topic modelling performed after preprocessing, which represented 37.2% of the sample.

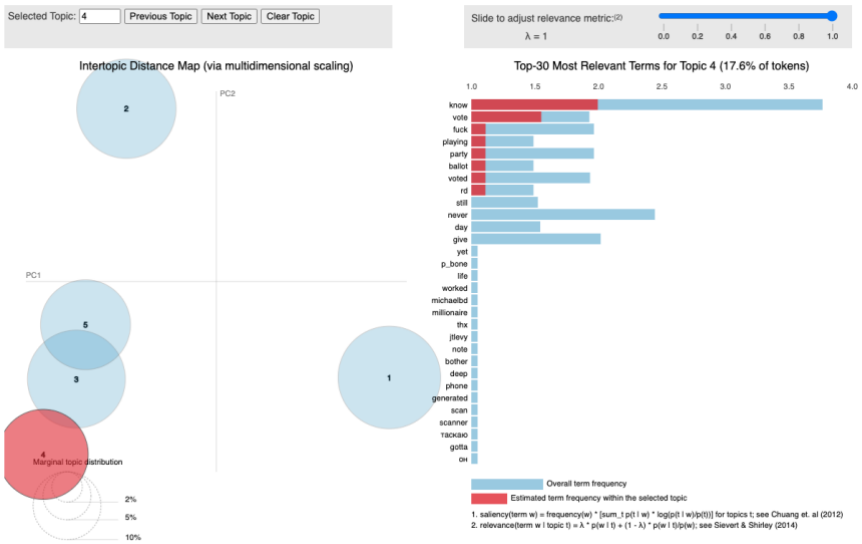

Figure 5.7. LDA visual and interactive representation of the most relevant terms and topics on the dataset before preprocessing.

The LDA of the spanish tweets yielded a coherence score of approximately 0.38%, after combining 5 topics and 20 passes. Since this score did not represent such a significant degree of similarity, which is visible by the results obtained when plotting the topics, we did not perform further exploration of the identified topics. Topic 1, in the spanish tweets topic modelling, provides similar results to topic 1 of the english tweets. The most frequent words, namely, “Si”, “ser”, “dios”, “ma”, “major”, “sempre”, “Pueblo”, “gracias”, “vida”, “asi”, “usted”, “pa”, “feliz”, “ahora”, “est”, “bien”, “va”, “gente”, “mismo”, “hermosa”, “trabajo”, “tener”, “presidente”, “hacer”, “tiempo”, “solo”, “nadie”, “persona”, “nunca”, “dia”, are commonly used in discussions, mainly among Latin-Americans that have a speech very characterized by their religious beliefs. On the second identified topic, the most frequent terms are similar to topic 1 namely, “si”, “gracias”, “da”, “tambin”, “aqu”, “ver”, “solo”, “ser”, “est”, “voy”, “dios”, “ahora”, “presidente”, “persona”, “va”, “bueno”, “ah”, “sempre”, “hacer”, “quiero”, “qu”, “pa”, “gt”, “tan”, ”veces”, ”modo”, ”verdad”, ”amigo”, ”nadie”. E.g. Words such as “presidente”, “ahora”, and “gracias” are present in both topics. Topic 3 does not clearly indicate any subject of a discussion, rather it shows many slang words, which is expected from Twitter comments. Topic 4 points to the terms “biden” and “trump”, which may indicate a possible political trait in tweets discussion. The top 30 most relevant terms of this topic are “da”, “bien”, “hoy”, “buenos”, “pa”, “qu”, “trump”, “feliz”, “semana”, “ser”,

“gajaja”, “amore”, “st”, “bella”, “si”, “estn”, “mundo”, “besitos”, “princesa”, “toda”, “bendiciones”, “biden”, “etapa”, “mejor”, “nunca”, “viernes”, “sigue”, “bueno”, and “fingente”. Topic 5 terms “new”, “york”, “numero”, “foto”, “acaba”, “hoy”, “publicar”, “felicidades”, “ganhador”, “da”, “ganhadores”, “vídeo”, “quiniela”, “si”, “jajajaja”, “aos”, “solo”, “pany”, “jersey”, “Manhattan”, “hace”, “via”, “ser”, “tiempoa”, “pm”, “do”, “nueva”, “city”, “puede” displays a similar structure as topic 4 of the english tweets, mentioning New York life and city sites.

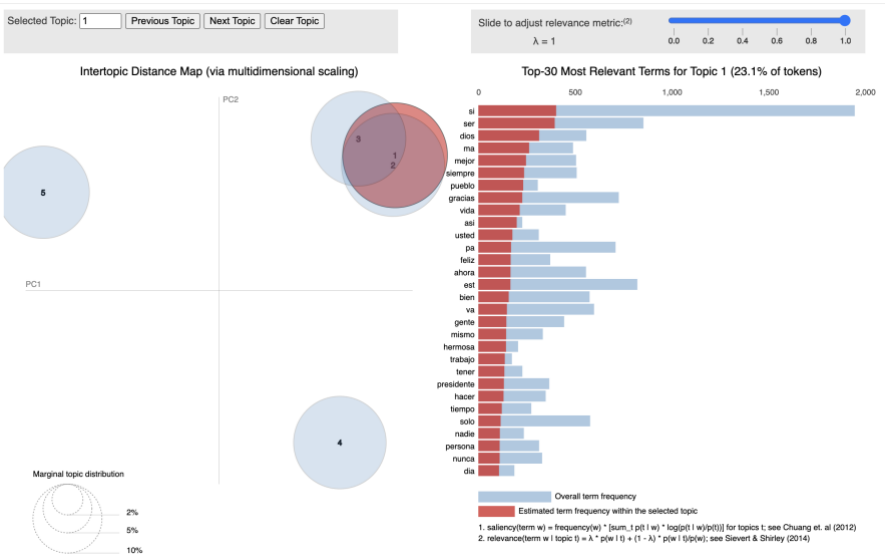

Figure 5.8. LDA visual and interactive representation of the most relevant terms and topics in the Spanish tweets.

Defining the viral topics within the English corpus was of extreme importance as it allowed us a greater understanding of the morality extracted, since other dimensions can be considered, namely the context of the analysis being conducted, and the occurrence of events beyond social media. This topic modelling proved our assumption regarding the influence of real-life events on social media, and thus the decision to study in more detail political annotation tweets. On the example in table 5.2, we can see the tweets that contained the word “biden” that were mapped to the political topic.

Table 5.2. English Tweets mapping to the topic label, extracted from the LDA analysis, in which the Term "Biden" is present.

extended_tweet	topic
An odd insult. RT Trump: Biden will 'listen to the scientists' if elected <link>	2
This means we are now in a fairly likely scenario where after spending the entire summer making the central premise of his campaign that Joe Biden literally has dementia, Trump will have gotten his ass kicked in the first debate and then dropped out of the final 2	2
Why These Voters Rejected Hillary Clinton but Are Backing Joe Biden <link>	2
You folks could only hope you Biden supporters and socialist government supporters!!! Bind put more laws on the books against black people than any other person in politics!! The man is a disgrace to societee in a disgusting sexual predator. Trump 2020!!! <link>	2
yes, the brunch crowd is getting way too excited, but im fine with not hearing from them if/when biden takes office	1

The analysis of the moral foundations' distribution across viral topics, showed a prevalence of care-vice foundation across all topics, while all other foundations varied greatly.

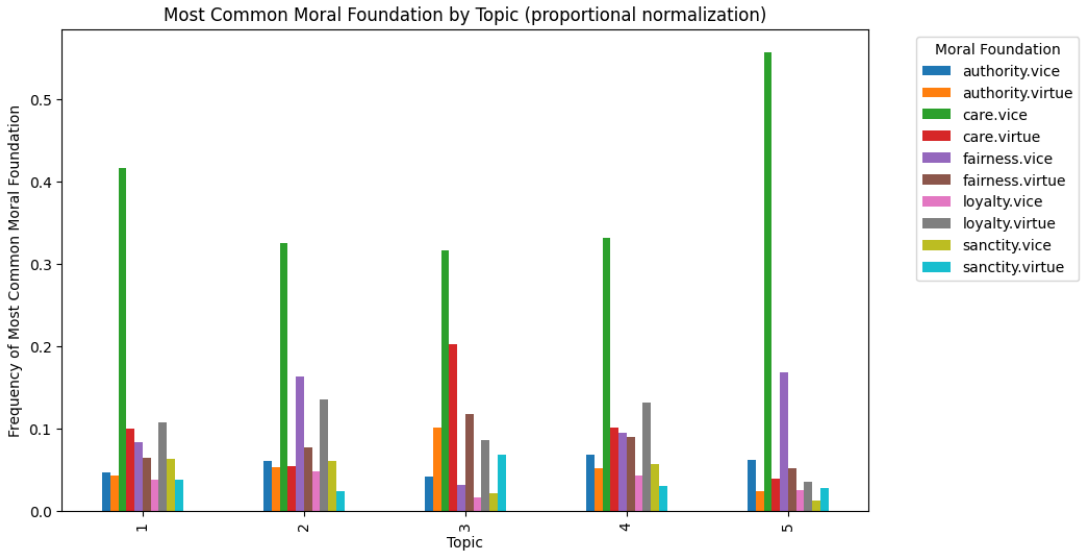

Figure 5.9. Moral foundations per topic (x axis), identified in the LDA analysis, in which each moral foundation is represented by a color, and each topic contains these foundations in side-by-side bar charts. The y axis represents the frequency of the most common moral foundations out of the total of tweets of each topic.

Regarding topic 2, fairness-vice and loyalty-virtue are among the most frequent foundations, as seen by figure 5.9, which can be linked respectively, to a democratic speech and political party affinity speech.

Our analysis on the political dataframe, which corresponds to one of the viral topics identified, was conducted by filtering tweets that had political annotation. To filter the tweets, we tried to identify political terms on the hashtags extracted. E.g. the tweet "#Peace is an indispensable precondition to #development.As such,urges the int. community to call upon Israel to immediately end its occupation of Palestinian territories;cease its appalling and colonising practices;& immediately comply with all relevant <namepii> resolutions\n\n@ #2C <link>" which is mainly associated with fairness-vice and authority-vice foundations, was identified as political.

Using only hashtags to find political annotations in tweets imposed the challenge of missing additional rows within the same topic. The percentage of hashtags in the dataset was 15.59% and, out of these tweets, only 9.68% were mapped to a word in the political dictionary.

Table 5.3. Top 10 Hashtags in the Tweets with political annotations, descending in relative frequency of occurrence.

Hashtags of Political Subset	Relative Frequency
vote	1
debates2020	0,522508
bidenharris2020	0,418499
election2020	0,304625
nyc	0,267215
vote2020	0,254676
voteearly	0,253443
electionday	0,188489
presidentialdebate2020	0,1852
vote	1

Regarding the moral foundations' distribution, when looking into the political filtering, the care-vice trend extends to these tweets, and we can see that fairness-vice, loyalty-virtue, and authority-vice are among the most frequent foundations of the dataset. From the society lens, these foundations fit into the political context mainly with loyalty-virtue due to the sense of loyalty with political figures or leaning during the speech, and authority-vice which can reflect the imposition of ideas, which usually happens when sharing thoughts on a subject in social media.

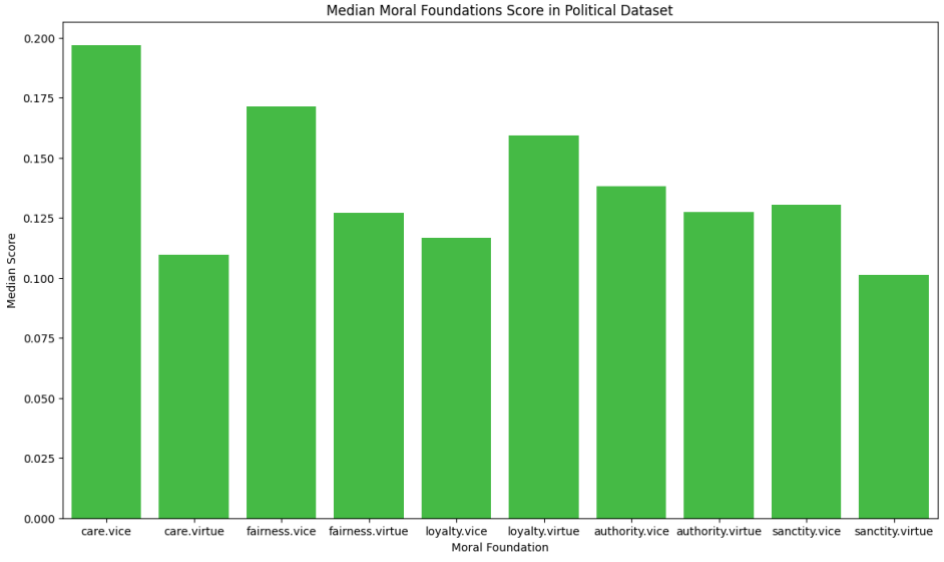

Figure 5.10. Median of moral foundations in the Tweets with political annotations.

In order to identity political leaning in tweets, the distribution of moral foundations in tweets that contained the word "Trump" and separately the word "Biden" were plotted.

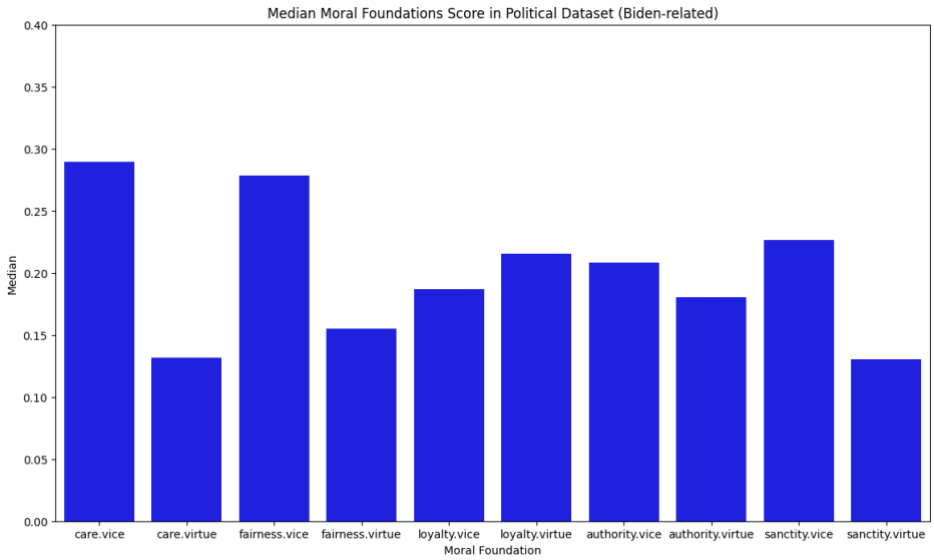

Figure 5.11. Median of moral foundations in the Tweets with political annotations and that contain the term "biden".

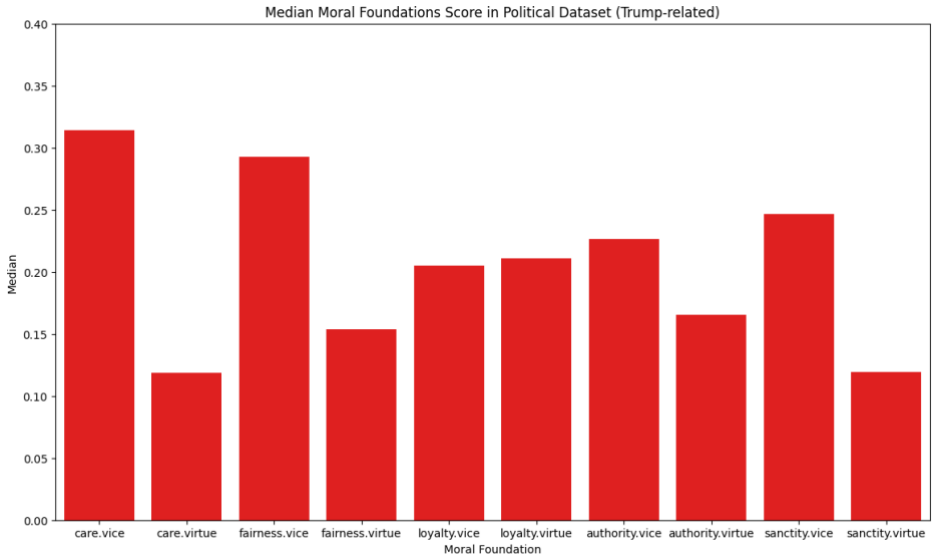

Figure 5.12. Median of moral foundations in the Tweets with political annotations and that contain the term "trump".

When comparing foundations between the candidate last names, “trump” has a prevalence of vice related foundations, and “biden” has a bigger virtue foundations score. This can be assumed when analyzing the presidential campaign results, since President Joe Biden won the elections, which can justify more support and positive speech associated to the tweets that mention him, while Donald Trump had less supporters which was reflected in the elections result. E.g. the tweet “I spoke to many enthusiastic Biden-Harris supporters from the state of Ohio during this evenings NY for Biden phone bank! I feel a blue wave in Ohio! #OhioForBiden <namepii> #BidenHarris2020Landslide #VoteBlue” has a positive compound score of 0.6486 and loyalty-virtue was the foundation with highest association score to this tweet, compared to “<namepii> NOT JUST THE FEDS,CIA ALSO IF U HAVE DOUBT ABOUT UR FED AND LAW

INSTITUTIONS THEN THANK TRUMP,HE GOT U THINKING THAT WAY,WHICH ISN'T NORMAL,THAT'S BULL WE BELIEVED IN THEM WAY BEFORE TRUMP, NOW UR CONFUSED,HAVE A MIND OF UR OWN,STOP CONSPIRACY THEORIES <link>", which has a negative compound score of -0.7845 and the strongest association is with authority-vice foundations. Despite these differences, both terms retrieve very similar scores as many times the two names are present under the same tweet, thus yielding the similar moral foundation scores. Further filtering is required to understand the principal corpus subject of each tweet mapped to the term "biden" and "trump".

For the consequent analysis, we assume and compare the speech of each political figure, Joe Biden and Donald Trump, reflected on the speech of Twitter users. Regarding the Republican-Democrat variability of moral foundations, which share similar values of Conservative-Liberal and Right-Left spectrums, we could not reach any major conclusions on the morality of these spectrums' viewpoint. The tweets mapped to the term "Biden", which belonged to the Democrat party, had slightly bigger scores for care and fairness virtue foundations, when compared to "Trump" and higher scores for authority and sanctity vice, when compared to foundations in virtue scale, which proves the theory proposed in the Political Annotations in Data section, when considering the above mentioned assumption. However, tweets with "Trump", only had loyalty-virtue with a score higher than the vice foundation. These results also defuse the conclusions of Lewis P.G. (2019), when considering our assumption, however multiple factors such as the low number of characters in certain tweets, the lack of additional filtering of "biden vs trump" discussions and the low variability between foundations scores, as well as higher frequency of terms identified as "care" in the MFD could justify these results. Thus, it was not possible to conclude political spectrum morality association entirely on our sample, and only a Twitter user moral-speech leaning between Biden and Trump.

The user cluster analysis against the political tweets indicates, mainly within cluster 0 and 1 a similar result from hashtags.

Table 5.4. Top 10 Hashtags in the Tweets with political annotation of Cluster 0, descending from relative frequency of occurrence.

Hashtags of Cluster 0	Relative Frequency
vote	1
bidenharris2020	0,359058
election2020	0,340882
vote2020	0,271752
electionday	0,261025
nyc	0,188319
voteearly	0,177294
trump2020	0,109952
votehimout	0,109952
2020election	0,096544

Table 5.5. Top 10 Hashtags in the Tweets with political annotation of Cluster 1, descending from relative frequency of occurrence.

Hashtags of Cluster 1	Relative Frequency
debates2020	1
vote	0,586383
presidentialdebate2020	0,354585
bidenharris2020	0,32625
nyc	0,262889
voteearly	0,250295
voteearlyday	0,182212
debatetonight	0,177883
debate2020	0,14758
election2020	0,131444

Table 5.6. Top 10 Hashtags in the Tweets with political annotation of Cluster 2, descending from relative frequency of occurrence.

Hashtags of Cluster 2	Relative Frequency
skilledtrade	1
rochester	0,978723
warehouse	0,680851
upstateny	0,659574
distributionjobs	0,489362
ny	0,446809
securityofficer	0,425532
securityguard	0,404255
vote	0,404255
newyork	0,319149

Cluster 1 shows for sanctity-vice a more relevant representation of the foundation, however, cluster 2 has always higher medians in the majority of foundations. This information was plotted as median to allow a more accurate comparison between clusters since these represent different shape sizes.

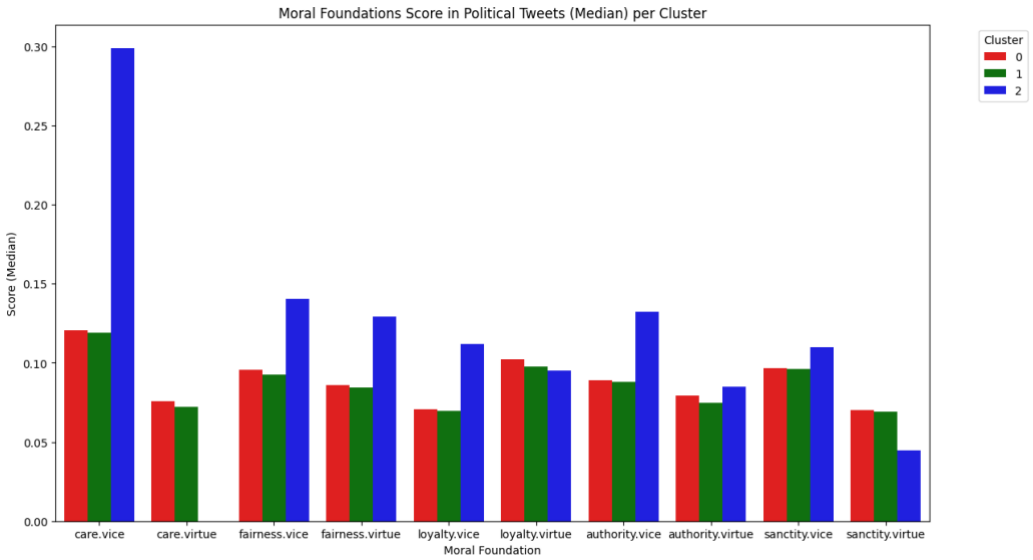

Figure 5.13. Moral foundations median distribution per user cluster. Legend in the figure identifies each cluster by color, namely cluster 0 as red, cluster 1 as green, and cluster 2 as blue.

One of the biggest downsides with the use of “contains” method to assess tweets’ political annotation from hashtags, or even tweets, is the fact that it will scan through the complete word and return the value if it contains the term. E.g. In the political dictionary, one of the political figures outlined is Bill Clinton, former US President. When analyzing the political tweets subset, we can see many references to the hashtag #BillsMafia, *“Alone with my beer, wings, and hopefully a bills win. Ahhh. Tonight is the fucking night! #letsgo #BillsMafia”* . The hashtag "BillsMafia" is a term used by fans of the Buffalo Bills, a professional American football team based in Buffalo, New York, and thus has a significant presence among political tweets as it matches “Bill” from Bill Clinton. This result is further portrayed through a false positive, meaning that the hashtags retrieved are not all related to the political context when using the method contains.

When looking at the hashatgs available in the Spanish subset, these accounted for 10% presence in the overall tweets. The hashtag frequency in table 5.7, depicts some terms that point to the elections, namely “Debates2020”, “Elecciones2020”, “Elections2020”, and “Vote”. Considering that some of these terms are in Spanish, the political dictionary is in English, and the fact that hashtags did not have any significant representation, we did not perform the case study analysis over this subset. Some libraries such as googletrans, allowed a programmatically conversion of the Political Dictionary, however these frameworks have usage limits and service terms and therefore were not considered for the purpose of this work.

Table 5.7. Top 10 Hashtags in the Spanish Tweets, descending from relative frequency of occurrence.

Hashtag Spanish	Relative Frequency
Giro	1
VueltaAGuatemala	0,639241
LaVuelta20	0,632911
Carapaz	0,518987
NYC	0,455696
Election2020	0,398734
Ecuador	0,398734
LaVuelta	0,386076
COVID19	0,322785
blazemusicnet	0,272152

6. CONCLUSIONS AND FUTURE WORKS

This work has made significant developments to leverage Twitter data and uncover the underlying morality in discussions related to viral topics. The application of the MFT has provided valuable insights into the variations in moral speech, particularly in the context of political ideologies and cross-cultural differences, such as the case of Spanish speaking users.

The lack of user details, the low percentage of geolocated tweets, the application of topic modelling to the Spanish tweets, which did not allow an analysis over the political discussions case study, and the limitation of words on the MFD, that even though has been expanded, does not reflect all social media corpus, imposed several challenges for the research. The study also grappled with the complexities of identifying political discussions, namely due to the methodology used to extract political annotations, and due to the lack of a concise and complete database of political vocabulary. Finally, the lack of available and standardized moral foundations dictionary across other languages, containing a similar number of terms, and an unbalanced foundation distribution in the moral foundation dictionary itself, posed challenges for general comparisons. These limitations underscore the need for further refinement and expansion of the MFD, to capture the full range of human moral experience, cross-culturally and ideologically. The length of the texts within the tweets do not allow for conclusive results regarding morality since the meaning which is being portrayed by the tweet may be taken out of context due to the lack of words. The consideration of a bigger and more time-varied Twitter sample may also help yielding better conclusions.

Despite these challenges, the research successfully developed a robust methodology capable of providing insights and predictions on diverse dimensions, such as discerning the presence of morality in specific topics. This was achieved by advanced preprocessing and NLP techniques, including topic modeling and sentiment analysis, as well as with the application of user segmentation and visual analysis. The outcomes of this study are manifold, as it has contributed to the evolving landscape of social science studies by developing in-context learning, which can be used to find semantically similar natural language utterances. This can be particularly beneficial for understanding and predicting public sentiment and behavior in response to various societal issues.

Looking ahead, the developed study could be applied to other social media platforms and used to explore the impact of morality on other aspects of social behavior. The application of the MFT could also be explored in different language sets and countries, broadening the scope and applicability of this research. This study could be extended to include a more comprehensive analysis of the role of morality in shaping the lifetimes of topics and the dynamics of user interactions on social media platforms, which could be achieved with a larger dataset that gathers other important time periods in history. Furthermore, using social networks that collect longer texts may help achieving more accurate conclusions from morality, when using the proposed approach.

In summary, this work has demonstrated the potential of using social media data to uncover the underlying morality in discussions related to viral topics. Despite the challenges, the study has provided valuable insights and developed a data mining approach that can be used for future research and applications.

BIBLIOGRAPHICAL REFERENCES

10 Most Dangerous Neighborhoods in NYC. (2024, February 5). PropertyClub. https://propertyclub.nyc/article/most-dangerous-neighborhoods-in-nyc

10 things you need to know today: November 3, 2020 | The Week. (2024, February 4). https://theweek.com/10things/947491/10-things-need-know-today-november-3-2020

50 Most Expensive Neighborhoods in NYC. (2024, February 5). PropertyClub. https://propertyclub.nyc/article/most-expensive-neighborhoods-in-nyc

Adnan, M., Longley, P. A., & Khan, S. M. (2014). Social dynamics of Twitter usage in London, Paris, and New York City. *First Monday*. https://doi.org/10.5210/fm.v19i5.4820

Albanese, F., & Feuerstein, E. (2021). *Improved Topic modeling in Twitter through Community Pooling* (arXiv:2201.00690). arXiv. http://arxiv.org/abs/2201.00690

Americans and Twitter: Key facts as it rebrands to X | Pew Research Center. (2024, February 4). https://www.pewresearch.org/short-reads/2023/07/26/8-facts-about-americans-and-twitter-as-it-rebrands-to-x/

Annual Report—MOIA. (2024, February 9). https://www.nyc.gov/site/immigrants/about/annual-report.page

Araque, O., Gatti, L., & Kalimeri, K. (2020). MoralStrength: Exploiting a moral lexicon and embedding similarity for moral foundations prediction. *Knowledge-Based Systems*, *191*, 105184. https://doi.org/10.1016/j.knosys.2019.105184

Ashford, J. R., Turner, L. D., Whitaker, R. M., Preece, A., & Felmlee, D. (2022). Understanding the characteristics of COVID-19 misinformation communities through graphlet analysis. *Online Social Networks and Media*, *27*, 100178. https://doi.org/10.1016/j.osnem.2021.100178

Azank, F. (2022, October 17). Clustering—Conceitos básicos, principais algoritmos e aplicação. *Turing Talks*. https://medium.com/turing-talks/clustering-conceitos-b%C3%A1sicos-principais-algoritmos-e-aplica%C3%A7%C3%A3o-ace572a062a9

Bureau, U. C. (2024, February 26). *The Chance That Two People Chosen at Random Are of Different Race or Ethnicity Groups Has Increased Since 2010*. Census.Gov. https://www.census.gov/library/stories/2021/08/2020-united-states-population-more-racially-ethnically-diverse-than-2010.html

Buskirk, T. D., Blakely, B. P., Eck, A., McGrath, R., Singh, R., & Yu, Y. (2022). Sweet tweets! Evaluating a new approach for probability-based sampling of Twitter. *EPJ Data Science*, *11*(1), 9. https://doi.org/10.1140/epjds/s13688-022-00321-1

Capitol riots timeline: What happened on 6 January 2021? (2021, February 10). *BBC News*. https://www.bbc.com/news/world-us-canada-56004916

Capitol riots timeline: What happened on 6 January 2021? (2021, February 10). *BBC News*. https://www.bbc.com/news/world-us-canada-56004916

Critiques | Moral Foundations Theory. (2024, February 4). https://moralfoundations.org/critiques/

Davis, D. E., Rice, K., Van Tongeren, D. R., Hook, J. N., DeBlaere, C., Worthington, E. L., & Choe, E. (2016). The moral foundations hypothesis does not replicate well in Black samples. *Journal of Personality and Social Psychology*, *110*(4), e23–e30. https://doi.org/10.1037/pspp0000056

Doan, T. M., Kille, B., & Gulla, J. A. (2022). Using Language Models for Classifying the Party Affiliation of Political Texts. In P. Rosso, V. Basile, R. Martínez, E. Métais, & F. Meziane (Eds.), *Natural Language Processing and Information Systems* (Vol. 13286, pp. 382–393). Springer International Publishing. https://doi.org/10.1007/978-3-031-08473-7_35

gal007. (2019, February 7). *NLTK available languages for stopwords* [Forum post]. Stack Overflow. https://stackoverflow.com/q/54573853

Garza, A. (2021, May 25). Segmentation Analysis with K-means Clustering. *Analytics Vidhya*. https://medium.com/analytics-vidhya/segmentation-analysis-with-kmeans-clustering-93d05565a9f8

Gehman, R., Guglielmo, S., & Schwebel, D. C. (2021). Moral foundations theory, political identity, and the depiction of morality in children's movies. *PloS One*, *16*(3), e0248928. https://doi.org/10.1371/journal.pone.0248928

Geller, M., Vasconcelos, V. V., & Pinheiro, F. L. (2023). Toxicity in Evolving Twitter Topics. In J. Mikyška, C. De Mulatier, M. Paszynski, V. V. Krzhizhanovskaya, J. J. Dongarra, & P. M. A. Sloot (Eds.), *Computational Science – ICCS 2023* (Vol. 10476, pp. 40–54). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36027-5_4

Götz, F. M., Stieger, S., Gosling, S. D., Potter, J., & Rentfrow, P. J. (2020). Physical topography is associated with human personality. *Nature Human Behaviour*, *4*(11), 1135–1144. https://doi.org/10.1038/s41562-020-0930-x

Graff, M., Moctezuma, D., Miranda-Jiménez, S., & Tellez, E. S. (2022). A Python library for exploratory data analysis on twitter data based on tokens and aggregated origin–destination information. *Computers & Geosciences*, *159*, 105012. https://doi.org/10.1016/j.cageo.2021.105012

Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2021). The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods*, *53*(1), 232–246. https://doi.org/10.3758/s13428-020-01433-0

How Many People Use Twitter? (2019–2024) [Jan 2024 Update]. (2024, February 4). https://www.oberlo.com/statistics/how-many-people-use-twitter

Jiang, J., Thomason, J., Barbieri, F., & Ferrara, E. (2023). Geolocated Social Media Posts are Happier: Understanding the Characteristics of Check-in Posts on Twitter. *Proceedings of the 15th ACM Web Science Conference 2023*, 136–146. https://doi.org/10.1145/3578503.3583596

Kaur, R., & Sasahara, K. (2016). Quantifying moral foundations from various topics on Twitter conversations. *2016 IEEE International Conference on Big Data (Big Data)*, 2505–2512. https://doi.org/10.1109/BigData.2016.7840889

Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter Data Analytics*. Springer New York. https://doi.org/10.1007/978-1-4614-9372-3

Lambert, S. (2017, August 17). The 5 Moral Foundations. *The Center for Artistic Activism*. https://c4aa.org/2017/08/the-5-moral-foundations

Left–right political spectrum. (2024). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Left%E2%80%93right_political_spectrum&oldid=1198290472

Lewis, P. G. (2019). Moral Foundations in the 2015-16 U.S. Presidential Primary Debates: The Positive and Negative Moral Vocabulary of Partisan Elites. *Social Sciences*, *8*(8), 233. https://doi.org/10.3390/socsci8080233

madams. (2020, May 28). The Evolution of Social Media: How Did It Begin, and Where Could It Go Next? *Maryville University Online*. https://online.maryville.edu/blog/evolution-social-media/

Matsuo, A., Sasahara, K., Taguchi, Y., & Karasawa, M. (2019). Development and validation of the Japanese Moral Foundations Dictionary. *PLOS ONE*, *14*(3), e0213343. https://doi.org/10.1371/journal.pone.0213343

Medianeuroscience/emfdscore. (2024). [Jupyter Notebook]. Media Neuroscience Lab. https://github.com/medianeuroscience/emfdscore (Original work published 2019)

miku. (2009, November 24). *Answer to "What is the difference between lemmatization vs stemming?"* Stack Overflow. https://stackoverflow.com/a/1787121

Moral Foundations Theory | moralfoundations.org. (2024). Retrieved February 4, 2024, from https://moralfoundations.org/

Moral Foundations Theory—The Behavioral Scientist. (2024, February 4). Https://Www.Thebehavioralscientist.Com. https://www.thebehavioralscientist.com/glossary/moral-foundations-theory

New Yorkers in Need: A Look at Poverty Trends in New York State for the Last Decade | Office of the New York State Comptroller. (2024, February 6). https://www.osc.ny.gov/reports/new-yorkers-need-look-poverty-trends-new-york-state-last-decade

Ntompras, C., Drosatos, G., & Kaldoudi, E. (2022). A high-resolution temporal and geospatial content analysis of Twitter posts related to the COVID-19 pandemic. *Journal of Computational Social Science*, *5*(1), 687–729. https://doi.org/10.1007/s42001-021-00150-8

O ano em Pesquisa da Google—Google Trends. (2024, February 4). https://trends.google.com/trends/yis/2020/US/

Sklearn.impute.SimpleImputer. (2024, February 7). Scikit-Learn. https://scikit-learn/stable/modules/generated/sklearn.impute.SimpleImputer.html

Spanish-mfd/dicts/dict_es_final.dic at master · abruegger/Spanish-mfd. (2024, February 4). GitHub. https://github.com/abruegger/Spanish-mfd/blob/master/dicts/dict_es_final.dic

Topic Identification with Gensim library using Python—Analytics Vidhya. (2024, February 4). https://www.analyticsvidhya.com/blog/2022/02/topic-identification-with-gensim-library-using-python/

Vidiyala, R. (2021, June 9). *Topic Modelling on NYT articles using Gensim, LDA*. Medium. https://towardsdatascience.com/topic-modelling-on-nyt-articles-using-gensim-lda-37caa2796cd9

Viganola, D., Eitan, O., Inbar, Y., Dreber, A., Johannesson, M., Pfeiffer, T., Thau, S., & Uhlmann, E. L. (2018). Datasets from a research project examining the role of politics in social psychological research. *Scientific Data*, *5*(1), 180236. https://doi.org/10.1038/sdata.2018.236

Viganola, D., Eitan, O., Inbar, Y., Dreber, A., Johannesson, M., Pfeiffer, T., Thau, S., & Uhlmann, E. L. (2018). Datasets from a research project examining the role of politics in social psychological research. *Scientific Data*, *5*(1), 180236. https://doi.org/10.1038/sdata.2018.236

Weerasooriya, T., Perera, N., & Liyanage, S. R. (2017). *KeyXtract Twitter Model—An Essential Keywords Extraction Model for Twitter Designed using NLP Tools* (arXiv:1708.02912). arXiv. http://arxiv.org/abs/1708.02912

Zúquete, M., Orghian, D., & Pinheiro, F. L. (2023). A Moral Foundations Dictionary for the European Portuguese Language: The Case of Portuguese Parliamentary Debates. In J. Mikyška, C. De Mulatier, M. Paszynski, V. V. Krzhizhanovskaya, J. J. Dongarra, & P. M. A. Sloot (Eds.), *Computational Science – ICCS 2023* (Vol. 14073, pp. 421–434). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35995-8_30

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa