INTERACTIVE TOPIC-GUIDED THEMATIC ANALYSIS FOR SOCIAL MEDIA DATA

by

Subham Kumar Sah

A thesis submitted to the faculty of The University of North Carolina at Charlotte in partial fulfillment of the requirements for the degree of Master of Science in Architecture and Master of Science in Information Technology

Charlotte

2022

Approved by:

Prof. Eric Sauda

Prof. Wenwen Dou

Dr. Alireza Karduni

©2022 Subham Kumar Sah ALL RIGHTS RESERVED

ABSTRACT

SUBHAM KUMAR SAH. Interactive Topic Guided Thematic Analysis for Social Media Data (Under the direction of PROF. ERIC SAUDA)

Social media plays a significant role in modern-day information dissemination, owing to its ubiquity and ease of sharing. There has been an increased focus on analyzing topics frequently occurring in social media.

We refer to social media as a platform that encompasses all media types through which user groups engage in dispersed, networked, and parallel information production, sharing, and augmentation activities. The most well-known examples include blogs, discussion forums, and Twitter (brief text messages) (commentary and discourse). Users can create interest groups or other forms of connections using these services, giving birth to relationship qualities in addition to content (for example, leader/follower connections on Twitter). Over a billion people use social media now to interact, produce, and share information, ideas, and insights after it first gained popularity among a small group of enthusiasts. Because of the diversity and openness of its numerous cultural venues, online social spaces are becoming acknowledged as significant areas for qualitative social science research. Novel commercial, scholarly, and governmental applications that evaluate "these datas" in order to learn new things and make better decisions have resulted from this. Customer interactions, the financial market, demographics, etc. are a few examples. Users create material in the social media space in a variety of formats, including geographical data, text, photos, videos, and photographs. This information may be used for a variety of things, such as by businesses to streamline operations, by policymakers to spot trends in public opinion, and by public health officials to track the spread of infectious diseases or organize relief efforts in the wake of a natural catastrophe. Along with other scholars, social and political scientists investigate social media as a cultural mirror.

Leveraging social media data, however, comes with a lot of difficulties. It is typically sent in high-frequency streams and has a big volume. The data is also multimodal, frequently unclear in substance, and heavily reliant on context and user. Communication patterns within and across the various social media platforms very quickly. The goal for the thesis is to address this challenge by demonstrating the value of a 'visual analytic' approach to capturing and exploring the qualitative and subjective facets of Social Media data as a socio-technical research clustering.

TABLE OF CONTENTS

LIST OF FIGURES	V						
CHAPTER 1: INTRODUCTION							
CHAPTER 2: BACKGROUND & LITERATURE REVIEW	3						
CHAPTER 3: IMPLEMENTATION	5						
3.1 First Prototype	5						
3.2 Evaluation of Application	8						
3.3 Results from the Interviews	9						
3.4 Literature Review: Thematic Analysis	10						
3.5 Researchers' Interview & User Journey	16						
3.6 Final Prototype	18						
CHAPTER 4: RESULTS FROM THE STUDY	24						
CHAPTER 5: DISCUSSION AND FUTURE WORKS	25						
CHAPTER 6: CONCLUSION	27						
BIBLIOGRAPHY							
APPENDIX							

LIST OF FIGURES

Figure 1: Shows the user journey of first prototype	11
Figure 2: Shows the user journey of the Second prototype	12
Figure 3: Shows the user journey of Third prototype	13
Figure 4: Shows the Final Prototype Developed for Interviews	13
Figure 5: Shows the Themes developed for the paper Enmeshing	18
Figure 6: Shows the Themes developed for the paper I Will Not Drink With You today	y. 19
Figure 7: Final user Journey for Thematic Analysis	21
Figure 8: User uploads the dataset	21
Figure 9: User performing the topic modeling.	22
Figure 10: User search with keyword	22
Figure 11: Tweets associated with that keyword	23
Figure 12: User-provided themes	24
Figure 13: User merges themes.	24
Figure 14: User adds themes to tweets	25
Figure 15: User explores added themes	26
Figure 16: User explores visualization.	27

CHAPTER 1: INTRODUCTION

Social media plays a significant role in modern-day information dissemination, owing to its ubiquity and ease of sharing. There has been an increased focus on analyzing topics frequently occurring in social media. Interpretive researchers develop information from text corpora by manually collecting documents, applying codes, refining and collating codes into categories, and refining and collating codes into categories until relevant patterns emerge. Machine learning might assist scale this data gathering and analysis given a big corpus, but previous research reveals that professionals are usually worried about algorithms possibly disrupting or driving interpretative scholarship. Scholastic, which integrates a machine-in-the-loop clustering algorithm to scaffold interpretive text analysis, is built using a human-centered design approach to address concerns about machine-assisted interpretive research. When a researcher applies and refines codes to texts, the resulting coding schema serves as structured metadata, constraining hierarchical documents and word clusters derived from the corpus. Interactive representations of these clusters can assist academics in selectively sampling papers to get deeper insights. through interactive topic modeling and document grouping, human-centered algorithm design and visualizations using familiar metaphors may enhance inductive and interpretative research approaches. Since the emergence of "Big Data" research a few years ago, social media analytics has started to amass studies using social media as a source and instrument for study efforts. Yet, the creation of techniques for dealing with this sort of data has received very little attention, Even while this is becoming more widely acknowledged, using such data to analyze what it has to say about social life presents significant methodological hurdles. Social networking sites offer a type of user-generated data that can be uninvited and unscripted and that is frequently conveyed in several ways (i.e. through combinations of text, hyperlinks, images, videos, etc.). Therefore, it's crucial to take into account the difficulties that this data presents for researchers in terms of making them suitable for analysis and in figuring out the kinds of research issues that such data may usefully address[1]. To help users with social media data analysis, Social media is a massive social networking site tuned towards fast communication. More than 140 million active users publish over 400 million 140- character "Tweets" every day[2]. Twitter has become a vital communication tool for people from various fields because of its speed and simplicity of dissemination. Twitter has been heavily involved in socio-political events, such as forecasting election results and people's hesitation to get the COVID vaccination[3, p. 19]. During

significant natural disasters, Twitter has also been used to disseminate damage updates and information on disaster preparedness. The purpose of this thesis was to create a visual analytics online application that applies interactive visualization and computational knowledge discovery to social media data. The right balance is a collaborative approach, where automatic mechanisms assist humans in extracting and interpreting useful information. There is a scope for an Interactive Data Analysis tool as an iterative and interactive process that applies computational methods to understand data, refine questions, and cycle through the steps.

CHAPTER 2: BACKGROUND & LITERATURE REVIEW

It is observed that researchers who are interested in social media are mainly interested in performing a few main analysis tasks that include topic modeling, sentiment analysis, Emotion analysis, Hashtag analysis, and hate speech analysis. Topic modeling is a method for finding the main theme that includes a large and unstructured collection of documents that can arrange a dataset in accordance with the themes found in it. Topic modeling algorithms can be applied to large numbers of documents and can be adapted to various types of data, among applications, topic modeling has been used to find patterns in genetic data, and social networks[4]. Twitter is a well-known platform for microblogging where users may post status updates (called "tweets"). These tweets are filled with several examples of how people express their likes and dislikes as well as their contributions to various issues. Systems created by Sentiment Analysis aim to recognize and extract opinions from text[5]. Sometimes only sentiment analysis does not do the justice for the analysis task and in some cases, sentiment analysis is insufficient, necessitating emotion detection, which precisely ascertains a person's emotional/mental state. The degrees of sentiment analysis, different emotion models, and the method of sentiment analysis and emotion detection from the text are all explained in this review study[6]. Hashtags are a community-driven Twitter standard for providing tweets with more context and metadata. By simply prefixing a word or phrase with a hash symbol, such as "#hashtag," they are naturally formed by Twitter users as a tool to classify messages and highlight issues. The widespread usage of hashtags enhances Twitter's expressiveness and popularity among users. This study showed how sentiment analysis and hashtag analysis on Twitter have a lot of promise[7]. Also, there is many threat and hate speech components on social media and many times hate speech and threats are not controlled on social media for example on Twitter many times users post hate speech and it is important to analyze and observe them. Also a key component of linguistic and

legal threat assessment frameworks, but one that Twitter's Violent Threats policy mostly ignores, is how people interpret and respond to a tweet[8].

CHAPTER 3: IMPLEMENTATION

3.1 First Prototype

It was found that there is lots of interest in leveraging social media data for research in social science, urban planning, epidemiology, and other fields, but most researchers lack the programming skills necessary to use existing NLP toolkits. There is a scope of Interactive Data Analysis tools as an iterative and interactive process that applies computational methods to understand data, refine questions, and cycle through the steps. To start with the application the first prototype was designed to do several analysis tasks like:

- Sentiment analysis and other text classification tasks
- Topic Modelling
- Hashtag Analysis
- Emotion Analysis
- Hate Speech Analysis

And there are several Existing Python libraries for data analysis and natural language processing Transformers, SpaCy, Stanza, Flair, NLTK, etc. which were used to design and develop the App. The first prototype that was designed could only do Sentiment Analysis and Topic Modelling. The very first prototype was designed in such a way that users could bring the dataset and once the data is uploaded users can explore the dataset and understand what analysis tasks need to be performed, and perform Sentiment Analysis and Topic Modelling as shown in fig1.



Fig1: Shows the user journey of first prototype

Then for the Second prototype few more analysis tasks were added like hate speech, Emotion and Hashtag Analysis.And for the second prototype it was also further developed to help Researchers to download the results of the analysis tasks as shown in fig2.



Fig2: Shows the user journey of Second prototype

And for the Third prototype some explainability tasks were added to help researchers build some confidence on the pre-trained deep learning models used in the backend of the Application and pivot table feature was also added so that users could merge to or more analysis tasks if necessary for the users analysis as shown in fig3.



Fig3: Shows the user journey of Third prototype

Below is the image showing the screenshot of the web app developed as shown in fig4.

upload file

Twitter Data Analysis Tool

This application is a dashboard used to analyze sentiments, Hastag, Emotions of tweets and Topic Modelling

Drag and drop file here Limit 200MB per file • TXT, CSV	Browse files
data.csv 19.0MB	×
Add Analysis Tasks	+
Expand Me to see the Data	+
Expand Me to see the DataFrame and Visualize	+
Expand me to understand How to work with pivot table	+
Download Results as CSV	

Fig4: Shows the Final Prototype Developed for Interviews

After the development of the application we interviewed several participants to understand the requirement and see if there are further modifications needed in the application.

3.2 Evaluation of the application

Evaluation of the application was done in an interview study with four interpretive researchers. The study was divided into three phases: an overview of the objectives, a think-aloud analysis, and a follow-up interview to gather more input. All participants have experience with text-based interpretative research and have written about it. The data sampling and sensemaking loops of qualitative analysis are both integrated with a constraint-based interactive ML algorithm. Our design decisions were based on the idea that a successful tool for human-AI collaboration should be able to: 1) adding human input during data sampling and sensemaking, and 2) incorporating human inputs.

First, a short introduction was presented outlining the intention to investigate the ways in which a social scientist may utilize the application to study a sizable corpus in order to get input on its design. Then, get basic demographic data and informed consent for participation. Questions asked to the participants are added in (Appendix). Also, The following script served as an introduction to the researchers and a pertinent research question:

"We are seeking participants for a study examining the Social Media data analysis and Exploration tool. And your expertise in the social science & data analysis domain will be of great help in understanding the usability and development of the App. Participants are mainly from the social science and data analyst domain, and prior programming and coding experience are not required. If interested, participation would involve using a web-based application. In the analysis

tasks, participants will go through the application to analyze the social media data. While performing the tasks, the performance will be recorded. The study will be conducted remotely.

We will ask you to explore the application based on your understanding of the social media questions, which include questions about demographics. Based on your question which you may have about the insights on social media data, you can explore the application and see if you can get useful information from the data. After your exploration, we will be interviewing you. The interview will be recorded and will be divided into two parts. The first part focuses on background information, while the second part focuses on the usability of the app. Both parts of the interview will be conducted in one session which should take from 35 to 45 minutes. "

3.3 Results from the Interviews

After doing the interview with the participants I found that users wanted to search topics with their own keywords.

- Users were more interested in qualitative analysis.
- Users wanted to see the tweets in clusters and wanted to understand what makes them fall into a particular topic/cluster.
- One of the users wanted to understand the political polarization and is the reason the research now has to move forward to thematic analysis, because each user might have their own set of research questions.
- Also understand that users wanted to sample the dataset on different themes.

3.4 Literature Review: Thematic Analysis

What is thematic Analysis?

Thematic Analysis is a method for identifying and analyzing patterns in qualitative data. Since first being named as an approach in the 1970s [9], a number of different versions of thematic analysis have been proposed within psychology[10]. Thematic Analysis can be applied within a range of theoretical frameworks, from essentialist to constructionist; thematic discourse analysis.

Thematic Analysis can be used to analyze communications such as letters, memoranda, reports, or social media texts to identify the intentions of the communicators, to reveal the focus of the individual, communal, institutional, or societal attention, to describe trends in communication content and to examine attitudes, interests, and values.

Thematic Analysis is widely used in a range of research interests and theoretical perspectives and is useful because:

- It works with a wide range of research questions, from those about people's experiences or understandings to those about the representation and construction of particular phenomena in particular contexts.
- It can be used to analyze different types of data, from secondary sources such as media or transcripts of interviews.
- It works with large or small data sets.
- It can be applied to produce data-driven analyses.

There are six phases of thematic analysis as per [11].

1. Familiarization with the data: is common to all forms of qualitative analysis – the researcher must immerse themselves in, and become intimately familiar with, their data; reading and re-reading the data (and listening to audio-recorded data at least once, if relevant) and noting any initial analytic observations.

2. Coding: Also a common element of many approaches to qualitative analysis, this involves generating pithy labels for important features of the data relevant to the (broad) research question guiding the analysis. Coding is not simply a method of data reduction, it is also an analytic process, so codes capture both a semantic and conceptual reading of the data. The researcher codes every data item and ends this phase by collating all their codes and relevant data extracts.

3. Searching for themes: A theme is a coherent and meaningful pattern in the data relevant to the research question. If codes are the bricks and tiles in a brick-and-tile house, then themes are the walls and roof panels. Searching for themes is a bit like coding your codes to identify similarities in the data. This 'searching' is an active process; themes are not hidden in the data waiting to be discovered by the researcher, rather the researcher constructs themes. The researcher ends this phase by collating all the coded data relevant to each theme.

4. Reviewing themes: Involves checking that the themes 'work' in relation to both the coded extracts and the full data set. The researcher should reflect on whether the themes tell a convincing and compelling story about the data, and begin to define the nature of each individual theme, and the relationship between the themes. It may be necessary to collapse two themes together or to split a theme into two or more themes, or discard the candidate themes altogether and begin again the process of theme development.

5. Defining and naming themes: Requires the researcher to conduct and write a detailed analysis of each theme (the researcher should ask 'what story does this theme tell?' and 'how does this theme fit into the overall story about the data?'), identifying the 'essence' of each theme and constructing a concise, punchy, and informative name for each theme.

Why thematic analysis?

Because understanding human discourse is important (give examples of why) and requires in-depth qualitative analysis (give examples of good thematic analysis). The problem is that for social media and other large datasets, the data is too large so it's impossible for people to read everything. So an app that uses topic modeling and clustering of text or images to enable thematic analysis (cite topic-guided thematic analysis).

A few of the examples of thematic analysis that required in-depth qualitative analysis on social media are

1. Paper on "Examining consumer behavior in the UK Energy sector through the sentimental and thematic analysis of tweets" [11]: Consumer engagement with brands on social media has been empirically proven. However, little is known about consumers' natural behavior on social media, as literature on this topic is still in the early stage of its evolution. Accordingly, this study aims to investigate and understand the group interactions of consumer behavior, with a specific focus on tweets within the UK energy sector. This paper draws on social capital theory to investigate how UK consumers engage with their suppliers, as well as the knowledge-sharing capabilities of the Twitter community. The authors performed a qualitative insight into the factors shaping consumers' behavior as they engage with brands on social media. A thematic model emerges in the form of an interrelated conceptual theory comprising three stakeholders,

the key relationships between them, and their natural behaviors and offers a contemporary, essential, and interconnected understanding of consumer behavior online with a focus on the energy sector.

2. Another paper by Ozge Ozduzen "Enmeshing the mundane and the political: Twitter, LGBTI+ outing and macro-political polarization in Turkey"[12] did a thematic analysis of data collected from Twitter. She collected data collected on Twitter through hashtags and mentions, first overviewed the Tweets in Excel to grasp the broader picture, employing content analysis. She then coded a total of 10,030 Tweets for the purposes of content analysis in relation to sentiments towards LGBT+ identity and women to capture hate speech and solidarity for the individuals or the communities in question.

For our textual analysis(Thematic Analysis), she then used Nvivo with the following keywords Turk, morality, twisted, disgusting, pervert/perverted, Pride, marauder, terror, homeland, and go in order to code the data. She grouped Tweets with #Intizar hashtags and mentions them in four clusters: national identity, religion, political orientation, and sexual orientation (see Table 1). This allowed her to depict how Tweets sustain the dimensions of macro-political polarization inherent in Turkish society.

Sex Orie	ual entation	National Identity	Religious Belief	Political Orientation
• • • • •	Straight Normal vs Gay Lesbian Perverse Abnormal	 Proud of being a Turk Protecting pure Turkish identity against internal and external enemies vs Non-Turkish External 	 Muslim appropriate vs Non-Muslim (atheists, Alevi, Christians etc.) abnormal 	 Statist Supportive of the governing party AKP Vs <i>Çapulcu/marauder</i> Terrorist Marginal

Fig5: Shows the Themes developed for paper Enmeshing

3. An example of thematic analysis is from the paper "'**Refugees are not welcome': Digital racism, online placemaking and the evolving categorization of Syrians in Turkey**"[13] Ozge Ozduzen where she argues that digital publics unleash and bolster everyday racism, creating an unregulated space where anonymity and ubiquity enable the dissemination of racist messages. By creating broader visibility and wider reach of racist texts and facilitating more participation for racists, social media platforms such as Twitter normalize the gendered and place-based racialization of refugees. In this article, she identifies the common patterns and sentiments on the texts disseminated by actors on Twitter within meso-level social structures. To facilitate a thematic analysis, she used Excel to count and color code related words such as sexist swear words and words for location and place.

For her research work, she also argues that computational algorithms are incapable of reading data precisely the same way humans do. Computer-coding methods can efficiently deal with large data, but they have also been criticized for their ability to understand the subtle latent meanings of opinions expressed.

4. Another example of thematic analysis of by the authors of the paper "I Will Not Drink With You Today: A Topic-Guided Thematic Analysis of Addiction Recovery on Reddit"[14]: To perform our thematic analysis, authors first collected Reddit data from 2014 to 2017 for #stopdrinking and #OpiatesRecovery. then developed LDA topic models for each subreddit by iteratively cleaning, building, and evaluating the generated models, which produced purposive samples made up of keywords and threads for each topic. And finally performed the core of the thematic analysis by iteratively identifying codes, developing themes, and reviewing themes with the topics, keywords, and threads from LDA model.

Authors based thematic analysis on the 6 phases described by[11]. The familiarization phase occurred as part of the LDA topic modeling. Using what was learned from familiarization, Author then worked with the purposive samples to identify codes and to develop and review themes. Author would read each thread in its original context on Reddit to understand what data-driven codes were present then iteratively compared the threads and codes to develop themes.

Threads could contain multiple codes, codes could contribute to multiple themes, and thus threads could also contribute to multiple themes. Authors gathered supporting quotations for each theme from multiple threads that contributed supporting codes. Finally, the authors would review the themes by looking at additional threads from each subreddit to confirm that they were present and come to agreement that the themes fit the data.

Theme	Sub-Themes	Paraphrased Example Quote				
Social Relationships and Activities	Filling a Void	"Since I stopped drinking my life has felt empty. What do you suggest to handle the boredom?" » "You got to fill that emptiness! You can get started with anything you wanted to try in the past. It might be hard starting but motivation will come if you try."				
	Group Activities "During the movie club we chat while watching the linked str can."					
	Healthy Activities	"I picked up cross-training, biking, and running. I also found myself enjoying golf."				
	Leisure Activities	"If you're bored trying reading Game of Thrones. With five currently out and thousands of pages it will keep you occupied a while."				
Sharing Knowledge	Family and Friends	"tl;dr my older sibling is addicted to opiates and I have no idea how to help them."				
and Lived Experience	Managing Addiction Managing Consequences Managing Recovery Managing Withdrawal	"I know exercise sucks to get started during withdrawal but believe me when I say you will feel better after you finish. I read this is because of runner's high where the body releases endorphins which helps the opiate receptors feel less starved."				
	Understanding Addiction Understanding Recovery Understanding Withdrawal	"I can't understand how any amount of alcohol can be good for us as since I stopped drinking recently and I feel way healthier. Does anyone know if the claim the moderated drinking extends your life has been refuted by science? I can't believe this claim."				
Supporting Formal Treatment of	Female Support	"I want to find a group that has members who have common ground with me as a young female if possible."				
Addiction	Higher Power Concerns	"After going back to NA I remember why i stopped. I just cant figure out how to reconcile not believing win god with the 12-steps. I know if supposed to be spiritual not religious but i just have such a hard time with how that works in my head."				
	Newcomer Support	"As long as you go to an AA meeting you're at the best one. Any meeting should welcome you."				

Table 2: Example themes that show what the community members are discussing to support each other's recovery. Includes themes identified and paraphrased example quotations from the subreddits.

Fig6: Shows the Themes developed for paper I Will Not Drink With You today.

Looking at different examples one can conclude that performing Thematic analysis is a very manual and time-consuming process, so to help the researchers in topic-guided thematic analysis and so that they also would not have to manually annotate or define themes and save their time. And, also to understand the thematic analysis process, interviews with the researchers from the social science domain were recruited and asked about how they would perform thematic analysis and asked about the flow diagram and user journey.

3.5 Researchers' Interview and User Journey

After the interview with the researchers from the social science to understand how they would perform thematic analysis on dataset they were provided, in this case they were provided with the dataset from Twitter that has #blacklivesmatter,#vaccine hesitancy,and major steps that should be taken into consideration for the implementation were:

- They look for the effects present in the dataset.
- They try to find the motivation language present in the dataset.
- They use topic topic modelling to get a data driven sense of the whole dataset.
- They would like to explore specific topics sometimes if they find a topic interesting.
- Go through the most representative of tweets and see what might be going on.
- Look at each of the topics and look at some of the representative data units. Download the data units and rerun the analysis on the downloaded dataset.
- They would like to combine some of the topics if possible.
- How the conversation changes after certain events(e.g- put keywords and timeframe and how the tweets have changed after some events).



So considering these steps user journey can be mapped in the following way:

Fig7: Final user Journey.

3.6 Final Prototype

And then the application was developed to meet the user requirement for each of the steps.

1. User uploads the dataset.



Fig8: User uploads the dataset.

2. Once the dataset is uploaded they perform the topic modeling.

		BLM	Vaccine	people	mandates	- 1	Example View Explore & Visualize Write Themes
DATA IMPORT	۲	NYC	mandates	vaccine_mandates	nyc_ny	-	Scientists can lieBLM is a Marxist OrganizationPeople Xing the border illegally are illegalLove is Love unless you Love Trump apparentlyWomenå€"s rights are rights unless they donå€"t want the
Search for words	Q	blm	people	vaccine	vaccine_mandates	-	vaccineKindness is everything unless you have a dissenting opinion from the left
		pfizer	children	11	fda		trank for communism rascism socialism. open borders abortioni€F's and forced double mask and vaccine lawswith a booster every 3 monthsgod bless antifa bim george soros bill gates and the clintons
TOPIC LABEL TWEET LABEL	+	climate	crt_antifa	climate_change	change		please educate yourself on the definition of Marxism instead of BS of BS
Download file as CSV.	Ŀ	new	nyc	no vaccine mandates	new york		I was at this BLM communism rally and they were all like "Screw this fake vaccine" so now I believe all lives matter and fm a capitalist.
		maga	maga blm	bim maga	new york		Again the political party you support (Republican) fought harder to end mask mandates and vaccine mandates in public schooling than they did behade horizons in accented. Descenden warm BU M Kern More and the second
		racist	mandates racist	mandates	vaccine		thing? Yall made ALM to argue against Black rights.
		marxist	socialism	marxist blm	marxism	,	mandates and further oppression by the government? BLM should renounce their ties to communists. Freedom is the answer and you won't find it in the Democrat Party. Freedom? The Bill of Rights!
		radical	left wide	left wing	right wing		Aggression based-have Fun with that! Hey BLM, burning cities, vaccine mandates, open borderenjoy socialism —we all are made
		fox news	difference	spot difference fox	fox	,	poor. Mask and vaccine mandates are also Marxist tools.I'm against them just like I am against BLM.Screw the communist mindset. Well all be
		BIM	Vaccine	neonle	mandates		equal all right. Equally poor with zero way to lift yourself up, while the politicians and elite flourish and control us all.
				people			

Fig9: User performing the topic modeling.

3. User is interested in a topic with a particular keyword and they search for that with that

keyword.

		BLM	Vaccine	people	mandates		Example View Explore & Visualize Write The	mes
DATA IMPORT	۲	NYC	mandates	vaccine_mandates	nyc_ny	•	Scientists can lieBLM is a Marxist OrganizationPeople Xing the border illegally are illegalLove is Love unless you Love Trump apparentlyWomenå€"'s rights are rights unless they donå€"t want the	
left wing vaccine	Q	blm	people	vaccine	vaccine_mandates		vaccineKindness is everything unless you have a dissenting opinion from the left	
							thank for communism fascism socialism open borders abortionâ€"s and forced double mask and vaccine laws with a booster every 3 months god bless antifa blm george soros bill gates	
TOPIC LABEL	•	prizer	children	11	īda		and the clintons	
TWEET LABEL	•	climate	crt_antifa	climate_change	change	-	please educate yourself on the definition of Marxism instead of BS fear mongering. #facts #BLM	
Download file as CSV.	· •	new	nyc	no vaccine mandates	new york		I was at this BLM communism rally and they were all like "Screw this fake vaccine" so now I believe all lives matter and I'm a capitalist.	
		maga	maga blm	blm maga	new york	-	Again the political party you support (Republican) fought harder to end mask mandates and vaccine mandates in public schooling than they did school shootings in general.Remember when BLM first became a thing? Yall made ALM to argue against Black rights.	
		racist	mandates racist	mandates	vaccine		How is it that those connected to Marxism are against vaccine mandates and further oppression by the opperment? BLM should	
		manxist	socialism	marxist blm	marxism		renounce the first between the second of the government between our renounce the first co communists. Freedom is the answer and you won't find it in the Democrat Party. Freedom! The Bill of Rights!	-
		radical	left wide	left wing	right wing	1	Aggression based-have Fun with that! Hey BLM, burning cities, vaccine mandates, open borderenjoy socialism —we all are made poor.	
		fox news	difference	spot difference fox	fox		Mask and vaccine mandates are also Marxist tools. I'm against them just like I am against BLM.Screw the communist mindset. Well all be equal all right. Equally poor with zero way to lift yourself up, while the politicians and either flourish and control us all.	Ľ
		BLM	Vaccine	people	mandates			

Fig10: User search with keyword.



4. Users click on that topic to see the tweets associated with that keyword.

Fig11: Tweets associated with that keyword.

5. Users provide a label for that topic.

		BLM	Vaccine	people	mandates		Example View Explore & Visualize Write Themes
DATA IMPORT	۲	NYC	mandates	vaccine_mandates	nyc_ny		Scientists can lieBLM is a Marxist OrganizationPeople Xing the border illegally are illegalLove is Love unless you Love Trump apparent/Women&F** rights are rights unless they don&F**t want the
Search for words	Q						vaccineKindness is everything unless you have a dissenting opinion
		blm	people	vaccine	vaccine_mandates	- 1	from the left
							abortionã€ [™] s and forced double mask and vaccine laws with a
		pfizer	children	11	fda	1	booster every 3 months god bless antifa blm george soros bill gates and the clintons
TOPIC LABEL	•						
TWEET LABEL	►	climate	crt_antifa	climate_change	change		please educate yourself on the definition of Marxism instead of BS fear mongering. #facts #BLM
Download file as CSV.		new	nyc	no vaccine mandates	new york		I was at this BLM communism rally and they were all like "Screw this fake vaccine" so now I believe all lives matter and I'm a capitalist.
		maga	maga blm	bim maga	new york		Again the political party you support (Republican) fought harder to end mask mandates and vaccine mandates in public schooling than they did school shootings in general.Remember when BLM first became a
							thing? Y'all made ALM to argue against Black rights.
		racist	mandates racist	mandates	vaccine	1	How is it that those connected to Marxism are against vaccine
		Stigmatizing the left					renounce their ties to communists. Freedom is the answer and you
		marxist	socialism	marxist blm	marxism	1	won't find it in the Democrat Party. Freedom! The Bill of Rights!
		Stigmatizing the left					Assessing based have for with that the DLM huming sition
		radical	left wide	left wing	right wing	1	vaccine mandates, open borderenjoy socialism —we all are made
		fox news	difference	spot difference fox	fox	1	Mask and vaccine mandates are also Marxist tools. I'm against them just like I am against BLM.Screw the communist mindset. Well all be equal all right. Equally poor with zero way to lift yourself up, while the
							politicians and elite flourish and control us all.
		BLM	Vaccine	people	mandates	1	

Fig12: User provided themes.

 If users find that two topics are related to each other they provide that other topic a same label or assign a theme to that topic and merge them together.

		BLM	Vaccine	people	mandates	/	Example View Explore & Visualize Write The	emes
DATA IMPORT	۵	NVC	mandates	varcine mandates	DHC DV		Scientists can lieBLM is a Marxist OrganizationPeople Xing the border	
		NIC	manuates	vaccine_manuales	liye_liy	-	apparentlyWomenâ€ [™] s rights are rights unless they donâ€ [™] t want the	
Search for words	Q						vaccineKindness is everything unless you have a dissenting opinion from the left	
		blm	people	vaccine	vaccine_mandates	1		
							thank for communism fascism socialism open borders	
							abortionat "s and forced double mask and vaccine laws with a booster every 3 months god bless antifa blm george soros bill gates	
		pfizer	children	11	fda		and the clintons	
TOPIC LABEL	•							
TWEET I ADEI	•	climate	ort antifa	climate change	change		please educate yourself on the definition of Marxism instead of BS fear mongering #facts #BLM	
THEET DOLL				china ta go na ngo	chunge		······································	
Download file as CSV.	I ^{†I}	vaccine mandatory in	NYC					-
		new	nyc	no vaccine mandates	new york	/	I was at this BLM communism rally and they were all like "Screw this fake vaccine" so now I believe all lives matter and I'm a capitalist.	
							·····	
							Again the political party you support (Republican) fought harder to end	
		maga	maga blm	bim maga	new york	· · · ·	mask mandates and vaccine mandates in public schooling than they did school shootings in general Remember when RLM first became a	
		vaccine mandate targe	ets peoples of different	race			thing? Y'all made ALM to argue against Black rights.	
		racist	mandates racist	mandates	vaccine	- 1		
							How is it that those connected to Marxism are against vaccine mandates and further oppression by the opvernment? BLM should	□
		Stigmatizing the left					renounce their ties to communists. Freedom is the answer and you	
		marxist	socialism	marxist blm	marxism	1	won't find it in the Democrat Party. Freedom: The Bill of Rights:	
						_		
		and and	1.6.11	1.6	1.14 A.		Aggression based-have Fun with that! Hey BLM, burning cities,	Гľ
		radical	left wide	left wing	ngnt wing	1	poor.	
		fox news	difference	spot difference fox	fox	1	Mask and vaccine mandates are also Marxist tools. I'm against them just like I am against BLM.Screw the communist mindset. Well all be	Γℓ
							equal all right. Equally poor with zero way to lift yourself up, while the politicians and elite flourish and control us all	
							personano ano ence novinari ana colliti di tibiliti.	
		BLM	Vaccine	people	mandates			

Fig13: User merges themes.

7. Users add more themes by first exploring the topics and finding tweets that have similar themes to each other.

	BLM	Vaccine	people	mandates	1	Example View	Explore & Visualize	Write The	emes
DATA IMPORT C	NYC	mandates	vaccine_mandates	nyc_ny		fake propaganda again Scientists can lieBLM is a illegally are illegalLove is apparentlyWomenâ€"s ri vaccineKindness is every from the loft	st bim Marxist OrganizationPeople X Love unless you Love Trump ights are rights unless they don thing unless you have a dissent	ing the border 't want the ting opinion	
	blm	people	vaccine	vaccine_mandates		rebutting blm marxism	frame		
	pfizer	children	11	fda		thank for communism fa abortionâ€ [™] s and forced booster every 3 months . and the clintons	scism socialism open border double mask and vaccine laws . god bless antifa blm george si	s with a pros bill gates	Ľ
TOPIC LABEL Stigmatizing the left vaccine mandate targets peoples of	climate	crt_antifa	climate_change	change	,	rebutting blm marxism please educate yourself of fear mongering. #facts #	frame on the definition of Marxism ins BLM	tead of BS	
different race vaccine mandatory in NYC	vaccine mandatory in new	NYC	no vaccine mandates	new york		misinformation against I was at this BLM commu fake vaccine" so now I be	rally inism rally and they were all like lieve all lives matter and I'm a c	Screw this apitalist.	
Download file as CSV.	maga	maga blm	blm maga	new york		tying blm with political Again the political party y mask mandates and vace did school shootings in g	polarization ou support (Republican) fough cine mandates in public school eneral.Remember when BLM fi	t harder to end ng than they rst became a	
	vaccine mandate targ	ets peoples of different mandates racist	race mandates	vaccine	,	thing? Yall made ALM to tying blm with marxism How is it that those conn mandates and further op	argue against Black rights. ected to Marxism are against v pression by the government? Bl	accine LM should	⊡
	Stigmatizing the left marxist	socialism	marxist blm	marxism		renounce their ties to con won't find it in the Democ	nmunists. Freedom is the answ rat Party. Freedom! The Bill of F	er and you Rights!	
	radical	left wide	left wing	right wing		Aggression based-have l vaccine mandates, open poor.	Fun with that! Hey BLM, burning borderenjoy socialism —we a	g cities, all are made	Ľ
	fox news	difference	spot difference fox	fox	•	tying blm with marxism Mask and vaccine manda just like I am against BLM equal all right. Equally po politicians and elite flouri	ates are also Marxist tools.I'm a 4.Screw the communist mindse or with zero way to lift yourself sh and control us all.	gainst them t. Well all be up, while the	
	BLM	Vaccine	people	mandates	1				

Fig14: User adds themes to tweets.

8. Users can also explore to see what themes they have added.

		BLM	Vaccine	people	mandates		Example View Explore & Visualize Write Themes	
DATA IMPORT	•	NYC	mandates	vaccine_mandates	nyc_ny	,	Fake propaganda against bim Scientists can lieBLM is a Marxist OrganizationPeople Xing the border lilegallay are lilegalLaye is Low unless you Love Trump apparentlyWomenáč s rights unless they donáč "t want the underski Menera is avenditien unless weit bave a diversitien contain.	
Search for words		blm	people	vaccine	vaccine_mandates	•	thank for communities a betty limit unless you have a ussenting opinion rebutting bim marxism frame thank for communism fascism socialism open borders abortiona ^{CP} is and forced double mask and vaccine laws with a	P
	(pfizer	children	11	fda		booster every 3 months god bless antifa blm george soros bill gates and the clintons	_
TOPIC LABEL Stigmatizing the left vaccine mandate targets peoples of	•	climate	crt_antifa	climate_change	change		rebutting bim marxism frame please educate yourself on the definition of Marxism instead of BS fear mongering. #facts #BLM	
different race vaccine mandatory in NYC		vaccine mandatory in new	nyc	no vaccine mandates	new york		misinformation against rally I was at this BLM communism rally and they were all like "Screw this fake vaccine" so now I believe all lives matter and I'm a capitalist.	
TWEET LABEL rebutting blm marxism frame tying blm with marxism	•	maga	maga blm	blm maga	new york		tying bim with political polarization Again the political party you support (Republican) fought harder to end mask mandates and vaccine mandates in public schooling than they did school shootings in general.Remember when BLM first became a	ſ
fake propaganda against blm misinformation against rally		vaccine mandate targ	jets peoples of different mandates racist	mandates	vaccine		thing? Yall made ALM to argue against Black rights.	P
Download file as CSV ∟		Stigmatizing the left marxist	socialism	marxist blm	marxism	•	renounce their ties to communists. Freedom is the answer and you won't find it in the Democrat Party. Freedom! The Bill of Rights!	
		radical	left wide	left wing	right wing		Aggression based-have Fun with that! Hey BLM, burning cities, vaccine mandates, open borderenjoy socialism —we all are made poor.	
		fox news	difference	spot difference fox	fox	•	Ving bim with markin Mask and vaccine mandates are also Markist tools I'm against them just like I am against BLM. Screw the communist mindset. Well all be equal all right. Equally poor with zero way to lift yourself up, while the politicians and elife flourish and control us all.	ſ
		BLM	Vaccine	people	mandates	- 1		

Fig15: User explores added themes.

9. Users can explore the dataset and search for the tweets and also explore the visualization if needed.

	BLM	Vaccine	people	mandates	,	Example View Explore & Visualize Write Themes
DATA IMPORT						orign_text Clean_text user_name location public_metric topic_words
	NYC	mandates	vaccine_mandates	nyc_ny	1	there's a lot less police in places like NYC type tables in every tense standoff, Bi4() there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type there's a lot less police in places like NYC type
Search for words Q						Consider how much you dislike vaccine mandates. Now consider how much stuff the government & Comandates& for us, starting with BEE, E.B.M. Jobal SurvAR, the SABC and various other aspects of the & Contaction Democratic Revolution. &
	blm	neonle	vaccine	vaccine mandates		If democrats arent destroying America explain BLM, Critical Race Theory, Horomone blocking therapies in children, vaccine mandates & amp open border policies (list goes on & amp; on). If youre a D & amp; think youre on the side of the good gays, youre either in denial or will fifty
		people	Haddine	Taconic_mandateo		u uninoricaly support BLM. Lmao that reflects way worse on you. They're fucking laughing at you. Because you're a suckerActually the vaccine has shown to do more harm than good. Vaccine has lead to many spikes in cases in many countries.
						Keep focused on the real news: Ohislaine Maxwell trial (Clinton, Bill Gates) BLM terrorist kills 6, injured 48Raising inflationBideris failed OSHA vaccine mandateFauci tortures animals in labsBideris failing PresidencyIbeinglibertarian
	pfizer	children	11	fda	/	Not taking the vaccine is not a reason to go to prison. But if it becomes one, then people will quickly decide to give them a reason a real reason.Anti/a? BLM?Histos.What will happen when intelligent, affluent people decide to rist and destroy?
TOPIC LABEL						Advyourself this, If Corbyn and Labour had won in 2019 how many of those who have refused the vaccination would be sitting here fully jabbed. As with BLM, IB, ER and climate activists, the arti covid vaccine is a movement which is propelled by arti Tory feeling.
Stigmatizing the left	olimato	ort optifo	olimato obango	change		Ves the SJW climate in the furry fandom is bod but I see some very bistantly conservative furrys on here who are followed by and interact with people who have BLM, vaccine status etc on their profiles. In many cases (but "certainty" not all), drama lamas create their own problems
vaccine mandate targets peoples of	climate	cr_antila	climate_change	change		Processed 10 of 39,205 input rows
different race	vaccine mandatory i	n NYC				02
vaccine mandatory in NTC	new	nyc	no vaccine mandates	new york	/	0 0
TWEET LABEL						
Download file as CSV	maga	maga blm	blm maga	new york		
	vaccine mandate tar	gets peoples of different	race			
	racist	mandates racist	mandates	vaccine	1	© © 0
	Othersee Markey Markey 1-6					© (0)
	Stigmatizing the left					
	marxist	socialism	marxist blm	marxism		
						stigmatizing the left wing
	radical	left wide	left wing	right wing	1	blacklivesmatter_covid_19
						• • • • • •
						• •
	fox news	difference	spot difference fox	fox		the blm vaccine
	BLM	Vaccine	people	mandates	1	kayle is bim



CHAPTER 4 : RESULTS FROM THE STUDY

To evaluate the system, we interviewed experts from social science who focus on the problem of theme generation and analysis, each who provided different feedback:

When we are dealing with a complex set of textual data, we need to pay attention to each of the clusters/topics as well as the intentions of such producers of tweets/data. To better deal with this issue, our interviewees recommended allowing them to make modifications on the systems encoding. Our interviewees wanted to do the data cleaning process themselves. Define their own stopwords or baggerwords. Interviewees found our application useful as it allows them to go both top down and bottom up approach. They found topic modelling helped them in getting an overall idea of what's in the dataset helped them explore their research questions.

CHAPTER 5 : DISCUSSION AND FUTURE WORKS

This research aimed to explore the potential of topic-guided thematic analysis with the help of an interactive app. Presently this system only supports data generated from Twitter and we would like to expand this application to all forms of textual information. This application also doesn't support doing the data cleaning and removal of stopwords if the user wants to add/remove stopwords themselves or do the data cleaning. Presently this application works good for textual information but doesn't support any other forms of information, I would like to further expand this application to support image data as well. Interviewees struggled at the beginning with how to use the application where I had to explain to them what each function would do. I would like to tackle this barrier so that the application is self-explanatory and explainable. I noted that this tool does not aim to incorporate all of the coding features present in commercial tools like Nvivo/taguette that primarily support code management.

If a user wishes to apply multiple codes to the same topic/tweets in this current implementation, the tweets/topics must be segmented into unique keywords for each code to avoid overlapping assignments. For assigning multiple codes to a single passage for approaches treating text data as a bag-of-words input will require novel algorithmic support, unless the word tokens can be separated according to additional metadata such as semantics and syntax. Visualization plays a significant role in researchers' desire to use visualization for sampling data items. For example, one of the users noted that the tree mapping after merging topics was not there in the App. Future work will study how visual features emerging in cluster visualizations may play a role in learning and memory for qualitative data.

Even in the absence of explicit visual summaries, people were able to leverage, conceptualize and remember the information contained within clusters and guide their exploration. Future work will explore if this behavior has a limitation.

CHAPTER 6 : CONCLUSION

This application explores the application of computational techniques to support qualitative research. I tried to show how the complementary use of computational and qualitative techniques, and the use of topic modeling could be used to purposively sample for thematic analysis, and can yield insights into the types of discussions occurring in social media at scale, and allow human researchers/social scientists to more deeply engage with those discussions.

I expect to see in depth integration of qualitative research and computing moving forward, and I think this work can serve as one model of partnership between human and machine-guided analysis techniques.

BIBLIOGRAPHY:

- [1] Phillip Brooker, Julie Barnett1, and Timothy Cribbin, "Doing social media analytics".
- [2] "social media users", [Online]. Available: https://financesonline.com/number-of-twitter-users/
- [3] "COVID-19 Vaccine Hesitancy on Social Media: Building a Public Twitter Data Set of Antivaccine Content, Vaccine Misinformation, and Conspiracies", doi: https://publichealth.jmir.org/2021/11/e30642.
- [4] Edi Surya Negara, Dendi Triadi, and Ria Andryani, "Topic Modelling Twitter Data with Latent Dirichlet Allocation Method", doi: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8984523&tag=1.
- [5] Amrita Shela and Ching-yu Huang, "Sentiment Analysis of Twitter Data", doi: https://ieeexplore.ieee.org/stamp.jsp?arnumber=8947771&tag=1.
- [6] "A review on sentiment analysis and emotion detection from text", doi: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8402961/.
- [7] "Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach", doi: https://dl.acm.org/doi/pdf/10.1145/2063576.2063726.
- [8] Pooja Casula and Nassim Parvin, "We found no violation!': Twitter's Violent Threats Policy and Toxicity in Online Discourse", doi: https://dl.acm.org/doi/pdf/10.1145/3461564.3461589.
- [9]. Merton, R. K. (1975) Thematic analysis in science: notes on Holton's concept. Science as Culture, 188(4186), 335-338.
- [10].Aronson, J. (1994). A pragmatic view of thematic analysis. The Qualitative Report, 2(1), http://www.nova.edu/ssss/QR/BackIssues/QR2-1/aronson.html.
- [11].Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. Qualitative Research in Psychology, 3, 77-101.
- [12].https://onlinelibrary.wiley.com/doi/epdf/10.1002/cb.1820
- [13].https://www.tandfonline.com/doi/pdf/10.1080/13569775.2020.1759883?needAccess=true
- [14].https://journals.sagepub.com/doi/pdf/10.1177/1461444820956341
- [15].https://dl.acm.org/doi/pdf/10.1145/3491102.3502076

APPENDIX:

Background information/questions?

1. What's the user experience with the data analysis tools?

If they analyze social media data?

2. What are users' experiences with using software applications for analyzing data?

3. How frequently do they use software applications for data analysis and how much time do they usually spend on getting useful information out of the application?

4. A brief summary of what they understand about social media (Think aloud method)

Questions about the application

1. How was their experience navigating through the application, and are there any particular challenges they faced?

2. Does the application allow the investigation of important questions participants may have?

a. Are there any additional analyses participants were looking for?

3. We would also ask the participants to compare it with the existing methods or tools participants use currently for the exploratory data analysis.

a. Is it more extensive?

b. Does it allow for deeper investigation?

4. We would like to hear from the participants if this application could be useful in their research and how we could improve it to meet the needs of the participants.

5. We would ask participants how the visualization in the application affects their understanding of the data?

6. How much of your data exploration/analysis/processing do you do locally (i.e. on your computer or one you're physically accessing) versus remotely (on AWS, CoLab, etc.)?

7. Do you work with someone who does the coding?