Detecting Political Framing Shifts and the Adversarial Phrases within Rival Factions and Ranking Temporal Snapshot Contents in Social Media

by

Sultan Alzahrani

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2018 by the
Graduate Supervisory Committee:

Hasan Davulcu, Chair
Steven R. Corman
Baoxin Li
Ihan Hsiao

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

Social Computing is an area of computer science concerned with dynamics of communities and cultures, created through computer-mediated social interaction. Various social media platforms, such as social network services and microblogging, enable users to come together and create social movements expressing their opinions on diverse sets of issues, events, complaints, grievances, and goals. Methods for monitoring and summarizing these types of sociopolitical trends, its leaders and followers, messages, and dynamics are needed. In this dissertation, a framework comprising of community and content-based computational methods is presented to provide insights for multilingual and noisy political social media content. First, a model is developed to predict the emergence of viral hashtag breakouts, using network features. Next, another model is developed to detect and compare individual and organizational accounts, by using a set of domain and language-independent features. The third model exposes contentious issues, driving reactionary dynamics between opposing camps. The fourth model develops community detection and visualization methods to reveal underlying dynamics and key messages that drive dynamics. The final model presents a use case methodology for detecting and monitoring foreign influence, wherein a state actor and news media under its control attempt to shift public opinion by framing information to support multiple adversarial narratives that facilitate their goals. In each case, a discussion of novel aspects and contributions of the models is presented, as well as quantitative and qualitative evaluations. An analysis of multiple conflict situations will be conducted, covering areas in the UK, Bangladesh, Libya and the Ukraine where adversarial framing lead to polarization, declines in social cohesion, social unrest, and even civil wars (e.g., Libya and the Ukraine).

To my parents and my wife for their endless love, support and encouragement, without whom none of my success would be possible.

ACKNOWLEDGMENTS

I would like to express my gratitude to my research advisor Prof. Hasan Davulcu for his thoughtful guidance, funding support and encouragement during my Ph.D. journey, without whom, none of my work would have been possible.

I also would like to express my gratitude to Prof. Corman for his enthusiasm that has been rendered throughout various cutting-edge projects we worked on.

I thank my committee members Prof. Hsiao and Prof. Li for their feedback on the work presented in this dissertation, and ealier works: prospectus proposal and related oral presentations.

I am also thnakful for King Abdulaziz City for Science and Technology, the Ministry of Higher Education in Saudi Arabia represented in the Saudi Arabian Cultural Mission, and the Royal Embassy in the USA for providing me with Master's and Ph.D's fellowships.

I am also thankful for all CIPS members: Mert, Amin, Anvesh Reddy, Chinmay, Naveen and Nyunsu for their help, constructive discussions and valuable feedback.

Last and foremost, I sincerely thank my wife, Fatimah, for giving me unfailing support and encouragement during compilation of my dissertation. It was her unwavering enthusiasm and motivation that provided me with the eagerness to complete this work. I owe you a lot of time!

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION:

Since the emergence of humankind, groups of peoples who know each other have been milling around in their social circles and discussing various topics, ranging from day-to-day issues to long-term concerns about society, economy, and politics. Their forms of engagement in rational discussion set an example of how members influence others to adopt certain attitudes, and shape their nuanced opinions with narrower interpretation of the issues at hand. As this dynamic impact becomes recurrent more over time, members can become further polarized especially when they interact with those of a similar view and mobilized when exposed to messages that are aligned with a political agenda. [96]

Over time and with the advent of letterpress, radio and TV stations enlarged the public sphere discussion. Such technologies from 1800 to 1980s played a significant role in maximizing speed and volume of information dissemination in their reach of larger audiences, as well as infiltrating community barriers which were previously considered too far or unreachable. These technologies were sophisticated tools that delivered targeted messages with political agendas. Examples in history include recruitment and collective actions, both of which were fueled by peoples grievances and their desire for change. For instance, the civil rights movement of the 1960s was incited by radio and letterpress; it voiced the grievances of the black community, thus culminating into massive social changes in the US regarding ethnic minorities and the less fortunate. Another example is the feminist movement of the 1960s to the early 1980s, which resulted in womens right to vote under the Voting Right Act of 1965 and the prohibiting of discrimination against genders and minorities. That movement later achieved many changes in regards equality rights, and fought against female subordination and male supremacy ideologies in all US institutions and in society as a whole. There are many other similar stories wherein traditional media served as cornerstone in human mobility around the world. [46, 84]

With the early advent of the internet and the World Wide Web (WWW) in the late 1980s, and subsequent growth and spread in 1990s, Web 1.0 has allowed news media to ability to publish limitless content to digital sphere that is accessible anytime and anywhere. At the dawn of the new millennium, Web 2.0 3.0 [4, 64] introdued social media sites wherein ordinary people became the focal point and took primary charge of creating content at any given moment. Many microblogging sites, such as Twitter and Facebook and other similar Social Media Sites (SMS), worked to grab their users attention while actively appealing to their engagement and actions (e.g., posting, sharing, replying, liking, etc.).[121]

The aforementioned SMS provide a wealth of possibilities for different research in human behavior interaction, known as collective intelligence. This form of intelligence attempts to utilize these interactions possibilities to provide clues to subtle questions analytically and empirically concerning problems to various spheres of humanity where answers emerge after studying these related historical interactions. Enormous amounts of interaction between users from most social media sites are saved in data warehouses for later use by query contents based on a set of keywords or other attributes, such as GEO location, timestamp, users' profile textual contents reflecting account/personality traits. This resulted in various possible analytical application and research hypothesis, such as descriptive analytics which can provide insight and offer forecasting perspectives. Such applications can be applied to various contexts, including those in the political, socio-political and communication arena all of which are the core focus of this dissertation. Other contexts include marketing and target marketing, which aims to improve consumer satisfaction through the study of sentiments and demands which profitability is imminent by offering right service for largest segments of consumers and for leading new markets discoveries.[64]

From a political perspective, collective actions can be defined as actions taken together by a group of people whose goal is to achieve a mutual objective, and can be employed for benevolent and malevolent goals [96, 46]. Many studies found that collective actions are highly motivated by grievance and the demand for change, wherein masses organize themselves to achieve a common goal. and who lead these efforts are an influencer and play actors who employee crowd, share fear and hope to achieve predetermined goals [84]. In social media, a collective action is much easier to organize and framed with rhetoric devices to serve its propaganda. Even though in many historical eras many society struggled in conflicts and sought for higher human rights standards, yet now many movements can be turned into a dangerous soft weaponized instrument- other wording "cyberwarfare"[62] promoting social unrest in a deliberate hidden and deceivable way to the ordinary people. Such as a mass mobilization can be used by exogenous powers to undermine cohesion among members of the targeted society, under the guise of rights and freedoms. An example of this is the Arab Spring in Libya that lead to civil war and its complex political landscape with lack of safety and security in major areas. Basic human rights further deteriorated, despite the momentary euphoria that came with the fall of the government in 2011 and gave the illusion of full democratic transformation [79, 60]. Another example are the Ukraine conflicts instigated by Russia, with its goal of reclaiming Ukraine to the Kremlin orbit and of impeding the influence of countries part of NATO. Russias use of energy war [131]]. Russias use of energy war [10] and cyber warfare [93] as leverage resulted in a civil war, the Crimea annexation and a full invasion of Ukraines southwest region.

1.1 Objectives

The nature of disseminated information and the main players driving mobilization is the topic of this study. The interaction between users from varying ideologies is crucial for studying socio-politics and movements, the adversarial language among rival groups which polarize the conflict and the main players who influence the political landscape and the masses. These can be quite challenging to study as there is an enormous amount of data and relational interactions reflecting the structure of societal allegiance . However, with the use of analytical and empirical methods guided by social computation, an examination is possible, and findings can be concluded. The impact of increasing social unrest and the promotion of hate speech creates a deceptive fertile ground and charged environment, which trigger and exacerbate grievances for mass mobilization that aims to dismantle social cohesion in various regions. Such aggressive means can be achieved especially when societies suffer from severe inequality and economic distress, which may ignite political chaos and violence.

Social computing uses experimental research design. It involves creating social conventions and contexts by collecting, representing, preprocessing, and spreading of information. It is followed by the use of computational and machine learning approaches, such as community detection, classification, prediction, clustering, etc. [78, 42].In this dissertation, we determined how to find trending or breaking information at nascent stages through hashtags. Next, they search organizational accounts that act as rhetorical devices for political organizational narratives and propaganda. This work detects and studies contentious language between rival groups dominating the political landscape, including phrases that elicit reactions between them, thereby providing an overview of current controversies and points of convergence and divergence. It also presents how users attention shift according to events and significant developments, and what this can be reveals about users affiliation to political factions where their sentiments for various looming issues can be revealed and explained [112].Moreover, a visualization tool is proposed, which can enhance the ability of Subject Matter Experts in understanding events and users shift in a complex political scene with high speed while avoiding the need to sift through heaps of social interaction contents. Finally, an empirical study is presented and examines how exogenous powers promote social unrest among minorities for political gain, expansionist ambitions and the fight over ideologies believed to threaten their hegemony.

1.2 Dissertation Structure

The dissertation is organized as follows:

Chapter 2- Information Volume breakout in Social Media and Identifying Proper Running Window Size: It proposes a forecasting model using content independent features to predict when a hashtag will become trending or die out based on adaptive volume monitor that can be accustomed to each hashtag regardless of the significance (i.e. hashtag's tweets volume) indicated by users attention.

Chapter 3- Finding Organizational Accounts Based on Structural and Behavioral Factors on Twitter: The study of organizational accounts recently gained more traction as it can disseminate narratives aligned with ideologies. These accounts have also become more active in voting seasons and in places where severe political turmoil is present, as they can potentially increase social unrest or promote hate speech, their political agenda and narrow interests.

Chapter 4- Identifying Ideologies Frame of Adversarial Political Groups:

This chapter explains the classification of users and posts in social media based on ideologies. Furthermore, it demonstrates how to visualize phrasal ideology flow after mapping posts word level into social cohesion dimension as dictionary-based approach where each dimension has optimistic positive polar and negative pessimistic. This is followed by the examination of the ideology flow from Facebook pages that promote the political agendas of certain groups. We were able to gauge commentators posts based on perceived ideologies as well as social cohesion dimension, and liken it to Twitter posts where they represent mass of users followers. In this chapter, findings on the behavior of users across all ideologies are explained, as Facebook represents the source of messages and its effects are reflected into Twitter. Many examples of discovered patterns will be discussed and linked back to already published reports that shed light on these phenomena.

Chapter 5- Detecting Adversarial Political Phrasal and Classify them into Contentious Non-Contentious: This chapter provides an approach used for detecting phrases among two main factions in the political landscape. By using an empirical approach, we found the phrases that likely elicit responses by the opposing group and the other phrases even continuous, but it does not.

Chapter 6- Community Temporal Clusters and Ranking: It shows how users and the keywords they use in their posts can be clustered temporarily. This process recurs as new events happen and significant development arises in the political arena. Also, sentiment analysis has been applied to identical users inclination towards a group faction and how their attention diverges and emerges against each snapshot. This chapter also aimed to increase the subject matter experts understanding when sifting over contents to try and make meaning out of contents by providing a rank per post, which helped supplement the view snapshot clusters per events.

Chapter 7- Framing Shifts of the Ukraine Conflict in pro-Russian News Media: This chapter showed an empirical study examining exogenous power (i.e., Russia) and how it promotes social unrest among minorities for political gain, expansionist ambitions and the fight over unnecessary ideologies believed to be a threat to their hegemony over neighboring countries particularly Ukraine, and we showed that how frameshift through divergence computation of frames codebook categories can forecast the onset of hostility that can lead to annexations or even invasions.

Chapter 8- Dissertation Conclusion: This chapter concludes the dissertation and presents the future work.

Chapter 2

INFORMATION VOLUME BREAKOUT IN SOCIAL MEDIA AND IDENTIFYING PROPER RUNNING WINDOW SIZE

2.1 Preface

Online information propagates differently on the web, some of which can be viral [1] . In this chapter, a simple standard deviation sigma levels based Tweet volume breakout definition is introduced, and patterns of re-tweet network measures to predict whether a hashtag volume will breakout are then determined. Motivated by Explotary Data Analysis (EDA) with both numerical summarization and data visualization, this study develops a trend-detecting method and a visual tool to help trace the evolution of hashtag volumes, their underlying networks and both local and global network measures. Moreover, a random forest tree classifier was trained to identify effective network measures for predicting hashtag volume breakouts. This studys experiments showed that local network features, based on a fixed-sized sliding window, have an overall predictive accuracy of 76%, whereas the overall predictive accuracy of a sliding window based breakout predictor jumps to 83% when global features that utilize all interactions up to the current period are incorporated. This study also determined the suitable fixed window size in a running mean as well as a standard deviation for whole hashtags set for the UK region as of 20 days after it provides a high classification rate among several various days. This approach can be leveraged when examining frames and temporal clustering.

[1]This chapter work has been submitted and accepted as a conference paper [15]

2.2 Chapter Introduction

Online Social Networks (OSNs) such as Twitter have emerged as popular microblogging and interactive platforms for information sharing among people. Twitter provides a suitable platform for the investigation of properties in information diusion. Diusion analysis can be applied to social media to examine viral tweets and trending hashtags, and to propose early-warning solutions that can signal if a viral hashtag started developing in its beginning stages. In this work, the 68-95-99.7 rule [108] is used to dene a simple method in hashtag volume breakouts. In statistics, the 68-95-99.7 rule, also known as the three-sigma rule, states that nearly all values lie within three standard deviations (σ) of the mean (μ) in a normal distribution. A xed sized sliding window wd (of predened d) was used to compute a running average and standard deviation for each hashtag's volume distribution. Subsequently, non-overlapping episodes were defined within a time-series of daily volumes for each hashtag whenever its daily volume exceeded ($\mu + 1\sigma$) the previous d day periods. The d day periods preceding an episode were labelled as the accumulation period of an episode. If the hashtag volume went on to exceed ($\mu + 2\sigma$) without falling below ($0,\mu - \sigma$), the episode was categorized as breaking, and as a non-breaking episode otherwise. Next, multiple network metrics associated with the accumulation period of each episode were examined, and a classier that aimed to predict if an episode will or will not lead to a breakout volume was built. A network-based classication model was employed and the contributing factors that have resulted in the breakout of hashtag volumes was examined to identify latent patterns for the breakout phenomena. A visualization tool was also designed Trending Hashtag Forecaster (THF) that reveals the underlying network structures, patterns and properties that lead to breakout volumes, as well as its spatio-temporal contents shown in the visualization.

However, the experiments have shown that "local" network features during a period of accumulation have an overall predictive accuracy of 76%, whereas the incorporation of global features that use measures extracted from the whole network up to the current accumulation period has the overall predictive accuracy of 83%, according to Trending Hashtag Forecaster.

2.3 Hashtags, Retweeting, Mentioning, and Reply

Twitter is well-known as a microbiological platform wherein users share content, media and opinions. Anyone who wishes to share can create a new hashtag label or participate on an existing one to interact with topics that identify such topic of interest where other people may adopt and can create new variation of hashtags labeling topic that may include different sentiments . Hashtags can be both general or specic in describing a topic of concern; thus, few of them eectively help in propagating information among communities, as they can carry different sentiments and offer disparate opinions and perspectives [87, 134]. The goal of this study is to capture such hashtags with pre-dened keywords embedded within Twitter crawler, as these would lead to relevant sets of tweets for analysis from subject matter experts (SMEs). SMEs will subsequently make inferences, filter through propagating topics and hashtags, relate them to the heat map, and observe user networks. Twitters features, such as re-tweeting, mentioning and replying, have been exploited to observe users relations and interaction, as well as provide insight regarding communities. In this work, the reply, mentioning, and re-tweeting functions of Twitter were used to construct a social network into a graph $G(V, E)$, dened by E edges (relations) and V vertices (users). to distill dened centrality measure. Section 2.7.1, to build feature space matrix based on user interaction and to be included in Vector Space Model (VSM) for the classication problem.

2.4 Problem Formulation

Given a set of tweets $T = t_1, t_2, t_3, ..., t_n$ where n is number of tweets in our corpus. These tweets comprise textual contents, user interactions and additional meta data. We explore and analyze both textual contents filtered by a given hashtag from hashtags set H. Then we denote tweet volume as number of tweets per day. We then compute daily means ($\mu(w_d)$) and standard deviation ($\sigma(w_d)$) for each hashtag by utilizing its volume distribution during its previous w_d days window. We experimentally determined the best window size by experimenting all $d \in \{2, 3, \ldots, 30\}$ days windows while we training a classifier. Then we select the suitable window size that woudlbe accounted for maximum accuracy .

If the hashtag frequency rises above ($\mu(w_{size}) + 1\sigma(w_{size})$), then we label that period as an episode, and we mark its previous d days (i.e. w_d size as the accumulation period of an episode. We start observing a hashtag frequency for two possible outcomes:

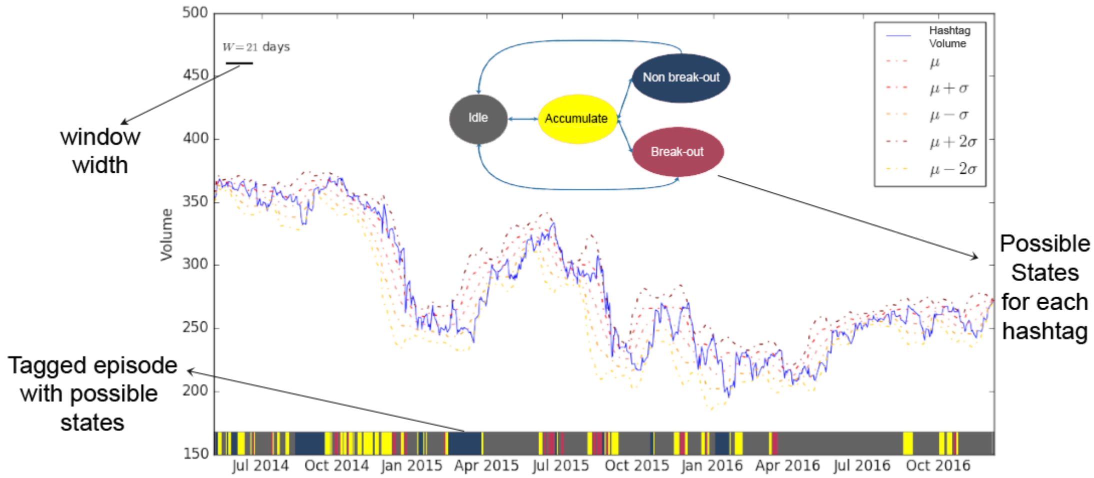
- a breakout if hashtag volume rises above($\mu(w_{size}) + 2\sigma(w_{size})$), without falling below $\mathbf{max}(0, \mu(w_d) - 2\sigma(w_d))$
- non-breakout, if hashtag volume falls below $\mathbf{max}(0, \mu(w_d) - 2\sigma(w_d))$, without rising above ($\mu(w_{size}) + 2\sigma(w_{size})$).

In a breakout scenario for an episode, no further overlapping breakouts are allowed until its volume falls below $\mathbf{max}(0, \mu(w_d) - 2\sigma(w_d))$. In both scenarios, an episode begins with its accumulation period and continues until the hashtag volume dies out (i.e. it falls below $\mathbf{max}(0, \mu(w_d) - 2\sigma(w_d))$, and possible states are idle, accumulate, breakout, non-breakout in which they are indexed to help in creating the ground truth inspired by semi-supervised learning approaches where training data are originally unlabeled. Figure 2.1 (a) shows both tweet volume as an example as well as possible state for each hashtag, and (b) shows the histograms of all daily hashtag volumes in the corpus.

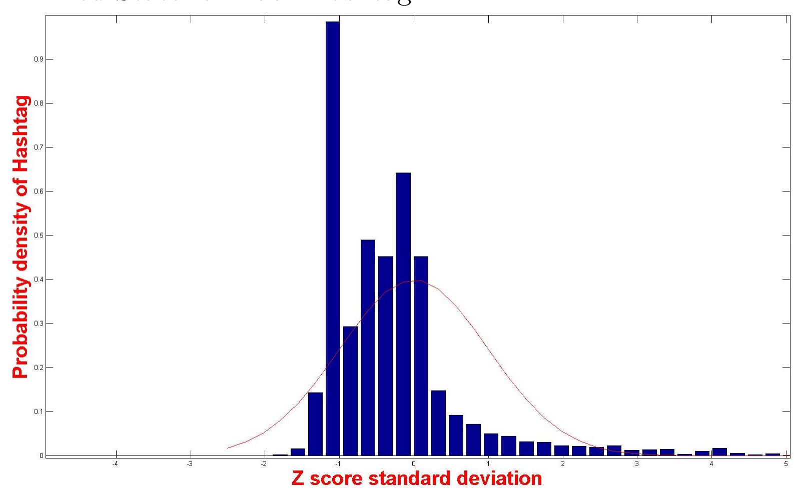
In Section 2.5 , related works are presented, followed by Section 2.6 , wherein the Tweet corpus is described. In Section 5, the Trending Hashtags Forecaster visualization tool is described, followed by Section 6, wherein the network based model, local and global network features to predict hashtag episode breakouts following accumulation periods are introduced. In Section 7, the experimental results and findings are presented. Finally, discussions regarding the model are presented in Section 8, followed by Section 9, wherein conclusions are made.

2.5 Related Work

The Twitter network has more than 271 million monthly active members, and 500 million tweets are generated daily. The vast size and reach of Twitter enables examination of potential factors that might be correlated with breakout events and viral diffusion. This study discovered that diffusion-related studies fall into two categories. First, many studies begin by analyzing social networks as a graph of connected interacting nodes, i.e. between users, friends or followers, and these studies investigate different factors that drive propagation and diffusion of information. Arruda et al. [43]proposes that network metrics play an important role in identifying influential spreaders. They examined the role of nine centrality measures on a pair of epidemics models (i.e. disease-spread phenomena on the SIR model and spreading rumors on a social network). According to the authors, epidemic networks are different from social networks, in that infected individuals in SIR are recovered by a probability μ, while a spreader of a rumor becomes a carrier through its contacts in social networks.



(a) Tweet Volume/ Window Width/ Running Mean and Standard Deviation/ Tagged Episodes with Possible Predefined State for Each Hashtag



(b) Distribution of the Z-score after Aggregated All Hashtags

Figure 2.1: Hashag Volume Example and Statistics Measures During 2014 to 2016

They found centrality measures, such as closeness and average neighborhood degree, are strongly correlated with the outcome of the spreading rumors model. The second category focused on the diffusion problem through content analysis by incorporating different natural language processing techniques. For instance, one study hypothesized that a specific group of words is more likely to be contained in viral tweets. Li et al. analyzed tweets in terms of emotional divergence aspects (or sentiment analysis) and noted that highly interactive tweets tend to contain more negative emotions than other tweets [86, 39]. This study argues that such keywords might not be generalized accordingly, as opposed to the centrality network measures of users that are typically numeric, more succinct, inherently have lower feature space dimensions and more are prone to classification over-fitting.

Weng et al. [137] investigated the prediction of viral hashtags by first defining a threshold for a hashtag to be considered viral, and then by examining metrics and patterns related to the community structure. They achieved a precision of 72% when the threshold was set statically to 70. Romero et al studied the diffusion of information on Twitter and presented several sociological patterns that make certain types of political hashtags spread more than others. Asur [20] presented factors that hinder and boost topic trends on Twitter. They found that content related to mainstream media sources tends to be the main driver for trends. Trending topics are further spread by propagators who re-tweet central and influential individuals.

This study proposes a model that predicts hashtag breakouts through adaptive dynamic thresholds, and uses generic content-independent network measures to draw information from (i) local networks corresponding to accumulation periods, and (ii) global networks corresponding to the entire network history preceding an accumulation period. The experiments showed that local network features yield an overall predictive accuracy of 76%, and that global network features yield an overall predictive accuracy of 83%. Consequently, this study determined the w_d size that yield to the highest accuracy.

2.6 Data Source

The dataset for this study is a collection of tweets from a region in the UK . These tweets have been crawled based on a set of keywords with the aim to capture political groups, events, and trends in the UK. The dataset consists of more than 3 million tweets, 600K users and more than 5.2 million interactions (both mentioning and retweeting) between users, along with 1,334 hashtags. To visualize and understand breaking hashtag phenomena, a visualization tool was built that helped facilitate exploring temporal dynamics of hashtags and their underlying networks during the accumulation period of each episode. Local and global network measures were also computed and displayed as network and node features. These network measures were used to train and test a predictive classifier, presented in the next section.

2.7 Methodology

In this study, tweets containing hashtags (case insensitive) which related to political groups in the UK from June, 2013 to July, 2014 were crawled. After crawling, hashtag episodes using techniques were detected as described in Section 2. Furthermore, the accumulation period and accumulation network of each episode were identified and the network measures corresponding to its accumulation network was extracted. Each episode was also labeled as breaking or non-breaking based on its spread.

The THF visualization tool reveals some of the discriminative patterns between breaking and non-breaking hashtags. Figure 2.3 shows the user interaction network for a non-breaking hashtag. The user interaction network denoted by the number 1 was captured during its accumulation period. Later, this hashtag did not breakout (i.e. did not cross its $\mu(w_{size}) + 2\sigma(w_{size})$, but fell back to zero volume, hence was considered as a non-breaking episode. Figure 2.4, illustrates a breakout hashtag. Following a w_{size} accumulation period, its volume exceeds $\mu(w_{size})+1\sigma(w_{size})$as denoted by network number 1. It also exceeds breakout levels by exceeding its $\mu(w_{size}) + 2\sigma(w_{size})$ threshold as denoted by network number 2. Network 3 shows the entire reach this

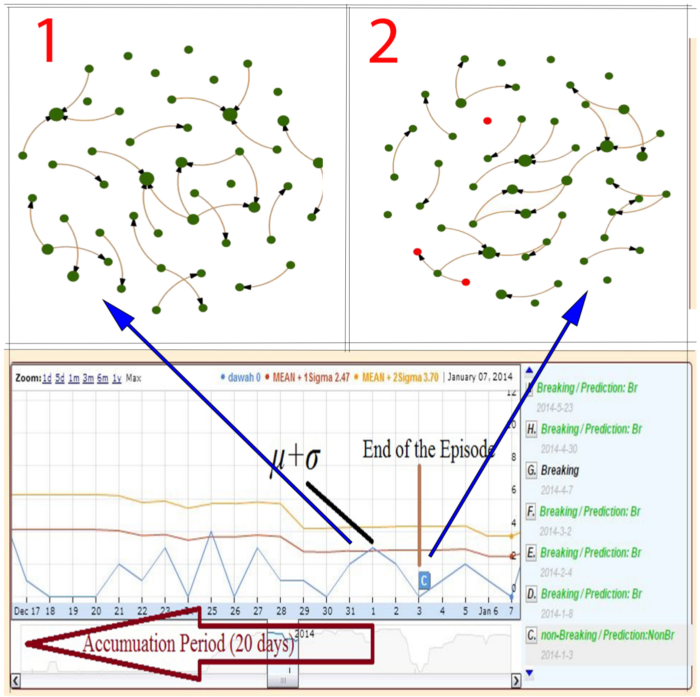


Figure 2.2: Non Breaking #brexit Hashtag Episode

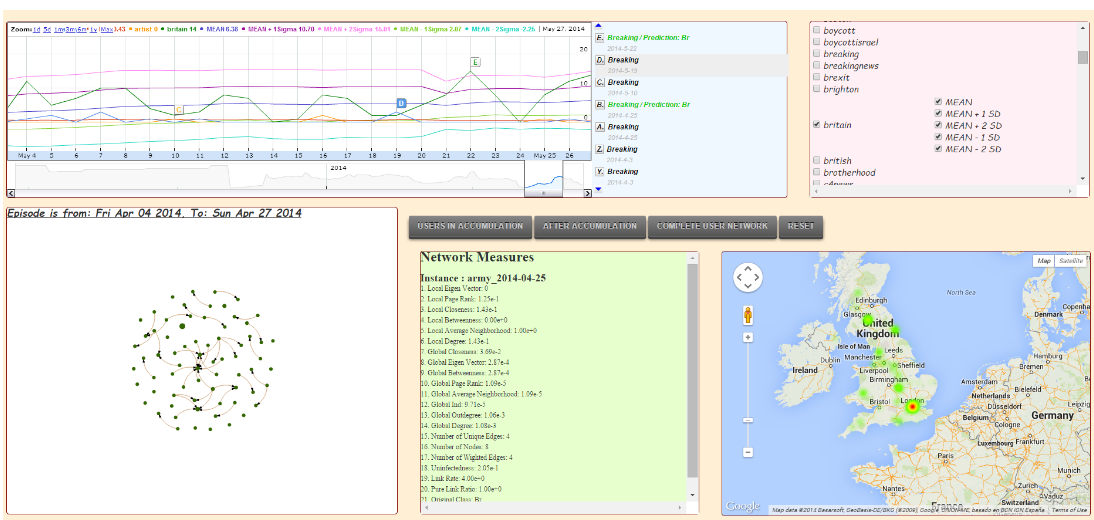


Figure 2.3: An Overview of the Trending Hashtag Forcater (*THF*)

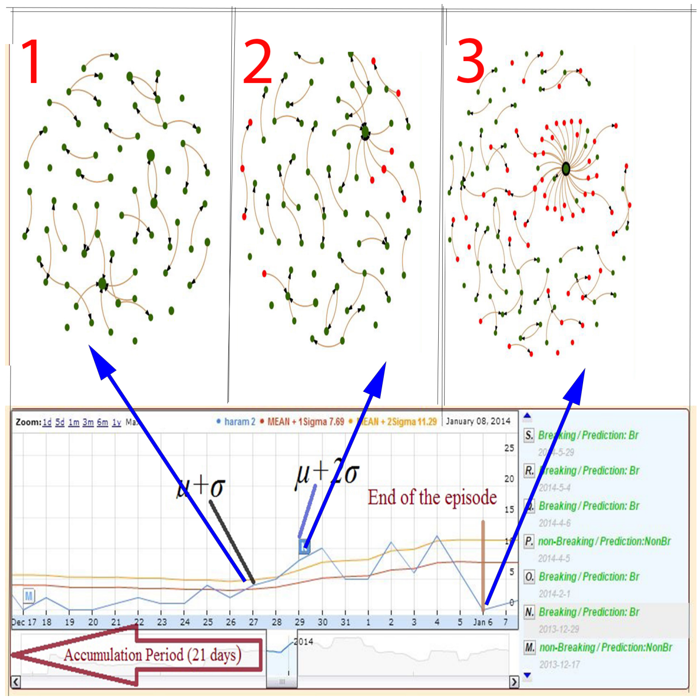


Figure 2.4: Breaking #Ukpolitics Hashtag Episode

episode before its demise (i.e. by falling below min(0, $\mu(w_{size}) - 2\sigma(w_{size})$). An interesting observation related to network 1 is the presence of a highly central green node, which attracts many new re-tweeters in network 2 and network 3. This observation indicates that existence of a large number of highly central nodes during the accumulation phase of an episode could be a good predictor for a following breakout. Patterns from similar instances could not be caught by the naked eye, yet they carry latent centrality measures correlated with the definition in this study.

2.7.1 Network Based Model

Using the network based model, this study aims to investigate how users get involved in a hashtag h by mentioning, replying or retweeting. Their interactions are depicted as a directed graph G_{h_i} , and normalized size-independent network features for directed graphs corresponding to accumulation periods of episodes are incorporated. The network graph is a pair $G = (V, E)$, where V is a set of vertices representing users together along with a set of edges representing the interactions between users. For instance, if a user $u1$ mentioned, replied, or retweeted one tweet of $u2$, then a directed edge from $u1$ to $u2$ is formed.

This study attempts to identify key features that contribute to the network based classification problem for breaking or non-breaking hashtags. Table 2.1 lists all features used for local and global measures. Local measures are associated with user interactions during the accumulation period only, whereas global measures draw their information from all interactions beginning from the start date of June 2013, until the end date of any accumulation period under consideration.

2.8 Centrality Measures

This section highlights the centrality measures used as features, or attributes, for the classifier or the learning model.

Eigen Vector Centrality: is a measure of the influence of a user in an interaction network. It assigns relative scores to all users where high scoring users are the most central, influential and important nodes.

$$C_e(v_i) = \frac{1}{\lambda} \sum_{i=1}^{n} A_{i,j} C_e(v_i)$$

The above equation can be transformed to aid in finding the eigenvector and eigenvalue, with the highest central and important nodes belonging to the highest eigenvalue. The computation is completed by the power iteration method by a given error tolerance multiplied by the number of users, with no convergence guarantee.

Page Rank: is a variant of eigenvector, and was first applied by Google [105] with the aim of ranking webpages and optimizing search engine algorithms. The basic difference between eigenvector and PageRank is that PageRank divides its value of passed centrality by the number of outgoing edges instead of passing the whole centrality as is the case in eigenvector centrality. Passing the whole centrality is less desirable since not every high central user will be mentioned or retweeted by another high central user.

$$C_p(v_i) = \frac{1}{\lambda} \sum_{i=1}^{n} A_{i,j} \frac{C_e(v_i)}{d_j^{out}} + \beta$$

A: is the adjacent matrix of users, and it can be noticed that C_p is is normalized by the outdegree of a node d_j^{out} of a user [31].

Closeness Centrality is a metric measuring the inverse of the sum of each users distances to all other users. Its closeness is defined as the inverse of the farness, whereas the farness is the sum of the distances. The intuition behind it, the influential and central users tend to quickly reach other nodes by interaction. In other words, it can be regarded as a measure of how long it will take to spread information form one node to all other nodes, sequentially. Hence, influential people can reach others and spread influence beyond average users. This centrality measure can be computed as follows:

$$C_c(v_i) = \frac{1}{SPs(v_i)}$$

where $SPs(vi)$ is the total sum of the shortest path from a user v_i to all others, and it given as follows:

$$SPs(vi) = \frac{1}{n-1} \sum_{j \neq v_i} SP_{v_i,j}$$

$SP_{i,j}$ is the shortest path between a user i and a user j. This measure can be computed by Dijkstra's algorithm ([99])

Betweeness centrality computes the number of shortest paths running through each user, compared to the number of users. The shortest path can be computed by Dijekstra's Algorithm. It is a way of looking at centrality by considering the importance of nodes in connecting other nodes. The below equation shows how this measure is computed:

$$C_b(v_i) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v_i)}{\sigma_{st}}$$

σ_{st} is the number of counts of the shortest path between user s and user t in the graph network, whereas $\sigma_{st}(v_i)$ is the a number of counts of the shortest path between user s and user t that specifically pass through v_i.

$$K_{nn,v_i} = \frac{1}{|N(v_i)|} \sum_{j \in N(v_i)} K_{v_j}$$

$N(v_i)$ are neighbors of user v_i. K_{v_j} is the degree of user v_j which belongs to the neighbor set of user v_i, $N(v_i)$

Degree centrality is a simple measure defined as the number of link (degree) incidents for a user (i.e., the number of ties that a user has). It is normalized degree distribution per node and is designated for an undirected graph, but can be used for a directed graph after conversion to undirected.

Indegree centrality is similar to the above measure but defers in the number of incoming links from a user, and is normalized for each node.

Indegree centrality is similar to the above two measure but differs in the number of outgoing links from a user, and is normalized for each node.

Number of uninfected neighbors of early adopters is the total number of retweets or mentioned edges a user has received globally, normalized by max-min retweets within the local network in a current period being measured [137]. These measures are converted from the users (vertices) basis to a graph basis by calculating the average of said measures after summing them up and dividing the number by the number of users (vertices).

Link rate this measure looks beyond just the user as a network of interaction, and deals with tweet content and URLs. It has been shown tweets that contain URLs tend to be more viral and more likely to be retweeted and mentioned. Thus, this measure is defined as follows: it is the number of URLs per Hashtag per period, normalized by the number of tweets for that Hashtag in that period.

In this study, the experiment was designed by identifying all the accumulative periods by running an algorithm, identifying its local and global network measures, as categorizing them as breaking or non-breaking.

To highlight what has been alluded to above, the following table lists the key points for the learning model:

Feature	**Description**
Eigenvector Centrality	A nodes centrality depends on its neighbors' centralities. If a node's neighbors are important, the node is most likely important too.
Page Rank	Variant of eigenvector where a node doesn't pass its entire centrality to its neighbors. Instead, its centrality is divided into the neighbors. [31]
Closeness Centrality	A node is considered important if it is relatively close to all other nodes in the network [99].
Betweeness centrality	Measuring the importance of a node in connecting other parts of the graph [58]. This measure possesses the highest space and time complexity.
Degree centrality	It measures the number of ties a node has in an undirected graph.
Indegree Centrality	It measures the number of edges pointing into a node in a directed graph.
Outdegree Centrality	It is similar to the two above measures but defers in the number of incoming links from a user, and is normalized for each node.

Link Rate	The number of URLs in the tweets during the accumulation period divided by the number of tweets.
Distinct Link Rate	Similar to link rate without the consideration of similar URLs.
Number of uninfected neighbors of early adopters	The total number of retweets or mentioned (edges) a user has ever received globally, normalized by max-min retweets within the local network in a current period being measured. ([137])
Neighborhood average degree	Measures the average degree of the neighborhood of each node [24]

Table 2.1: A Summary of All Features Space

The following three models were employed:

1. Network Interaction Approach: local, global, and a combination of both sets of features.

Also, we leverage Principle Component Analysis(PCA) for the three models' features as follows:

1. Local Features: eigenvector, pagerank, closeness, betweeness, average neighborhood degree, and degree centrality.

2. Global Features: eigenvector, pagerank, closeness, betweeness, average neighborhood degree, uninfected neighbors before break out, in degree, out degree, degree, and link rate.

3. Local and Global features combined: all features were combined in the hopes of better optimization.

Next, a 10-fold cross validation was performed for each model. The classification accuracies are presented in the next section. A PCA was conducted on the third model (which includes all local and global features) to find the best combination of weighted features. Tables 2.2 and 2.3 shows components variance weight and each variable weight to map variable into lower dimension space that accounts for the maximum variance.

Component	Eigenvalue	Variance	Cumulative Variance
1	5.785	36.157	36.157
2	3.299	20.621	56.778
3	1.669	10.43	67.208
4	1.237	7.728	74.937
5	1.01	6.305	81.241
6	0.886	5.537	86.778
7	0.663	4.142	90.92
8	0.481	3.007	93.928
9	0.423	2.646	96.573

Table 2.2: Feature Description

Feature	Component 1	Component 2	Component 3	Component 4	Component 5
PageRank Local	0.1385	-0.4467	0.1279	-0.1621	-0.0999
Closeness Local	0.0456	-0.5113	-0.0266	0.0029	0.2327
Betweeness Local	0.0493	-0.4385	0.0623	-0.0357	0.2053
Avg Neighbor Degree Local	-0.1095	-0.0442	-0.518	0.2465	0.41
Degree Cent. Local	0.1123	-0.4905	0.117	-0.1218	-0.0216
Uninfected Neighbor	0.1934	0.0186	-0.0362	-0.147	-0.3505
PageRank Global	-0.3552	-0.0977	-0.3211	-0.0954	-0.0848
Closeness Global	-0.2448	0.0922	0.4226	0.155	0.245
PageRank Global	-0.3552	-0.0977	-0.3211	-0.0954	-0.0848
Betweeness Global	-0.3576	-0.073	0.1938	-0.1173	-0.1965
Avg Neighbor Degree Global	-0.3552	-0.0977	-0.3211	-0.0954	-0.0848
In Degree Global	-0.3884	-0.0902	-0.1328	-0.1202	-0.1624
Out Degree Global	-0.2059	0.0201	0.4218	-0.0954	-0.0848
Degree Global	-0.3897	-0.0554	0.1024	0.0397	0.0353
Link Rate	-0.0336	0.1356	-0.0812	-0.5643	0.561
Pure Link Rate	0.0282	0.1889	0.1521	-0.6357	0.0816

Table 2.3: First Five Components Sorted from the Highest Corresponding Variance to the Lowest

2.9 Discovering the most appropriate window size

Figuring out the right window size poses a significant challenge, and is one of variables which helps to efficiently identify patterns and capture underlying signals. This specific problem has been of interest for various applications dealing with large streamline data with rules setting and decision such as trading algorithms (e.g. [125, 61]). Adjusting the right window size w_d for timestamps data analysis is necessary to distill useful training set among the data stream, and it is tied to labels/classes (based on out rule definition). One expected quality that obtained labels expected to posses a good proportional balance since they are w_d dependent. Larger window size causes the running average and standard deviation to lag and be less responsive to daily volume changes, causing large mis-informative instances and categories, while small window size leads to missing the real underlying signal. Thus, a trade-off approach is needed to locate w_d fit by examining various w_d settings and optimizing the maximum weighted accuracy value associated with w_d . Weighted accuracy was employed since the fitness function accounts for each class (i.e. breaking and non-breaking) rather overall accuracy for whole which the latter prone to inaccurate measure about imbalanced training data. In other words, weighted accuracy is best fit to penalize imbalance data sets even majority well predicted as the learning model tends to favor the dominant class over the other leading to bad models. w_d value holding highest accuracy ACC is given in the equation below (2.1)

$$\underset{w_d}{\operatorname{argmax}}\, ACC(w_d) \tag{2.1}$$

$ACC(w_d)$ is an average of the value of the outcome accuracy percentage of breaking accuracy (equation 2.3) and non-breaking accuracy (equation 2.4)

$$ACC(w_d) = average(ACC(w_d, X_{Brekaing}, Y_{Brekaing}, \hat{Y}), ACC(w_d, X_{Nonbreakaing}, Y_{Nonbreakaing}, \hat{Y})) \quad (2.2)$$

$$ACC(w_d, X_{Brekaing}, Y_{Brekaing}, \hat{Y}) = \frac{1}{n_{Breaking}} \sum_{i=1}^{n_{Breaking}} \mathbb{1}(Y_{Breaking(i)} == \hat{Y}_i) \quad (2.3)$$

$$ACC(w_d, X_{Nonbrekaing}, Y_{Nonbrekaing}, \hat{Y}) = \frac{1}{n_{Nonbreaking}} \sum_{i=1}^{n_{Nonbreaking}} \mathbb{1}(Y_{Nonbreaking(i)} == \hat{Y}_i) \quad (2.4)$$

The $\mathbb{1}(x)$ is an indicator function defined as follows:

$$\mathbb{1}(x) = \begin{cases} 1 \quad \textbf{if} \quad x \quad \textbf{is} \quad true \\ 0 \quad \textbf{if} \quad x \quad \textbf{is} \quad false \end{cases} \quad (2.5)$$

The learning model was defined as $F : X, Y \to \hat{Y}$, where $X = \{x_1, \ldots, x_n\}$ is the training set and $Y = Y_{Brekaing}|Y_{Nonbrekaing} = \{y_1, \ldots, y_n\}$ such that $\forall y_i \in \{-1, 1\}$ are the observed labels, based on the empirical rule where '-1' denotes the non-breaking class, while '1' denotes the breaking. $\hat{Y}$ is the prediction obtained by the learning model which aims to learn model parameters to fit Y.

2.10 Cross Validation and data dependent constraints

Cross validation is a technique which examines the goodness of fit of a given model to the data. In general, cross validation (e.g. the most common type **k-fold-cross-validation**) assumes data instances: $X = \{x_1, \ldots, x_n\}$ sample and are independent and identically distributed (i.i.d.), but this assumption cannot hold true with temporal data that is likely inclined to possess some level of dependency based on closeness among data, or at least the timestamp order. For instance, weather trade forecasts demand training sets to always be successive, in timestamps, training set always precedes testing with time [19]. Hence, many studies proposed dependent cross validation. Earlier, [41] proposed a methodology to alter k- fold-cross-validation to disregard data instances occurring in validation sets (fold) that meet dependence condition of any instance in the training set (other folds). The essence of dependencies between any two-given data point when both occur within a predefined **distance**. Meanwhile, the suitable definition of dependencies is proposed by [123], which is not related to cross validation but rather to time split of training and test set. They searched for $\hat{t}$ stamp that utilized to split data into two parts where the first is training and the second is testing, thereby no testing instance is successive in any of the training with a time order perspective. [75] used both cross validation (not randomly placed) and disregarded instances that meet dependency conditions proposed by [123]. This is called Time-Border-Split Validation. This study used what was proposed, but disregarded overlapping episodes of the same hashtags.

2.11 Experiment Results and Findings

A best fit of w_d setting was decided and it was found that $d = 20$reached the height accuracy stated in Section 2.9 after having tried the values from $d \in \{2, 3, \ldots, 30\}$. 2790 episode accounts were found for non-breaking, while 1331 episode accounts were found for breaking. Next, the correlation between features and breaking hashtags using the Principle Component Analysis (PCA) was examined. The PCA is a dimensionality reduction approach that analyzes datasets to find which features give the highest variance among instances and maps the given features into a lesser number of factors called components [107]. Finally, to predict whether a given hashtag will break or not, a supervised network based learning model was used.

2.11.1 Features Correlated with Breaking Hashtags

The PCA identified nine factors shown in Table 2.4 According to Kaiser Criterion [22], the factors to consider are the ones with eigenvalue above 1. This study focuses on the first two components since they reveal interesting insights. Table 2.5 shows the correlation between the studys features and the first two components shown in Table 1. The first component is strongly negatively correlated with global measures, whereas the second component is strongly negatively correlated with local measures. These two components proved that global features should be grouped together as and they contributed heavily (36%) to the variation in our dataset. Additionally, some of the local measures were also grouped together in a single factor and slightly contributed (21%) to the variation in our dataset.

Component	Eigenvalue	Variance	Cumulative Variance
1	5.79	36.16	36.17
2	3.30	20.62	56.78
3	1.669	10.43	67.21
4	1.24	7.73	74.94
5	1.01	6.31	81.24
6	0.88	5.54	86.78
7	0.663	4.14	90.92
8	0.48	3.01	93.93
9	0.42	2.65	96.57

Table 2.4: PCA Components

2.11.2 Network based model

For this model, two sets of features were measured: local and global. For local features included: eigenvector, pagerank, closeness, betweeness, average neighborhood degree, uninfected neighbors before break out, and degree centrality. The global features included: the previous features measured globally, plus in degree, out degree, and link rate. Next, a random forest classifier was trained and tested with 10-fold cross-validation (time series depended aware) using three approaches: prediction us-

Feature	Component 1	Component 2	Feature	Component 1	Component 2
PageRank Local	0.14	-0.45	PageRank Global	-0.36	-0.10
Closeness Local	0.05	-0.51	Closeness Global	-0.24	0.09
Betweeness Local	0.05	-0.44	Betweeness Global	-0.35	-0.07
Avg Neighbor Degree Local	-0.11	-0.04	Avg Neighbor Degree Global	-0.3552	-0.10
Degree Cent. Local	0.11	-0.49	Degree Global	-0.3897	-0.0554
Uninfected Neighbor	0.19	0.02	In Degree Global	-0.39	-0.09
Link Rate	-0.03	0.14	Distinct Link Rate	0.0282	0.1889
Outdegree Global	-0.21	0.02	-	-	-

Table 2.5: Correlation Between Table and Components

ing all features shown in Table 2.1, prediction using global features that are correlated with the first factor identified by the PCA shown in Table 2.6, and prediction using local features that are correlated with the second factor returned by the PCA as shown in Table 2.7. The highest precision of 84%, recall of 81% and F-measure of 82% for breakout prediction with the global features was achieved. Additionally, the highest precision of 82%, recall of 85% and F-measure of 84% for non-breakout prediction with the global features was also achieved. On the other hand, local features showed overall lower precision and recall of roughly 76%. These findings suggest that global measures outperform local measures in predictive accuracy.

Network	TP	FP	Precision	Recall	F-measure
Local	0.73	0.2	0.77	0.73	0.75
Global	**0.81**	**0.15**	**0.84**	**0.81**	**0.82**
All Features	0.8	0.16	0.83	0.8	0.81

Table 2.6: Break Out Results

Network	TP	FP	Precision	Recall	F-measure
Local	0.79	0.27	0.75	0.79	0.77
Global	**0.85**	**0.19**	**0.82**	**0.85**	**0.84**
All Features	0.84	0.20	0.81	0.84	0.83

Table 2.7: Non Break out Results

2.12 Chapter Conclusion and Remarks

A model for predicting breaking hashtags was developed using a content independent network model comprising both local and global network features drawn from an indicative accumulation period of hashtag volumes. For the network model, this study measured and experimented with the predictive accuracies of global and local features, and examined their importance and rankings using the PCA. Global features drawn for the accumulation period network showed higher predictive accuracy compared to the local features. The network based model with global centralities for the accumulation period network can be used as a general framework to predict breaking hashtags with an overall accuracy of 82%. This can help in fostering analysis and detecting the most relevant hashtags that may yield a breakout earlier in its nascent stages, especially when sheer volume of tweet content is examined. The next chapter will discuss how to detect and find ideology and frames, and will use window size to examining contentious frames among political factions.

Chapter 3

FINDING ORGANIZATIONAL ACCOUNTS BASED ON STRUCTURAL AND BEHAVIORAL FACTORS ON TWITTER

3.1 Introduction

Social media has emerged as an integral part of a connected lifestyle, in which people locate and interact with each other to stay informed on current issues and to help shape their own opinions. Twitter is a micro-blogging social network which allows users to post, like and share or retweet short messages or tweets. According to the most recent Twitter statistics [1] , more than 330 million active daily users post more than 500 million tweets daily, a quarter of which are tagged with hashtags. 80% of active Twitter users are outside the United States. Twitter supports over 40 languages and allows users to connect with their friends as well as organizational accounts such as religious, political and educational groups, NGOs, news outlets, public figures, and celebrities [2] .

Several works focus on developing methods to predict a Twitter account's demographic attributes such as age [100], gender [52], location [90], and political affiliation [50]. Related works on individual vs. organization (IvO) detection include those by De Silva et al., McCorriston et al. and Kim et al. [56, 47, 44]- bearing in mind the significant role of organizational accounts for being what is called "mouthpieces" of

[1] Twitter Usage Statistics Report, Internet Live Stats, `http://www.internetlivestats.com/twitter-statistics/`

[2] This chapter is an extension of earlier master thesis work by Chinmay Gore [63]. We redefined the model and the way we split of account into two classes (i.e. binary classification/detection problem: Individual vs. Others), as well as refined the research question, strengthened the literature review, supported it with a statistical test beside the visualizing features of the two classes and repeated the learning classification experiment, This work has culminated to an accepted paper [17] after the extension

disseminating ideologies and aiding mass mobilization within both political and communication context [97]. IvO detection algorithms were previously used for ground truth labeling in community detection, where key organizational accounts, such as political parties and their leaders, were located and their followers, who like and share their messages, were grouped into different communities. Unlike Facebook pages and groups, Twitter does not explicitly support the notion of an organizational account, hence detecting them remains an open problem.

In order to calibrate and evaluate a classification model that could detect organizational accounts on Twitter, we experimented with a 5-month Twitter corpus comprising over 7 million tweets from Bangladesh. We sampled 31,139 accounts using cold-start heuristics presented in Section 3.3 to locate and label nearly 200 organizational accounts classified as 68 NGOs, 62 news outlets, 35 political groups, and 17 public intellectual and iconic figures. The remaining 30,957 accounts were labeled as individual user accounts. We then experimented with a set of network-based, behavioral, temporal and spatial features, all independent of domain and language, to identify relevant features useful in differentiating between IvO accounts. Organizational accounts [3] correspond to a rare category of accounts in this corpus, with less than 1% frequency. Following this, we experimented with a set of linear and non-linear classifiers. The highest performing sparse logistic regression classifier achieved an accuracy of 68.2% and 64.4% recall leading to a 66.2% F1-score in detecting organizational accounts using the set of domain- and language-independent features presented in Section 3.4.

The rest of the paper is organized as follows: section 3.2 features a review of related works. Section 3.3 describes the Bangladesh tweet corpus that we used and

[3]Since public intellectuals, leaders and celebrities share similar spatial, behavioral and connectivity related characteristics with group and organizational accounts, we labeled them together as "organizational" - as opposed to "individual" accounts.

our cold-start heuristic methods-based ground truth collection process. Section 3.4 explains the content-independent, network-based, spatio-temporal and behavioral features employed in our experiments. Section 3.5 presents experimental evaluations and findings. Finally, Section 3.6 concludes the chapter and discusses future work.

3.2 Related Works in IvO Type Detection

Rao et al. [51] propose a method to infer a user's demographic account information, for e.g., age, gender, political affiliation and location, by detecting latent features in their messages. Late paper by Onur Varol et al. [53] works on resolving a bot-detection issue by employing a diverse set of features categorized into user profile, followers, network, temporal, content and sentiment related ones. They found roughly 10% of accounts as bots. However, organizational accounts seem to be a rarer category than bots. A paper by Wu et al. [55] focuses on categorizing Twitter accounts and then analyzing the flow of information between these accounts. Their focus is on how to validate the two-step communication flow model developed by Katz et al. [77]. In this model, elite users are divided into four categories - mass media, organizations, celebrities and other bloggers. A bag of discriminating indicative keywords was generated and a score calculated for unseen accounts based on their tweet content to enable classification. In most recent study by David Savage et al. [116], they develops anomaly detection methods suitable for use in social media to detect anomalous account patterns such as malicious spammers, fraudsters, cyberbullies and predators. Their methods also use language- and content-independent network and behavioral features for detecting anomalous accounts.

Three related studies examine the IvO account detection problem. De Silva et al. [44] propose a classification model that depends on multi-lingual content features tested on English and Spanish datasets. McCorriston et al. [47] employ a set of network (e.g., ratio of followers to friends, etc.), temporal (e.g. tweet volumes), spatial and content-related features to distinguish between individuals and organizations on Twitter. Kim et al. [56], utilizing network- and content-based features to identify organizational accounts, reported that content-based features were the key determiners of account types. The key contribution of our method is that, it employs a set of *content- and language- independent* network behavioral and spatial-temporal features for identifying those organizational twitter accounts.

3.3 Tweet Corpus

According to a 2016 census, Bangladesh's population is more than 168 million [70]. The number of Internet users in Bangladesh now stands at over 66.8 million, which means there is a 41% penetration. Facebook, with a usage of about 97.2%, is the most used social network while Twitter ranks second with 1.08% usage – approximately 1.7 million accounts. Our tweet corpus includes all tweets tweeted between June and October 2016 that were either geo-tagged as being from within Bangladesh or from users whose account location on their profile mention a city or a place in Bangladesh. The tweet corpus statistics are as shown in Table 3.1.

3.3.1 Ground Truth Labeling

The ground truth labeling followed a set of cold-start heuristics yielding 31,139 candidate accounts, which were further validated by a human labeler. Accounts were sorted in descending order according to their PageRank and degree centrality measures of retweet, follower and user-mentioned networks in an attempt to leverage the high importance of organizational presented in their accounts which are indicated by high centrality measures. We identified those accounts by their centralities as a preliminary step before identifying organizational accounts then a validating manual

Feature	Value
Number of tweets	7,090,560
Number of users	150,000
Minimum timestamps	June 1, 2016
Maximum timestamps	October 31, 2016
Tweets Location	Bangladesh
Languages used	English, Bangla

Table 3.1: Tweets Dataset

Account type	Count
Individual	30,957
NGOs	68
News	62
Political organizations	35
Celebrities	17

Table 3.2: Ground Truth Data

Feature	Retweet Network	Mention Network	Followers Network
Nodes	308,477	335,678	12,604,797
Edges	681,404	431,437	25,942,312
Connected components	4,305	11,755	1079
Average node degree	75	48	1078
GCC Nodes	299,890	298,337	12,600,824
GCC Edges	675,682	405,813	25,939,414

Table 3.3: User × User Behavioral Interactions Network

labeling step to follow on. The top 200 accounts on each list were retrieved and manually labeled. We also used heuristic matching rules by employing translations of the same keyword lists used by Wu et al. [55] to detect candidates to label, based on the following rules:

News: We curated a list known newspapers and TV channels in this country using Wikipedia since its wiki-pages are voluntarily available for users engagement, edits and updates. Twitter accounts for each aforementioned newspaper/TV channel on the curated list were after matched within the corpus.

Celebrities: this label is given to each member of a curated list of politicians,movie actresses & actors, and public name figures, who are well known in Bangladesh, was collated from Wikipedia and matching handles were located and verified in the corpus. Some additional celebrities and other types of organizational accounts were located and labeled by taking a closer look at accounts with the largest number of followers.

Political Organizations: We made a list of political parties and groups with the help of three political scientists from Bangladesh, matching their names with Twitter handles and profile information. Additionally, several Bengali keywords indicating political, social, religious groups and organizations were used to match and validate Twitter handles as belonging to them.

NGOs: A list of Bengali keywords indicating NGOs was created and matched, alongside local branches of globally active NGOs, for e.g., Red Cross, listed on Wikipedia.

Individuals: We created a list of regular expressions like "I am," "I'm," "I love," "I work at," "I like," etc. in Bengali to match and label roughly 30K accounts as individuals. Table 3.2 shows the frequency distribution of each type of of labeled account as per the above heuristic approaches expectedly revealing imbalanced data and reflecting the reality of the actual population of accounts; hence imposing challenges in our training models - addressed in the experiment section 3.5.

3.4 FEATURES

In this section, we describe the network, behavioral and spatio-temporal features, listed in Table 3.4, that were used to differentiate between IvO accounts. These features are content- and language-independent, i.e., they rely only on the non-textual features of the Twitter accounts. Retweet, follow and user-mentioned networks were also extracted and computed from the tweet corpus. The connectivity statistics of these networks are shown in Table 3.3.

3.4.1 Network Features:

We created three directed weighted graphs based on retweet, follower and user-mentioned information. This means a directed edge is added from user A to B if user A retweeted user B, or if user A mentioned user B, or if user A follows user B in the corpus. The following subsections show the various network centrality measures we

Features				
Network			Account Profile	Tweet Location & Timestamp
Retweet	Followers	User Mention		
-Degree Centerality	-Degree Centerality	-Degree Centerality	-Number of list of users	-Location Entropy
-Pagerank Centerality	-Pagerank Centerality	-Pagerank Centerality	-Number of favourites by users	-Location Variance
-K-core Centerality	-K-core Centerality	-K-core Centerality	-Ratio of friends to followers	-Timestamp Entropy
-Clustering Coefficient	-Clustering Coefficient	-Clustering Coefficient	-Hashtag centeralities and Clustering Coefficient	-Timestamp Variance

Table 3.4: Feature Sets Used in the Learning Model

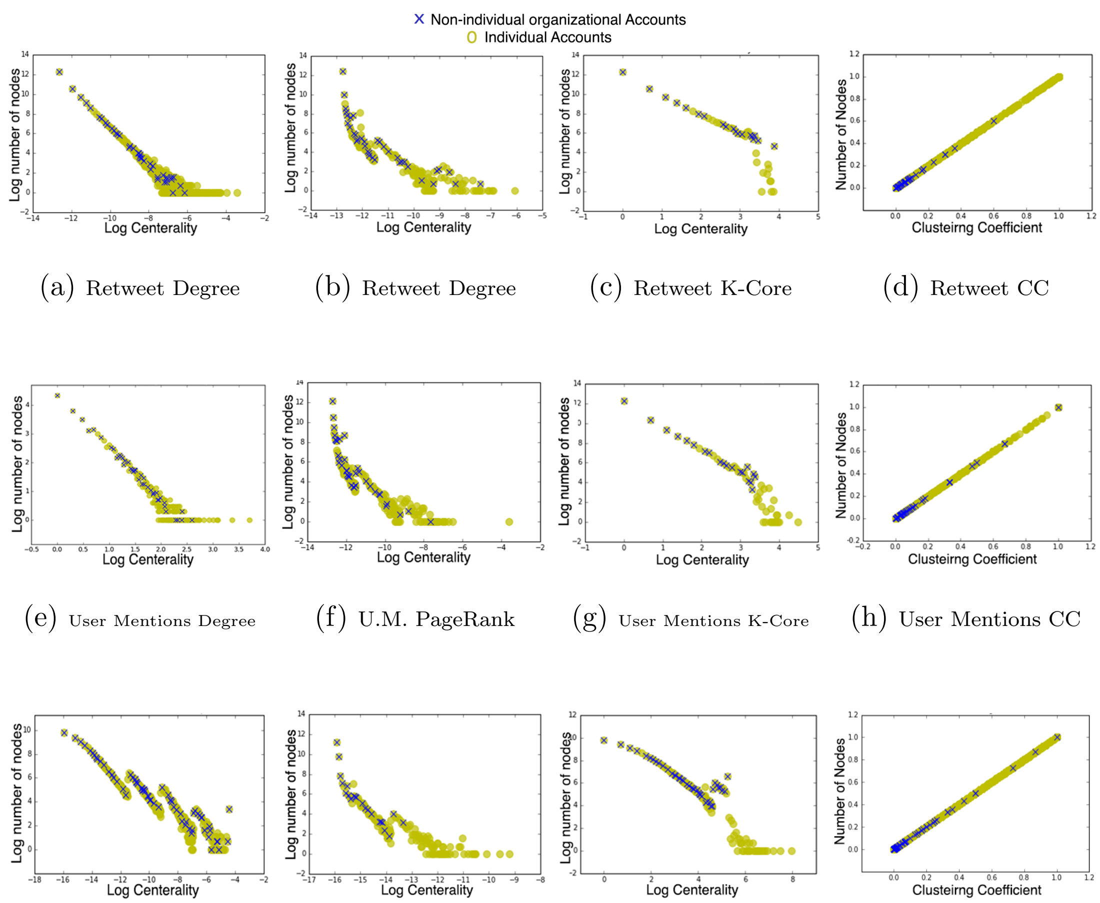
experimented with: subsubsection[Degree Centrality:] This measure of a given node represents the fraction of nodes it is connected to Sabidussi [113]. The higher number a node has, the more nodes it connected. This is further divided into in-degree and out-degree, where values are normalized to keep them bound between [0,1]. Figure 3.1 (a, e and i) charts the log-log centrality distributions for IvO accounts. From this, we observe that news organizations and celebrities have high in-degree and low out-degree centralities, and organizational accounts are located towards the head of the corresponding power law distributions. For organizational accounts, user-mentioned centralities tend to be lower as they are mentioned more often than they mention others.

PageRank Centrality:

This is an extension of the eigenvector centrality[30]. PageRank can be computed iteratively and the accounts' values explain their relative importance by using a damping factor until convergence. In figures 3.1 (f and j), we observe that organizational accounts tend to have lower PageRanks in follower and user-mentioned networks [105].

K-core Centrality:

Another important centrality measure [122], the k-core decomposition process is initiated by removing all nodes with the degree k=1. This causes new nodes with the degree k≤1 to appear. These are also removed, and the process is continued until the only nodes remaining are those of degree k>1. The removed nodes and their associated links form the 1-shell. This pruning process is repeated for the nodes of degree k=2 to extract the 2-shell, that is, in each stage the nodes with degree k≤2 are removed. The process is carried on until all higher-layer shells have been identified and all network nodes have been removed. In Figure 3.1 (c, g and k), it can be observed that organizational accounts' k-core values are clustered towards the head of the distribution, separated from individual accounts' k-core measures, which are clustered towards the tail.



(a) Retweet Degree (b) Retweet Degree (c) Retweet K-Core (d) Retweet CC

(e) User Mentions Degree (f) U.M. PageRank (g) User Mentions K-Core (h) User Mentions CC

(i) Followers Degree (j) Followers PageRank (k) Followers K-Core (l) Followers CC

Figure 3.1: Degree Centrality, PageRank, K-Core and Clustering Coefficient for three type of networks: Between, User Mention, and Followers Behaviors. Not all points are shown, 50 points were sampled for visualization, bearing in mind some data points overlap each other.

Clustering Coefficient:

This coefficient is a measure of the degree of which nodes in a graph tend to cluster together [135]. For instance, friends of friends are likely to have a high clustering coefficient with the coefficient's value close to 1, while sparse and rarely connected graphs display a coefficient value closer to 0 [81]. We hypothesized that followers' connectivity levels would vary based on the type of account. For example, organizations would have a diverse set of followers who are not connected to each other. The graphs in Figure 3.1 (d, h and l) show the clustering coefficients of IvO accounts where organizational accounts tend to cluster around lower coefficients with some outliers.

3.4.2 Tweets' both Timestamp and Location:

Tweets originate from various locations and have different timestamps. Especially in the case of individuals tweeting from their mobile devices, users' spatial-temporal behaviors might help differentiate them from more stationary organizational accounts. Descriptions of the entropy- and variance-based temporal and spatial features that we experimented with are as follows:

Temporal Features:

For every user, we have a distribution of timestamps marking times from when the user tweeted. Interestingly, entropy measures failed to capture a significant difference between IvO accounts, however, variance measures based on timing and locations of tweets showed a higher ability to display variations.

Spatial Features:

We utilized geo-tagged tweets to compute spatial entropy and variance measures for accounts with spatial information, and found that organizational accounts tend to show lower variance in geolocation indicating their posting locations are more stationary. Entropy, nonetheless, did not display enough discriminative power, suggesting that variance can be better explained by geolocation deviation around the means per account.

3.4.3 User Profile Features:

We gathered profile information for all the Twitter accounts stored in a database, such as user descriptions and their favorites count. We extracted the following features from user profiles:

Ratio of Followers to Friends :

This is one of the indicative features proposed by Ethem Can et al. [32] in their study, which showed a high correlation of user's posts being retweeted with the proposed ratio. We plotted the followers to friends ratio for all types of accounts and observed that organizational accounts have a lower friends-to-followers ratio.

Favorites Count:

Twitter allows users to like any tweet and this information is captured as their favorites count. We observed that individual users have higher favorite counts as compared to organizational accounts [36].

List Count:

Twitter lists make it possible for users to create a curated list of other Twitter accounts of their choice and interest. On their timeline, users can then view a feed of tweets originating from the members of their lists. This functionality was introduced in 2013, and was first used as feature by Yasugi et al. [57]. Users can create their own lists or subscribe to preexisting ones. List counts represent the number of lists users are members of, and it was observed that organizational accounts tend to have lower list counts.

Username Frequency:

Keywords in individuals' usernames tend to match many others, whereas there are keywords in organizational accounts' names that tend not to match frequently. When all keywords in a name are repeated multiple times among other Twitter handles, then the account is likely to belong to an individual. This variance motivated us to include it as feature in our learning model.

Hashtag Network:

Hashtags are common terms or phrases that identify messages related to a specific topic or an event. Users who tend to share prevalent hashtags, they usually engage in discussion on analogous topics. We made use of each accounts' hashtag usage to build a hashtag network as highlighted by Wagner & Strohmaier [133] study on distilling features from the social network. We started by gathering the hashtags used by every account. Then for every common hashtag between accounts, we added a weighted edge, where accounts sharing multiple hashtags had higher weights. Next, we calculated centrality measures and clustering coefficients for each account in the hashtag network.

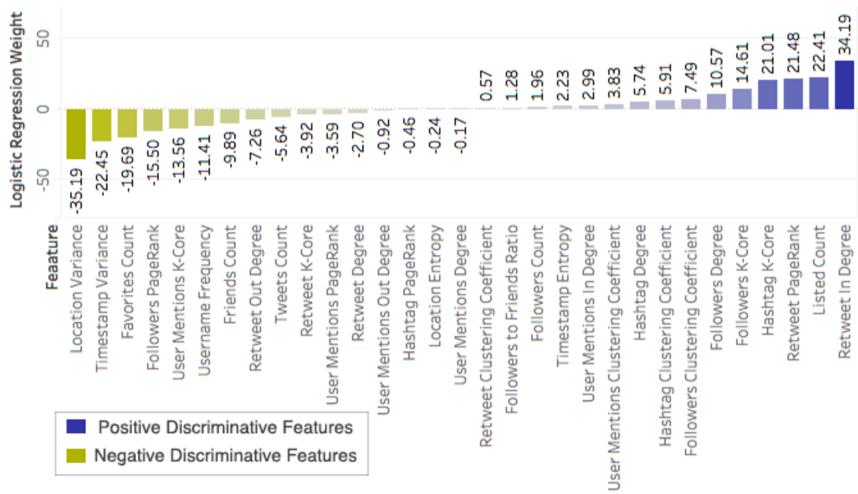


Figure 3.2: Logistic regression model weights.

3.5 EXPERIMENTS

This section highlights our experimental results and evaluations in detecting the rare organizational accounts in our tweet corpus – including political, social and religious groups, celebrities, public figures and NGOs.

3.5.1 Normalizing Computed Feature Measures:

We normalized all feature values to make them scale invariant. For these centrality measures and clustering coefficients, the values were already normalized to lie between 0 to 1, except for the k-core centrality. Features of User profile based as friends to followers ratio, favorites count and name frequency were normalized as well.

3.5.2 Detection of IvO Accounts:

Besides showing labeled data distributions across various features for visual inspection in Section 4, Fig. 3.1, here we also conducted a statistical test to show that individual accounts behave differently from organizational ones. We employed a multivariate two-sample test between individual and organizational accounts, as proposed by Baringhaus et al. [23]. We experimented with 1,000 permutation bootstrap-replications between two types of account distributions. The *critical-value* measure was obtained as 9362.8 based on 95%-value-of-confidence, and an *observed statistical value* of 38,9025 was calculated, which is significantly larger than the *critical-value*; thus confirming the alternative hypothesis that distributions corresponding to IvO accounts are significantly different from one another.

Next, we considered the labeled data to evaluate the performance of the classifiers. We reported precision, recall, and f-measure metrics for each category of accounts using 10-fold cross-validation on the resultant balanced data adapted by using the undersampling method [129] targeting individual dominant class. Table 3.5 shows that the logistic regression-based model outperformed other models such as Random Forrest, Multilayer Perceptron (MLB), AdaBoost and Gaussian Naive Bayes. Discriminating feature weights of the logistic regression model can be utilized to understand positively and negatively correlated features for individual and organizational account categories. In the logistic regression model, positively weighted features relate to organizational typed accounts, whereas negatively weighted features relate to the individual typed accounts. Figure 3.2 shows the corresponding weights of all features employed in the model.

Classifier	Precision	Recall	F1 Score
Logit Regression	**.682**	**.644**	**<u>.662</u>**
RandomForest	.563	.689	.62
MLP	.426	.611	.502
AdaBoost	.293	1.	.453
GaussianNB	.309	.622	.413

(a) Classification Performance the <u>rare</u> Organizational Accounts

Classifier	Precision	Recall	F1 Score
Logit Regression	**.989**	**.991**	**.99**
RandomForest	.991	.985	.988
MLP	.988	.976	.982
AdaBoost	1.	.93	.95
GaussianNB	.988	.96	.974

(b) Classification Performance for the Dominant Individual Accounts

Table 3.5: Classifier Performance for Individual and Organizational Account Types

3.5.3 Discriminating Features for Organizational Accounts:

Organizational accounts tend to have higher valued:

- Retweet pagerank: Organizational accounts tend to be retweeted by other central users
- Hashtag k-core: Organizational accounts use central hashtags towards the core of the hashtag network
- Followers k-core: Organizational accounts are located towards the core of the followers network
- Followers clustering coefficient: Followers of organizational tend to follow each other as well
- Retweet in-degree centrality: Organizational accounts tend to be retweeted more often
- List count: Organizational accounts have higher list counts

3.5.4 Discriminating Features for Individual Accounts:

Individual accounts tend to have higher valued:

- Spatial variance: Individuals tend to Tweet more from assorted locations through portable devices.
- Timestamp variance: Individuals tend to have an arbitrary tweeting activity, where as organizational accounts are more time structured
- Favorites count: Individuals like/endorse others' Tweets more often compared to organizational accounts
- Followers pagerank: Organizational accounts tend not to follow many others

- User mentions k-core: Organizational accounts tend not to mention many others

3.6 Conclusions

In this chapter, a classification model to detect organizational accounts on Twitter was proposed by employing a set of content and language independent network, temporal & behavioral and those spatial features. It has been known that community detection is a popular method for understanding the network structure. Nonetheless, Communities can further be described by locating their key influencer. The method used in this study can be employed to automatically detect organizational accounts such as religious, political, educational groups, NGOs, news outlets, public figures and icons located in any community, thus helping to name the key groups and actors driving such collectives.

Chapter 4

IDENTIFYING IDEOLOGIES FRAME OF ADVERSARIAL POLITICAL GROUPS

4.1 Chapter Introduction

The ubiquity of social media has facilitated political discourse and allowed politicians, activists, and normal users to engage in communication compared to any precedent time. Social media plays an integral part of the military operation of antagonists factions to serve as a tool to shape public opinion and be used in manipulation long-term communication strategies. It also serves daily logistics in organizing operations, mobilization, information dissemination and intelligence. An opposing opinion may arise attempting in portraying social media as a tool that can be leveraged positively in societies where members engage in discussions and address various social and political issues, yet many views suggest that it equivalently can lead society to be more fragmented and prejudiced, bearing in mind the attempts of political groups in pushing for political framing and narrative agendas. Hence, it would exacerbate the problem to further deterioration in the social fabric and cohesion among society members. For instance, [130] raises some concerns about the obscure shifts of the American culture living in the digital age to being more narcissistic than older generations as a result of the discreditation and disappearance of many rooted values and qualities. This premise can be generalized, to greater extents, in other cultures alike due to globalization. Since social media has been adopted by users of almost all countries, there are roughly 2.64 billion users in 2017 [1]

[1] Number of social media users worldwide 2010-2021- Statista https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Social media can be exploited for political gain and help serve in warfare strategic communication. This so-called Soft Power was introduced by Nye [102] in 1991. With the shifting forms of power form solely military, postpost–World War IIto new forms of hegemony, such as the predominant influence by media outlets, communication channels and social media, several examples show how social media can influence outcomes for political gain between rival nations. One prominent example is the allegation that the Russians meddled in the US presidential election. Trumps campaign became the focus of attention for any possible colluding with Russia, which came with the hacking breach incident of the National Committee. Both the US Intelligence and the Office of Cybersecurity and Communications have a broad consensus that Russian hackers covertly breached the Democrats. Later, the incident became an acceptable fact that Russia play a maim role in meddling the 2016 US election. By many observers note, Russia have been successful in the seizure of confidential documents and have spread rumors of the Democratic campaign in social media to defame them publicly, in the same time, virulently over- shadowed Republicans in favor for one candidate over the other especially peculiarly when the race margins are very narrow and election, and consequently outcomes can be swung to their favored candidate through public influence. This can serve as a clear example of how public opinions can be leveraged for a set of goals and agendas, either locally or by political groups or foreign governments.

Such an exploitation of social media is not only confined by the adoption between rival nations but also serves as a tool within sole countries and between its opposing parties. Almost all organizations and political groups shaped its identity by a set of values and vision work to make their agenda persuasive and apparent to the public consumption pushing for their causes and narrative interpretation. These efforts are consistent with making less margins of others points of view by entrenching the

"Othering" [101], which leads to further political turmoil and has aggravated the situation in civil war zones adding to it the own confirmation bias that finds its polarized breeding ground in the social media. This causes the audience to focus on facts that confirm their points of view and interpretations, and less likely to question other interpretations [89, 65] [2] . These media devices have a negative impact on the political turmoil in countries such as Egypt, Syria, Libya, and others, as these have allegedly led to the Arab Spring which culminated in humanitarian disasters and displacement of more than 11.5% of people were killed or injured in Syria and 11 million where displaced in and out of Syria . [3]

A look on the holistic online debates of the last decade will reveal that these gained more traction by social scientists with the advent of social media, known as Web 2.0 [132]. Users have become the cornerstone and the main drive of influence power decentralized contents generation than outlet media itself. This notion aligned with the quantum leap shift in the means of social communication means, as they mutated from conventional journalism incubating debates to the public forms of engagement with ease.

In this chapter, this study lays the foundations of perspective discovery techniques that are used to detecting issues embraced by the cohesive dimensions of social theory. The dataset was crawled from Facebook and Twitter, based on a list of well-chosen keywords (i.e. dictionary-based) that a crawler uses to retrieve the most relevant posts/tweets to the research study in terms of content, timing, and location of the posts.

[2] "othering". James Norris wrote "The "othering" is any action by which an individual or group becomes mentally classified in somebodys mind as 'not one of us'. Rather than always remembering that every person is a complex bundle of emotions, ideas, motivations, reflexes, priorities, and many other subtle aspects, its sometimes easier to dismiss them as being in some way less human, and less worthy of respect and dignity, than we are."[101].

[3] Syrian Center for Policy research- http://scpr-syria.org/publications/policy-reports/

4.2 Debate and frames definition

Debate, persuasion and framing are all related, and lead to the ignition and the animation of the political sphere. Debate is defined as a solemn discourse addressing a set of intriguing topics and issues to audiences (especially in online platform channels) [7, 8]. On the other hand, framing is an integral part of a debate. Framing refers to the rhetorical bundling of frequent keywords regarding issues to specifically encourage certain interpretations and discourage others for a political benefit. Thus, frames are the dimensions of debate that amenable to identification and vigorous analysis that underlying pattern and relations in posted text, ranging from syntactic and semantic lexical clues, and word cluster. Framing is effective when chosen words, phrases, memes, images favor an interpretation, while discouraging others [124]. This study will identify the frame set for each camp that leads to contentions, ranging from political camps to other camps.

This model was selected because it can explain both violence and radicalization engines as debate inherit marginalizing the othering. According to James Norris, the othering is any action by which an individual or group becomes mentally classified in somebodys mind as not one of us. Rather than always remembering that every person is a complex bundle of emotions, ideas, motivations, reflexes, priorities, and many other subtle aspects, its sometimes easier to dismiss them as being in some way less human, and less worthy of respect and dignity, than we are.” [101].

4.3 Data Source

The curated data represent politically related posts in social media originated by users in Libya, one of the most prominent examples of a contemporary state struggling to retain basic levels of economic, social and political stability after the end of Muammar Qaddafis forty-two-year rule. A salient division between two opposing powers with foreign backers has since cropped up, namely General National Accord (GNA) and House of Representative (HoR), located in Tripoli and Tobruk respectively as portrayed on the map in Figure 4.2 of Section 4.4. Both factions scrambled for power, seeking to delegitimize the other, by forming alliances from a mixed array of militia units, and tribal or territorial-based armed groups. Such a political landscape continued to bring more division, as partisanship-abetting radicals and extremists along with their accomplices grew as a result of rifts. The intent is to track and provide meaning of a complex situation with multiple players and a myriad of opposing and disruptive ideologies. This study adapted the social cohesion theory and extended the basic premise of this framework to construct the model pictured in Table 4.1, with multiple dimensions to reflect the human/internet/physical landscape of the Libyan conflict zones. The human landscape of Libya is divided into six major categories with three subcategories. With the consideration that groups and individuals are often complex beings with more than one affiliation and possibly inconsistent sets of values and ideologies, these dimensions allow us to look over the issue from a predefined dimension for better and concise understanding. Our data contains two types of documents: (1) Facebook posts and other related blogs and (2) Tweets and the related metadata, such as timestamp, GEO location, hashtag, etc. The number of Facebook posts is 27,024, and the tweets are two million tweets in the span of 22 months. The figure below illustrates the volume of tweets, and its collection was from January 2016 to December 2017. Figure 4.1 - from different data collection warehouse.

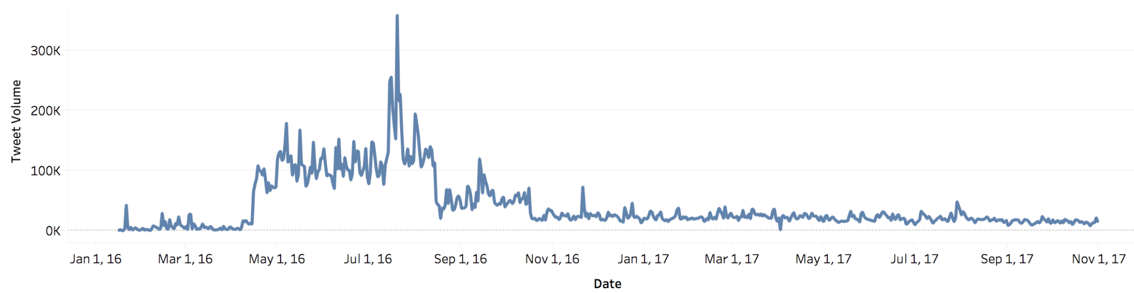


Figure 4.1: Tweet Volume Chart Show Temporal Daily Volume and the Start and End Date of Data of Data Collection: Feb 2016 until Nov 2017

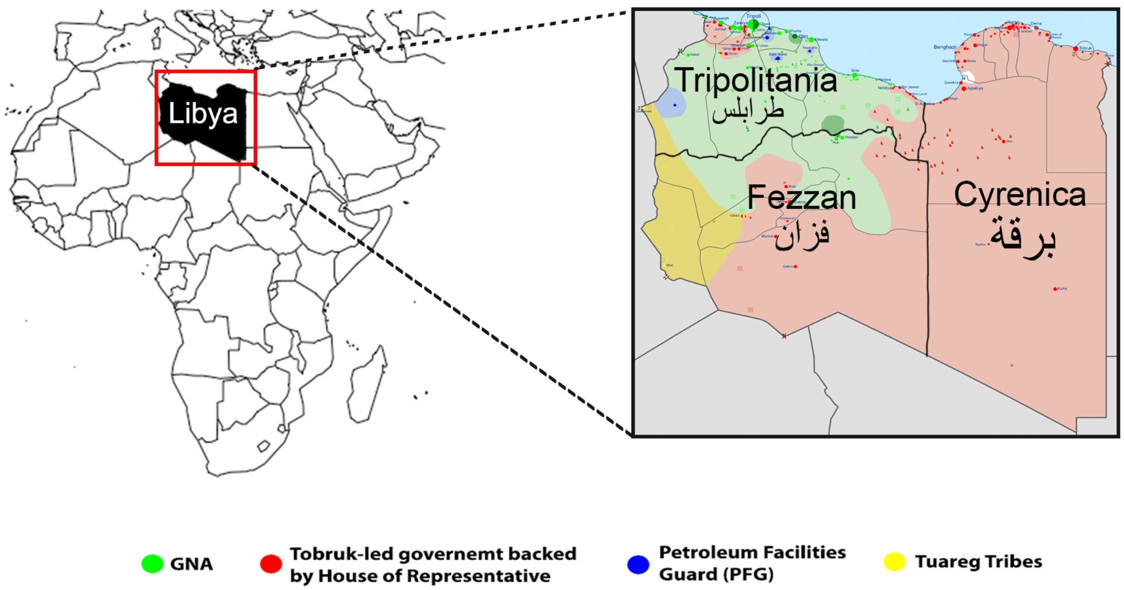


Figure 4.2: Libya Is Located in the Northern Africa, East of Algeria and Tunisia and West of Egypt. The Northern Side of the Border Is the Mediterranean Sea. Sources and References: [45, 141]

4.4 Factions background

Based on the conversations in the online spaces, the major findings are as follows:

1. The pro- Government of National Accord [حكومة الوفاق الوطني] (GNA) is an interim government formed and supported by United Nations Security Council, with headquarters in the Tripoli. This faction has been attempting to gain traction within a deeply divided nation by capitalizing on their internationally-

Model		
(1) Social Cohesion	(a) Political (POL)	Legitimacy & Participation (LP)
		Illegitimacy & Non-involvement (IN)
	(b) Economic (ECON)	Inclusion (INC)
		Exclusion & Discrimination (EXC)
	(c) Social (SOC)	Recognition & Belonging (RB)
		Rejection and isolation (RI)
(2) Justice vs. Injustice (JST)		Justice
		Injustice
(3) Safety & Security vs. Insecurity (SFSC)		Safety & Security (Security)
		Insecurity
(4) Autonomy & Sovereignty vs. Foreign Influence (ASFI)		Libyan Nationalism (LN)
		Western Influence (WI)
(5) Culture (CLTR)		Secular / Liberal (SC)
		Religious / Islam / Sharia (RIS)
(6) Well-being (WB)		Prosperity/High Quality of Life (PGL)
		Poor / Unemployed / Under-Employed (PUU)

Table 4.1: The Social Model

designated legitimacy and their implementation of a cohesive judicial and monetary system. They likewise claim their faction can contribute to the stability/security of the people, and mend the division of the nation and reconcile with forces on the ground. They are considered as having some affinity with Muslim Brotherhood and independent militia on the ground, which is considered the major impediment pushing its legitimacy across society bodies.

2. The Libyan Civil Society Organization (CSO) Non-Govermental Organizations (NGOs)[المؤسسات الغير حكومية \ مؤسسات المجتمع المدني]) is a non-governmental organization (NGO) present in Libya, and is largely organized by youth groups concerned about social services and in favor of a United Libya built by their own people. They are also concerned with the safety of the nations population, and are skeptical of the GNAs ability to solve this issue. However, they show a high level of solidarity toward the GNA, as the latter allegedly embraced democratic principles.

3. The Dignity Allied forces [عمليات الكرامة] are heavily focused on the Cyrenaica province and propose the division of Libya into its three constituent regions in line with the federalist system. Their legislative branch in Tobruk, the House of Representatives, has a healthy competitive drive and has created offices within the government in response to both the GNA and NGNC to maintain a sense of legitimacy. They have built alliances with local militias, which include ties to Salafist, to fight ISIS and al-Qaeda, as well as increase security in Eastern Libya. They pursue controlling region and resist avalanche of Muslim Brotherhood potential influences with other Islamist insurgent groups, and their top priority is to put national stability first against any other set of cross-border ideology.

4. Libya Dawn [فجر ليبيا]and the GNC are mostly anti-GNA, and are unlikely to cede power willingly, regardless of popular opinion. They also have a large number of posts discussing individual politicians and political niche parties, suggesting that there is much scheming and division behind the scenes. This faction has been complicit as having some ties with extremists and Muslim Brotherhood extremists and their sleeper cells.

5. The Salafist groups [السلفيين], Muslim Brotherhoods animus, are largely non-violent and skeptical of the three functioning governments. They have a surprising number of young adherents who are anti-establishment and anti-professional politicians. Much like the younger generation involved in the current US presidential election (especially those who support the Bernie Sanders movement), they claim establishment parties and politicians are not advocating for enough changes to the governmental system. This group pushes to instill continence among followers by depoliticizing people for the sake of retaining stability, and assembling followers under one ruler (i.e. in Libyas case: the army led by Haftar and the HoR) to fight extremists. Many foes of the Salafists consider them to be a conservative Islamic group, and non-violent.

6. ISIS [داعش] and al-Qaeda [القاعدة] have a diminishing online presence, which is reflected in their waning physical presence in Libya. They also have flexible alliances with local militias, and each other depending on their needs in a given moment, and are claiming they contribute to the security of local communities in the face of increasing national instability. Both are known for attracting and recruiting young men and women including foreigners to propagandize a narrative for redemption with full dedication to martyrdom. Both are well known for atrocity and wreaking havoc, as well as for illegal trades and smuggling oil to fund their operations. Despite their diminished presence, they remain one of the main factor for social unrest in Libya, but they startling grow up quickly in power vacuum calling for utopia society of Caliphate State.

4.5 Methodologies

We attempt tp make meaning out of a complex situation with multiple major players (holding a myriad of different ideologies and goals) is derived from the theory that cooperation and dissent between factions that can be predicted by several factors, including economic disparity/parity, cultural biases, political disenfranchisement or empowerment, feelings of security, justice, etc. In our case, we have added to/modified the basic premise of this theoretical frame. We found the broad dimensions (i.e. (i) Alienation/Belonging, (ii) Marginalization/Inclusion and (iii) Dispossession/Social Justice) and tailored to reflect the human/internet/physical landscape of the Libyan conflict zones as pictured above.

In particular, as the model above shows, we have divided the human landscape of Libya into six major categories with three subcategories (bearing in mind that both individual and groups individuals are often complicated process such that many prominent issues can be related to more than one dimension such as injustice and in security where the both are very are related.

The first category is devoted to the three most important elements of social cohesion [74] as outlined by the initial theoretical model. These are defined by feelings/impressions under the following categories:

1. **The legitimacy or illegitimacy** of the political system.

2. I**nclusion or exclusion** in economicissues.

3. **Belonging or isolation** in the larger Libyan society.

4. **Justice or Injustice**. In Libya's case, the keywords in this category have reflected the actions of local militias (are they oppressing the people of their communities) and the actions of international organizations such as the United Nations and Western nations (particularly in the cases of ISIS and AlQaeda).

5. This category has to do with impressions of **Safety and Security** . How do the populations of these groups articulate their state of mind about the

safety of their communities/daily lives or the nation as a whole. These online conversations are tied to spiritual matters or the stability of the region since security so often depends on the governments (or governments in Libyas case) military and police forces. Unstable governments, of course, most often have inefficient, corrupt, or ineffective security forces.

6. This category **Autonomy & Sovereignty vs. Foreign Influence** Influence is focused on the relative autonomy of the nation, and is especially applicable in Libya because of the United Nations interference in the second civil war, and their involvement in the toppling of the Ghaddafi rule. There is also the leftover negative outlook on the political and military involvement of Italy, due to the nations colonial rule of Libya.

7. The seventh category **Culture** (especially the conflict between secular and religious),is particularly of concern for the non-ISIS Salafists, ISIS, and Al-Qaeda, since they are the most likely populations to favor a very conservative Islamic government and local laws based on sharia.

8. The final category (eight) is centered on the expression of feelings of **Well-Being**, is particularly of concern for the non-ISIS Salafists, ISIS, and al-Qaeda, since they are the most likely populations to favor a very conservative Islamic government and local laws based on sharia. Well-Being is related to good or bad social/civil service and unemployment and underemployment. Like other conflict zones (Kosovo, in particular), there is a high rate of unemployment, especially for youth populations, consistent medical care, water treatment, food distribution for the poor, garbage pick-up, etc. These are all practical, everyday matters that affect the life and livelihoods of local populations. For ISIS and al-Qaeda, these impressions of well-being can also encompass religious matters

the right way of living in the path of God to inspire and mobilize followers to disseminate the redemption narrative and serve their political agenda.

Social model dimensions are combined with extracted keywords from Facebook pages after they were indexed. First, salient controversial keywords were categorized through a shallow search approach of entire keyword space and propagated through multiple rounds of vetting by social scientists to create the most representing balanced set in size of each dimension (i.e. 1000 keywords for each). A combination of most frequent keywords, LDA topic models [29] as well as topical phrase mining [48]was used. These keywords are given blue/red colors (blue being a positive feeling/impression and red being a negative feeling/impression on the part of the group members being measured). Then, they were used to measure both the amount of dialog devoted to one of the eight categories by members of each group and the sentiment of the target group towards a certain topic. Figure 4.4 shows an example of these bigram keywords. When we refer to bigram as$n$-**gram** with $n = 2$ of observing two consecutive keywords in a corpus. $n = 2$ seemed to be the best setting to trade capturing the context of the post with the undesirable sparsity which is proportioned to the n [88]. Tables 4.3 and 4.4 shows the quantify the amount for each dimension and each coalition [4] .

For example, the initial impression for all seven of the political groups, from GNA supporters to al-Qaeda and ISIS, are devoting over 20% of their posts on social media to discussing their positive feelings about the security/stability of the country (Category 5, measured by +secure and -insecure) as opposed to other issues of itics/government (POL and Category 1 in the model), the economy (ECON, Category 2), society (SOC, Category 3), justice (Category 4), the autonomy of Libya (Category

[4]Figure 4.4: Some English translations show more than two consecutive words due to the approach of extracting the bigrams of Arabic contents but not the English, therefore some of the Arabic bigrams trivially mapped into two bigrams English keywords

6), religious and national cultures (Category 7), and small-scale economic/social/civil senses of well-being (Category 8). Interesting phenomena was observed of how information propagation flows, and influence crowds opinions disseminate. It starts with news and ideologies emitted from Facebook pages that show a high level of articulation and plausibility among the audience, touching their causes and sentiments in different formats such as articles, memes and photographic media. This effort at the source is managed by the political cohort and its echo ripple to resonate into Twitter feeds by the audience showing the high presence of Twitter users. This is one explanation for this phenomenon. Facebook represents a more accurate and reliable information outlet for groups usage, but users favor Twitter users to interact with. Twitter provides anonymity, allowing opinions to be expressed openly with less censorship when reacting to the news. One explanation is by [69] wherein Facebook and Twitter tend to be used differently by users. Additionally, per the Arab Social Media Influencer Summit, Facebook is sitting astride the social means in almost all Arab countries (including Libya), except a few countries, whereas WhatsApp swaps positions with Facebook. However, Twitter is within the average position of other social media outlets for 2015 [128].

4.5.1 Identifying Ideology from Tweet

An ideology is the affiliation of a user to any of the seven aforementioned cohorts. This study aims to build a supervised classification model. Hence, learning models require labeled data be available with a primary goal that classifiers can learn from patterns presented in the text. The only useful information about groups affiliation are the Facebook pages where monitoring of the current Facebook pages and the new emerging ones is completed periodically. Subject matter experts manually identify association so that all the posts which are crawled, inherit a label of hosting page affiliation which is then inserted into the learning model. To sum up, Facebook pages are leveraged to learn from and to classify users tweets based on the learned model after training.

We trained four models to benchmark and select the outperform model. We used (1) the Linear regression with Ridge Regularization adaptation (equation 4.1) (2) the logistic regression by employing the logit transform to the A matrix as shown in Equation 4.2 where logit loss is used, and we use (3) the SVM model 4.3 with and without and with kernal to compare and to extract the most informative features and to build a classification model that can be used for classifying tweets contents. Table 4.2 shows list of parameters for the classifier. We leveraged Facebook to build that model since we alrady maintain the list of Facebook pages affiliations, so each post inherits the affiliation of its hosting pages.

Kernel function used for the SVM is $\phi(x) = x$ if kernel is the identity ,but if RBF $K(a, \hat{a}) = \phi(a)^t \phi(\hat{a}) = \mathbf{exp}(-\gamma \|a - \hat{a}\|^2)$

$$\underset{x}{\operatorname{argmin}} \sum_{i=1}^{M} w_i (y^i - (x^t a_i + c))^2 + \lambda_1 |x| + \lambda_2 x^t x \tag{4.1}$$

$$\underset{x}{\operatorname{argmin}} \sum_{i=1}^{M} w_i \log(\exp(-y_i(x^t a_i + c)) + 1) + \lambda_1 |x| + \lambda_2 x^t x \tag{4.2}$$

$$\underset{x}{\operatorname{argmin}} \frac{1}{2} \|x\|^2 + C \sum_{i=1}^{M} \mathbf{max}(0, 1 - y_i \hat{y}_i)) \tag{4.3}$$

$$\text{such that } y_i(x^t \phi(a_i) + b) - 1 \geq 0, \forall i = 1, \ldots, N$$

$A_{M\times N}$	Data matrix representing a Vector Space Model (VSM) where each row represents a Facebook post and columns represent features as bag of words. We transform A to **tf-idf** [106]
Y	The ground truth labels as a vector or response that classifier aims to learn, Y, as binary class we assume $y_i \in \{0, 1\}$ for regression and $y_i \in \{-1, 1\}$ but for SVM
$\hat{Y}$	The predicted response given by the learner/classifier trying to mimicking
λ_1	Lasso regularization parameter [127] known as L_1 norm
λ_2	Ridge regularization parameter [67] known as L_2 or the Euclidean norm
X	The returned weighted vector for each feature/attribute. It can be leveraged to learn the polarity for each feature for discriminative analysis goal
M	Number of instance (Facebook posts)
N	Number of features (e.g. bag of words, **bigrams)**
i	$i \in \{1 \dots M\}$ and used as asubscript to indicate a row a_i within A matrix or an element such x_i or y_i within X or Y respectively.
j	$j \in \{1 \dots N\}$ and used as subscript to indicate a column a_j within A matrix or an element x_j within X vector.
W	Weight vector for each row. The default setting $w_j = \frac{1}{M}, \forall i \in \{1 \dots M\}$
C	A constant used with SVM Hinge Loss for penalizing misclassified posts and since they are non-separable due to the nature of textual content
ϕ	Denotes a kernel function
c	A constant used in regression

Table 4.2: Nomenclature

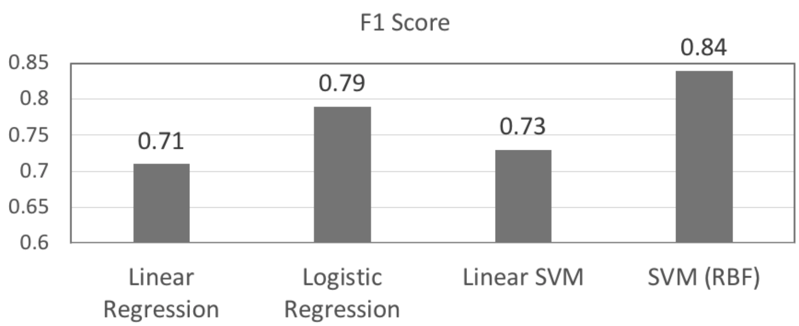
4.6 Experiment

The four classifier models were used to (1) benchmark four types of classifiers widely used with text (2) select the one corresponding to the highest l performance scaled with F1 scores as a harmonic means of the precision and the recall. Figure 4.3a shows that the SVM with the RBF kernel outperformed other classifiers, and hence, is adapted to classify ideologies of all tweet posts. See Figure 3.3a. Figure 3.3b subsequently breaks down classifier performance to depict each ideologys precision, recall, and F1 score corresponding to highest average F1-score (i.e. 0.84). It is evident that the classifier has an average F1 score of 0.8 with the exception of ISIS, AlQaeda, and Salafists. This is due to low documented presence, as well as a common presence of prayers and minor religious topics as discussed in Facebook pages which do not necessarily show a level of political framing. This study used One-Verse-Other strategy for a multiclass learning scheme [86, 11]. The same classifier is then adopted to categorizing Facebook posts, and is used to predict ideology of tweet posts, as shown in Figure 4.5 and 4.6 . Facebook posts were color-coded based on ideology, and the Sankey diagram was used to explain the ideology flow and the proportion when comparing Facebook and Twitter.

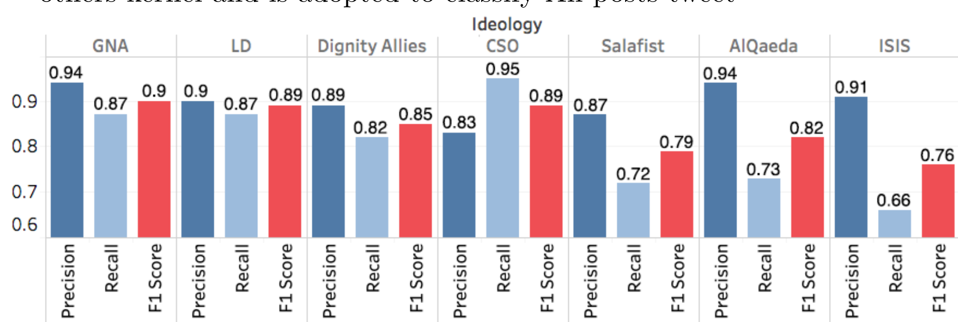
We utilized the four classifier models for (1) benchmark four type of classifiers widely used with text (2) to select the one corresponding the highest l performance - scaled with F1 score as harmonic mean of the precision and the recall. Figure 4.3a show that the SVM with the RBF kernel outperformed other classifiers, and hence, is adapted to classify ideologies of all tweets posts. Figure 4.3b.

In the chart showed in Figure 4.3b the performance is broken down classifier performance to show each ideology precision, recall, F1 score corresponding to highest average F1-score (i.e. 0.84) as depicted in the Figure. One may observe that the classifier has an average F1 score > 0.8 except for the ISIS, AlQaeda, Salafists due to low documents presence as well as a common presence of prayers and minor religious topics as discussed in Facebook pages which is not necessarily show a level of political framing. We used One-Verse-Other strategy for a multiclass learning scheme [111, 14]

The same classifier is, next, adopted to categorizing Facebook posts, and the same classifier used to predict tweet posts ideology as it shown in Figure 4.5 and 4.6 as Facebook posts were color-coded based on ideology, and used the Sankey diagram to explain the ideology flow the proportion when comparing The Facebook with the Twitter.



(a) F1 Score of four classifiers. SVM with the RBF kernel outperforms others kernel and is adopted to classify All posts tweet



(b) Ideology Classifier Performance: SVM (RBF)

Figure 4.3: Classifier Performance of Ideology Detection- of Facebook Posts.

[120] made a holistic survey of several approaches of classification, and results show SVM was one of the best classifier approach in terms of accuracy and continues to be one of the most important methods for text categorization in many streamlines of research. It essentially tries to find the best $M-1$ hyperplane that separate data, as though it were a binary classification problem. Thus, this study first employs linear SVM and fine-tune the C parameter used to control Hinge Loss (as opposed

to Log loss in logistic regression) that penalize misclassification in the relaxation optimization problem that goes beyond merely widen the hyperplane margins. Then, an experiment is conducted with non-linear SVM through Radial Basis Function (RBF) kernel and its parameter, γ, controlling parameter Kernel (Gaussian) that map features into higher dimension space.

	Arabic	Translation	
LP	الانتخابات التشريعية	Legislative elections	
LP	العملية الشرعية	Legitimate operation	500 Keywords
LP	مراكز الاقتراع	Voting centers	
⋮	⋮	⋮	
IN	الغير دستوري	Unconstitutional	
IN	الرفض الشعبي	Crowd's Rejection	500 Keywords
IN	الغير قانونية	Illegality	
⋮	⋮	⋮	

(a) Political (POL)

	Arabic	Translation	
INC	للتنمية المستدامة	Sustainable Development	
INC	الاجتماعي والتنمية	Social Development	500 Keywords
INC	تنمية إقليم	Region Development	
⋮	⋮	⋮	
EXC	الإقصاء والتهميش	Exclusion and Marginalization	
EXC	استبعاد الأقليات	Excluding Minorities	500 Keywords
EXC	التهميش الممنهج	Systematic Marginalization	
⋮	⋮	⋮	

(b) Economic (ECON)

	Arabic	Translation	
RB	المجتمع المدنى	Civil Society	
RB	أطياف المجتمع	Spectrums of Society	500 Keywords
RB	لخدمة المجتمع	To Serve Communities	
⋮	⋮	⋮	
RI	تهجير السكان	Displacement of population	
RI	تهجير الأفارقة	Displacements of Africans	500 Keywords
RI	قوارب المهربين	Smugglers' Boats	
⋮	⋮	⋮	

(c) Social (SOC)

	Arabic	Translation	
Justice	العدالة والمساواة	Justice and equality	
Justice	الحقوق والحريات	Rights and Freedoms	500 Keywords
Justice	العدالة والكرامة	Dignity and Justice	
⋮	⋮	⋮	
Injustice	الظلم والعدوان	Oppression and Aggression	
Injustice	الطغيان والاستبداد	Tyranny and Despotism	500 Keywords
Injustice	والظلم والقهر	Injustice and Iniquity	
⋮	⋮	⋮	

(d) Justice vs. Injustice (JST)

	Arabic	Translation	
Security	الأمن والاستقرار	Security and stability	
Security	استقرار الأوضاع	Stabilize the situation	500 Keywords
Security	والأمن القومي	National Security	
⋮	⋮	⋮	
Insecurity	الاضطرابات السياسية	Political Turmoil	
Insecurity	جرائم قتل	Murder Crimes	500 Keywords
Insecurity	الانتقام للشهداء	Retaliation for the Martyrs	
⋮	⋮	⋮	

(e) Safety & Security vs. Insecurity (SFSC)

	Arabic	Translation	
LN	استقلالية السلطات	Independence of authorities	
LN	استقلال الوطن	The Independence of country	500 Keywords
LN	كف التدخلات	Stopping Interventions	
⋮	⋮	⋮	
WI	العلمانية والإنحلال	Secularism and Decadence	
WI	الانفتاح للغرب	Opening up to the West	500 Keywords
WI	العلمانية اللادينية	Secularism and nonreligion	
⋮	⋮	⋮	

(f) Autonomy & Sovereignty vs. Foreign Influence (ASFI)

	Arabic	Translation	
SC	العلمانية والدولة	Secularism and the State	
SC	ليبرالية الإسلام	The Liberalism in Islam	500 Keywords
SC	حكم العلمانيين	Secular Rule	
⋮	⋮	⋮	
RIS	تحكيم الشريعة	Arbitration of Sharia Law	
RIS	المنافقين والمرتدين	Hypocrites and Apostates	500 Keywords
RIS	أرض الخلافة	Caliphate Land	
⋮	⋮	⋮	

(g) Culture (CLTR)

	Arabic	Translation	
PGL	التخطيط والتنمية	Planning and Development	
PGL	التنمية الإدارية	Management Development	500 Keywords
PGL	الجودة الشاملة	Total Quality	
⋮	⋮	⋮	
PUU	الفساد والمخدرات	Corruption and Drugs	
PUU	الأوضاع المتردية	Deteriorating Conditions	500 Keywords
PUU	ضعف الرقابة	Weak Oversight	
⋮	⋮	⋮	

(h) Well-being (WB)

Figure 4.4: A Sample of Keywords Bigrams Belonging to Each Social Model Categories

4.7 Visualizing Ideologies and Issues Dissemination

In attempt to find meaning out of volume counts presented in Facebook and Twitter, Tables 4.3 and 4.4. The Sankey Diagram was used to chart the ideologies and issues flow from the left side (Facebook) to the right (Twitter), as shown in the Figures 4.5 and 4.6. [119]. The Sankey diagram was first created by Riall Sankey in the 19th century to represent energy flow in factories, but such a visualization was widely adopted in different knowledge fields, such as finance and other social sciences- tracking opinions and social movements mentoring [138, 110]. Moreover, a visualization tool was made available to serve subject matter experts exploration need such that s/he can click and browse on the chords . It was adopted in this chapter, and depicted ideology and energy flow as representing the model category counts that disseminate from political cohort to their along the way to affiliated users in twitters. Figure 4.5 explains all political groups/factions that disseminate information related to all dimensions. Nonetheless, the figure clearly presents the group size and their followers. Furthermore, Facebook posts by political groups and commentators appear to be more positive than tweets from Twitter users. On the other hand, Figure 4.6 presents more significant details as this study breaks down the Sankey diagram per political group to reveal the detail volumes, particularly for the smaller political groups. It is clear that the CLTR (i.e. the Culture) dimension is thicker within the Sankey diagrams of those groups claiming religious ideology.

Social Dimensions			Coalitions							
			GNA	LD	Dignity Allies	Salafist	CSO	AlQaidah	ISIS	Total
Social Cohesion	(1) POL [LP,IN]	Legitimacy & Participation	1,230	8,543	8,543	1,275	6,184	32	242	16,034
		Illegitimacy & Non-involvement	94	2,361	2,361	1,167	2,411	21	13	5,960
	(2) ECON [INC,EXC]	Inclusion	726	5,132	5132	8478	7281	81	57	20972
		Exclusion & Discrimination	44	4,862	4,862	529	724	8	19	6,123
	(3) SOC [RB,RI]	Recognition & Belonging	462	3,838	3,838	1,863	3835	27	85	9,563
		Rejection & isolation	291	8,557	8,557	2,956	6691	74	307	18,278
(4) Justice vs. Injustice [Justice,Injustice]		Justice	1,119	5,047	5,047	3,328	4,128	25	163	12,528
		Injustice	35	1,645	1,645	2,575	1,826	6	8	6,052
(5) Safety & Security vs. Insecurity ss [Security,Insecurity]		Safety & Security	4,235	20,213	20,213	13,053	22,948	316	681	56,530
		Insecurity	311	9,250	9,250	4,003	5,375	120	192	18,748
(6) Autonomy & Sovereignty vs. Foreign Influence [LN,WI]		Libyan Nationalism	362	4,983	4,983	2,038	4,794	25	89	11,840
		Western Influence	315	6,310	6,310	8,674	8,543	69	124	23,596
(7) Culture [SC, RIS]		Secular/Liberal	255	5,441	5,441	8542	7,660	65	105	21,708
		Religious/Islam /Sharia	235	6,717	6,717	11,107	4,850	148	355	22,822
(8) Well-being [PGL, PUU]		Prosperity/good quality of life	797	5,224	5,224	8,095	8,091	83	75	21,493
		poor/ under-employed/ unemployed	290	2,236	2,236	2,361	2,998	21	32	7,616
Total			**10,801**	**100,359**	**100,359**	**80,044**	**98,339**	**1,121**	**2,547**	**279,863**

Table 4.3: Social Dimension Tracking of Facebook pages used by political cohorts to orient and mobilize the crowds

Social Dimensions			Coalitions							
			GNA	LD	Dignity Allies	Salafist	CSO	AlQaidah	ISIS	Total
Social Cohesion	(1) POL [LP,IN]	Legitimacy & Participation	574	2,954	7,019	156	7,789	62	131	18,685
		Illegitimacy & Non-involvement	227	711	2,465	133	2,346	50	20	5,952
	(2) ECON [INC,EXC]	Inclusion	256	1,125	2,285	56	3,333	15	23	7,093
		Exclusion & Discrimination	217	437	1,986	130	2,424	38	21	5,253
	(3) SOC [RB,RI]	Recognition & Belonging	1,149	2,965	9,141	743	16,753	125	297	31,173
		Rejection & isolation	1,086	8,557	8,557	2,956	6691	74	307	18,278
(4) Justice vs. Injustice [Justice,Injustice]		Justice	512	1,462	4,586	229	6,908	71	33	13,801
		Injustice	287	950	4,556	606	6,734	202	3	13,328
(5) Safety & Security vs. Insecurity ss [Security,Insecurity]		Safety & Security	1,038	4,421	12,470	858	15,324	214	176	34,501
		Insecurity	2,961	17,760	42,233	2,327	49,711	1,600	394	116,989
(6) Autonomy & Sovereignty vs. Foreign Influence [LN,WI]		Libyan Nationalism	846	2,651	7,266	185	9,171	564	90	20,773
		Western Influence	265	1,295	6,034	571	9,346	161	32	17,704
(7) Culture [SC, RIS]		Secular/Liberal	112	871	4,594	538	7,812	147	11	14,085
		Religious/Islam /Sharia	327	1,438	6,031	1,347	7,569	267	40	17,019
(8) Well-being [PGL, PUU]		Prosperity/good quality of life	128	611	1,018	29	2,227	14	12	4,039
		poor/ under-employed/ unemployed	192	982	2,735	222	3,410	30	23	7,594
Total			**10,177**	**43,147**	**125,251**	**8,299**	**168,610**	**3,777**	**1,392**	**360,653**

Table 4.4: Social Dimension Tracking of Tweets volume of Uses due to the Mobilization

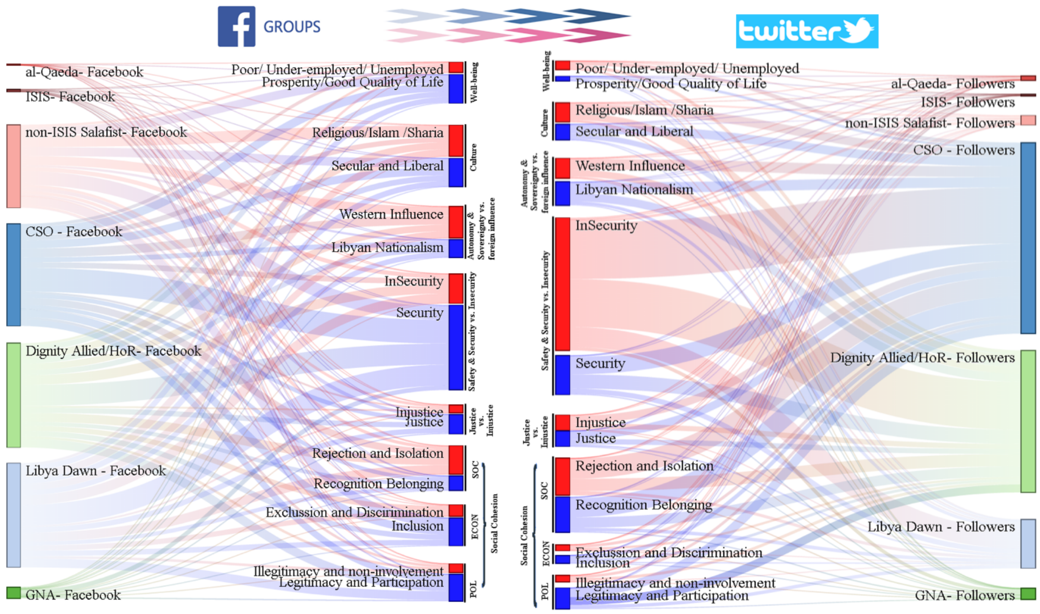
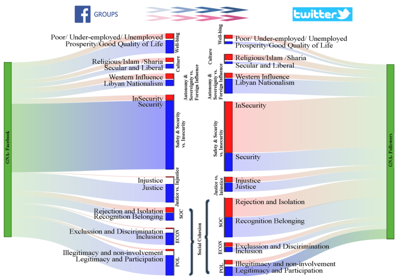
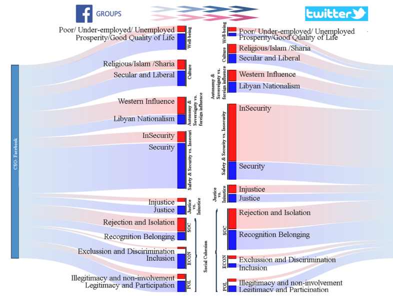


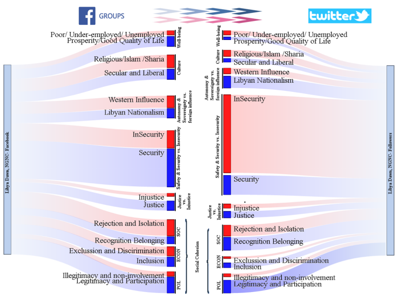
Figure 4.5: A Summary of the 7 Cohorts Mobilization for Affiliated Users. The Flow of Mobilization Mainly Originated from Left (Source as Facebook Pages) and Its Reflection Expand to the Right of Users as Twitters Accounts (Destination). The Chord Thickness Represents the Volume of the Flow.



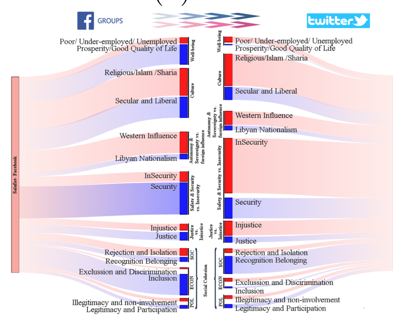
(a) GNA



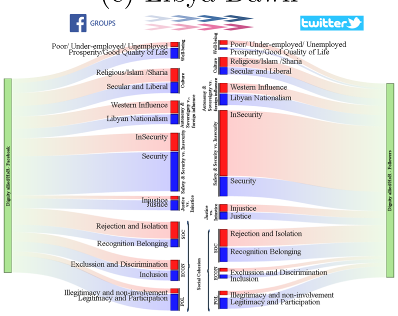
(b) CSO



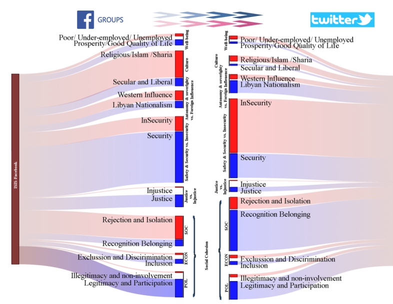
(c) Libya Dawn



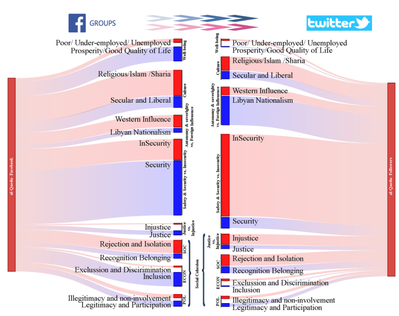
(d) Salafists



(e) Dignity Allied forces



(f) ISIS



(g) AlQaeda

Figure 4.6: A Summary of the 7 Cohorts Mobilization for Affiliated Users Separately. The Flow of Mobilization Mainly Originated from Left (Source as Facebook Pages) and Its Reflection Expanded to the Right of Users as Twitters Accounts (Destination)

4.8 Chapter Conclusion and Remarks

Social media can be used to extract ideologies and issues. However, these issues can be mixed when presented in a text which sets a challenge to make meaning out of them. This study proposed a social model and its dimensions, aiming to gain more insight and track volumes. It also showed how to build a learning model for Facebook content to detect the ideology of a post, and demonstrated the performance of classifier as a multiclass learning model. The same model was subsequently used for tweets. This study also formulated the Sankey diagram as a visualization method to chart ideologies and issue flow from Facebook to Twitter based on each model dimension/category, and explained these phenomena. In the following chapter, this study show how to use a discriminative model to automate detection tweets from a faction (i.e., GNA, Dignity Allied, Libyan Dawn, etc.) not only regarding issues and ideologies as a general form, but also to detect those that pose adversarial level yielding another opposing camp to follow up and react in a temporal scheme.

Chapter 5

DETECTING ADVERSARIAL POLITICAL PHRASAL AND CLASSIFY THEM INTO CONTENTIOUS NON-CONTENTIOUS

5.1 Introduction

The focus of this chapter is to detect adversarial phrases. In Chapter 4, this study showed one aspect of analyzing frames in how to map them (when they are presented as bigrams) into eight dimensions, so that the ideology volume flow from one mean of communication (i.e., Facebook) to the second mean of communication (i.e., Twitter) is shown. This chapter, on the other hand, presents another political frames analysis by relating frames to political adversarial factions, wherein one camp creates some content that may elicit an opposing camp to do one of the following: (1) respond or (2) ignore, but a temporal factor can be augmented to explain better one of the most perspectives of political communication which is the adversarial frames that can reflect communication strategy of each rival opposing camp. The proposed model of this study is an extension of the co-authored work at [11] as its aim is to identify the frames which arouse a contentious reactionary response and the other which does not- called ignored, among antagonistic parties. In [11], forum posts and some related website articles from Slovakia were studied, curated per organization whether Radical or Liberal, whereas this study extends my part contribution of LDA distance based measures and distributions of similarities implementation and building a context for a learning model that includes instances of structured features and their labels. This work aimed to cover social media networks allowing more topological features derived from OSM (Online Social Media) to be added into a learning model as well the spikes detection was approached differently due to the nature of data .Earlier work, I contributed in, [11] the spikes detection was motivated by the hashtag breakout study [15], and that former work have used later in different other studies (i.e. [59, 13]). As mentioned in Chapter 4, two rival groups in Libya were selected: (1) Libya Dawn, which fosters a movement approach that aligns with Muslim Brotherhoods subversive changing narrative that does little regard national identity and interest in a favor of global political Islamic movement (2) Dignity Allies, which adopts the doctrine of the army, and aims to protect national interest as well as fight insurgent political groups. Both vie for power and compromise for a narrow interest partisan sake that consequently yielding to disintegration among the three bodies of political power: legislative, executive, and judicial along with extreme lacking both cooperations toward national interest and seriousness in tackling imperative looming Issues. Such political turmoil has been bolstering debate in social media, where major players and iconic figures become polarized and take sides in the upheaval. They aim to mobilize and influence the public, pushing for ideologies and political agenda frames through the means of social media. This involves a systematic approach to obtain insight into comparative and latent temporal discriminative analysis, which can reveal strategies of cohorts and their stance towards paramount issues.

5.2 Prominence Issues and Idiologies Microdynamics

It is crucial when it comes to explaining the macro level salient issue qualitatively to dig deeper into micro-level qualitative issues that spur political debate among groups in Twitter. We aim to leverage a computational model that can detect those phrases generated by a group that drives an opposing group to either reacts or ignores. With the advent of automated content analysis tools, Subject Matter Experts (SMEs) have been able to conduct various discourse studies on large-scale data collections. These efforts are also extended to research and create social, behavioral models and handle its analysis to serve better social context understanding. Socio-political and computation scientist have been developing tools to enable the systematic study of textual and user-based content, propose hypotheses and examine them. One of the earliest work [9] formulated a goal taxonomies concepts:{purpose, action, state} and used artificial intelligence to perform computations on text, which lead to a more semantic level work in [117]. Such work has been an incentive to streamline in the political as agent computational framework to build cognitive structure and narrative analysis from the theoretical perspective that lacks practical experimentation (e.g. [33, 34, 35]). For text content indexing and information retrieval and similarity measures are first comprehensively summarized with different weighting approach for the keywords as features [115, 82]. They provide varieties of numeric representation for the Vector Space Model (VSM) useful for various hypotheses and applications such as clustering and classification.

On the comparative analysis side, one of most analysis work done related to dynamic partisan dynamic in political discourse by [95, 80, 76]. [95] developed a probabilistic Bayesian model to identify features (words) analyze conflict between republicans and democrats in the U.S. Senate. Their probabilistic approach is to identify features (words) that capture partisan dynamics and analyze conflict between the two political sides. Thus, similar to my contribution in [11] which its data are collected from forums and blogs posts, the contribution here is an extension, of previous part and to allow model to scale up for data curated from Twitter, and follow a new approach of detecting spikes based on density where its parameters are data dependents- More to follow in the methodology section.

5.2.1 Discriminative Classifiers and Generative

Discriminative and generative classification models agree on a principal goal to give an X as an instance a label y, yet they differ in approaches. For instance, the generative goal is to determine the joint probability of $p(x, y)$ as an intermediate step to inference x's label by estimating $p(y|x)$ through Bayes' theorem $p(y|x) = p(x, y)/p(x)$ which require to fit a classifier model such as Gaussian mixture model, Hidden Markov Model, or Latent Dirichlet allocation.. On another side, discriminative classifiers which are more simple such as **logistic regression** and **Support Vector Machine** alternately try to shortcut the classification learning problem by directly estimating the posterior probability $p(y|x)$ without the need to the middle steps such as $p(x, y)$. From the practical point of view, discriminative classifiers with text are widely used since x can be as simple as a bag-of-words (BOW) $n - grams$ and shows plausible results in various text mining concerns about classification or categorization. This does not imply discriminatingly made generative irrelevant for text application, generations can be best applied to more rigorous applications such as language translation and topic modeling and other probabilistic approaches. In the previous chapter (i.e. chapter 4) we have used the discriminative classifier to identify ideologies and issues and to discriminate features itself. We also will use the generative classifier to obtain topic distribution for each spike as a way to decompress text. Next section will address this more thoroughly.

5.3 Methodology

A highlight step of collecting and structuring data before delving into the methodology is listed as follows:

1. Tweets are indexed and retrieved by each dimension of the social model. Each dimension's related tweet will be handled separately for the training model.

2. Each instance (data point) ingested to the learning model has to include: daily volume (i.e. a rise in the volume of tweet) per dimension.Following conditions has to met where variables are labeled on Figure 5.1a for a consequence spike condition:

   (a) Time constraints: $t_1 < t_2 < t_3$

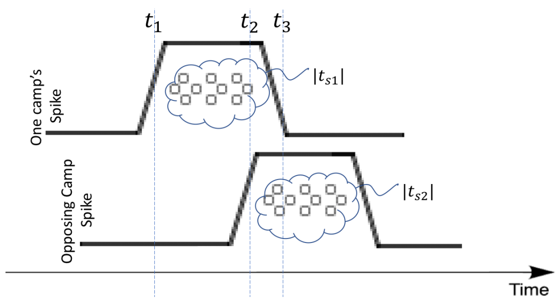
   (b) Time constraints: $(t_3 - t_1) < \beta$

   (c) Tweet Volume: $t_{s1} > \alpha_1$ and $t_{s2} > \alpha_2$

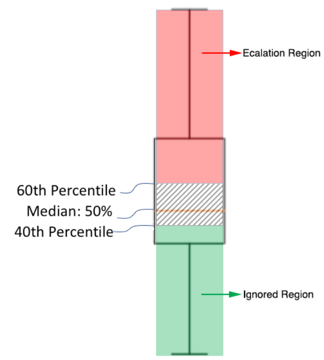
   $\aleph_1, \alpha_2$ and β are set during the experiment after examining distributions of spikes.

3. A binary label is assigned to each opposing consecutive spikes as either **reactionary** episode or **ignored**. The assignment is based on similarity measure or closeness measure of features presented in the text of tweets. Determining the threshold is done by examining the distribution and setting that suitable thresholding in a manner to allow clear discriminative data falling into the two classes: **reactionary** (i.e. highly similar/close to each other in the similarity measure as both spikes have high commonality and addressing similar topics/issues) or the **ignored**. section 5.3.1. Figure 5.1b shows the conceptual distribution of a similarity measure. We consider any data-point fall into the red distribution range (i.e. $[60thpercentile, maximumvalue]$) to be a reactionary, meanwhile, the green to be ignored and lastly if none of the two conditions are met, we opt to exempt related instances from the dataset allowing for further distance to

avert model perplexity and to allow for more distinguishable instance between the two classes for better model durability. It is experimentally determined and a "trading of" arises between the adequate number of samples fed to the model versus more clear discriminative pattern favored given to a learning model.



(a) Two consecutive spikes based on density that represent a single data-point in a learning model



(b) Similarity Value Distribution among all consecutive spikes, allowing to set the right thresholds to determine each spike label

Figure 5.1: Two conceptual approach: (1) Detecting spike (2) Labeling Data instances of consecutive spikes.

5.3.1 Reactionary and Ignored Spikes detections

With the spike density based, we are able in detecting prominent tweet volume spikes that reflects high attention by users linked to their posting tweet behavior. "**Reactionary**" spikes that trigger a response from the other political side, or (ii) "**ignored**" spikes that lead to no response. Generally, a debate can be described as "a formal discussion on a set of related topics in which opposing perspectives and arguments are set forth' [1] '. Therefore, a simple general hypothesis can be outlined which facilitate capturing these dynamics analytically based on the following observation.

[1]Oxford Online Dictionary

1. The two political sides incline to discuss their views of issues and pushing ideology showing multiple vantage points on issues that can be captured on the dimentions of the social model.

2. Both sides likely comment on other's points of view by criticizing to or even could be stretched to degrading and demonizing other views, yet they then connect these points back to their point of view and their perspectives in a way to show that they have a legitimate point of view.

To capture this dynamics analytically. We can perform spike categorization as either reactionary or ignored based on similarity among distribution of the two consecutive spikes from opponent political sides. Next, we need to determine correspondence of topics between a pair of consecutive spikes from opposing political sides which indicates whether the two raising similar topics indicating the second spike is a reactionary of the first one. and similarly if no high similarity presented that would indicate the second spikes is an ignored of previous opposing camp;'s spike. Hence, we utilize the similarity measure Distance, $Sim(S_A, S_B)$ between two univariate distribution $\boldsymbol{A} = a1, ..., a_n$ and $\boldsymbol{B} = b_1, \ldots, b_n$ derived from topic distribution Then the similarity is measured by either equation 5.4 which is a conversion form of Kull-Leibler Divergence 5.3 or 5.1 show a two options further explained in proposed models coming next.

5.3.2 Proposed Models

For any machine learning task, we have to provide a context (i.e. instances and their labels) such that a model can learn from a mapping fuction from instances to given labels. We used three models to build the context, particularly the labels of spikes in a semi-supervised approach and then used the model to detect their phrasal frames (i.e., most informative features the model learned to discriminate between escalated or ignored spikes). Phrasal frame tend to be reactionary on the other opposing camp. These models are listed as follows: In our baseline model (Vector Space model), we directly modeled the similarity approach by using the cosine similarity over spikes frequent keyword vector representation of E and R, without requiring a *lower dimensional space* representation of the data, e.g. inferring topic distribution LDA or word embedding document to vector. The similarity measure is Cosine similarity which can be used through computing inner product of between two vectors and its resultant value is granted to be bounded between 0 and 1 (i.e. 1 is indicating the highest similarity and 0 the lowest)

$$Sim(S_A, S_B) = Cosine(A, B) = \frac{A.B}{\|A\|_2 . \|B\|_2} \tag{5.1}$$

Previous Cosine similarity will be used with two of the proposed models, but with an LDA model we will use the Kullback-Leibler distance measure after covering it to a similarity measures

Next, Kullback-Leibler Distance is another measure that fit right with vectors expressed as discrete probabilistic distribution (i.e. $\sum_{i=1}^{|V|} p_i = 1$). Its distance can be utilized as a similarity measure, and following the non-symmetric in equation 5.2 which can be used, but we used the symettirnc version to assure this distance measure can be converted to a similarity measure.

$$D(A||B) = \sum_{x \in X} A(x) \log \frac{A(x)}{B(x)} \tag{5.2}$$

Bigi et al. [28] proposed a symmetric version which will we use it through out the experiment.

$$D(A||B) = \sum_{x \in X} \left(\Big(A(x) - B(x) \Big) \log \frac{A(x)}{B(x)} \right) \tag{5.3}$$

For similairy, we invert the distance as follows:

$$Sim(S_A, S_B) = 1 - D(A||B) \tag{5.4}$$

Next, three models will be used as semi-supervised learning approach in labeling our data, namely: Lexical Model based, Word Embedding based, and an LDA based.

5.3.3 Lexical Model

This models is represented by is a sparse vector of frequent keywords. We computed all distances spikes using eq. 5.1, and by thresholding we determined the labels of spikes whose measure is larger than or equal to the mean, indicating an 'escalated' spike from the opposing camp, otherwise they are labeled as 'ignored'.

5.3.4 Word Embedding

Word embedding utilizes neural networks to encode the context into a denser lower dimensional space. It is a highly effective method in capturing semantic relations where each document is represented by a real number vector such that similar documents are closer to one another than dissimilar documents in a geometric space. We employed the Paragraph Vector Distributed Bag of Words (PV-DBOW) proposed by [94, 85] to infer the real number vector of a document (a.k.a. doc2vector). After training the PV-DBOW model over our corpus, we infer vectors for each spike and computed all distances using eq. 5.1. Cosine fits here better as each document is represented by a point in a geometry space, and through thresholding approach, we determined labels of spikes.

5.3.5 LDA

The LDA model can be viewed as a three level Bayesian probabilistic model to learn distributions of topics over documents and words. After training an LDA model, we infer topic distribution for each spike's topics. Next, we determine the labels of spikes based on the Kullback-Leibler (KL) measure, which captures the divergence of distributions between two consecutive spikes (eq. 5.3).

5.3.6 Text Decompression through Topic Models

Text decomposition is primarily approached through generative models. In many other machine learning problem, Principle Component Analysis (PCA) or Linear Discriminant Analysis (LDA) are widely used to reduce dimensionality but as linear reduction. This linearity could potentially lead to high bias and significant distortion where text poses immense variability as well as high sparsity and nonlinear patterns. Nevertheless, text decomposition can still be approached but differently to tackle its curse of dimensionality by looking into probabilistic soft membership of a lower dimension through transformation. This transformation projects data points into a lower-dimensional space or what so called "subspace." Assume that we have **term-docuement-fearture matrix** A of $(m \times n)$ dimension where m is number of documents, and n is number of features (i.e. terms sequence). The goal is to project data into a lower dimension, say d where $(d << m)$ which help in looking beyond the words as merely lexical level to higher level semantical where terms and phrases can be related to mixture of topics. various models have been proposed in the literature where each assumes its observed variables and other hiddens (such as topic distribution of a given keyword), and then alllow a computational model to estimate hidden variables from observed one in the corpus.

One of most popular topic model is the Probabilistic Latent Semantic Analysis (pLSA) [68]. It is a generalization of LSA that assumes topics to be orthogonal, but pLSA relaxed this condition and approached it through a generative model. On the another side Latent Dirichlet Allocation (LDA) came after and enhanced the quality of obtaining results if compared with the pLSA due to harnessing Dirichlet priors that impede overfitting. The two intends to find lower dimensions of topics/concepts that go beyond lexical sparse term-document matrix to higher semantic mapping after estimating hidden variable of the selected model. For the framework, we choose LDA over pLSA, and the following would highlight hidden latent parameter of the model.

LDA is a generative probabilistic model that assume documents are a mixture of topics, where a topic is a probability distribution over words. It is based on a Dirichlet distribution that has been categorized as a family of continuous multivariate probability distribution parameterized by α and η as priors as well as number of the topic, K. This LDA is a probabilistic unsupervised learning model and can be explained by a graphical form, Figure 5.2. That figure shows plates, arrows, and nodes where plates represent replication; Nodes represents variables, and arrows represent dependencies. We denotes T as number of tweets; N as number of words per a given tweet;K as number of topics selected for the model. It is worth mentioning that K is highly dependent on the data nature and some intuitions expectations about data corpus with its most prominants topics. Pragmatically, most practitioners sets K to any value within 100 range. LDA approach it by approximating the posterior conditional probability globally as an optimization problem. Next, we leveraged the estimations of unobserved value

- For each topic (set of related keywords) in the K dimension, we extract and rank most top keywords pertaining to each topic.

- For any collection of tweets (even unseen in training set), topic distribution $\theta_{1:K}$ can be inferred through utilizing corresponding words in $z_{(1:K,1:T)}$.

$w_{t,n}$	the specific word ith in tweet t which is the only observed variable in the model.
$z_{t,n}$	per-word topic assignment; the topic distribution for ith word in tweet t can be inferred.
θ_t	per-tweet topic proportions; the topic distribution for tweet t.
β_k	per the whole tweet collection topic distributions
α	parameter of Dirichlet prior of per-tweet topic distributions
η	parameter of Dirichlet prior of word distribution of a topic
N	number of words for a given tweet
K	number of topics
T	number of tweets

Table 5.1: LDA Nomenclature

The LDA has many hidden variables (i.e. $\beta_{(1:k)}, \theta_{(1:k)}, z_{(1:K,1:T)}$) to be estimated by some variables observed which is $w_{(1:K,1:T)}$. Approximating the posterior conditional probability $P(\beta_{(1:K)}, \theta_{(1:K)}, z_{(1:K,1:D)}|w_{(1:K,1:D)})$ globally as an optimization problem. Next, we use the posterior to sort of words per topic from highest to lowest value. Gippps sampling with conjugate priors are employed for the posterior inference. The generative process can be captured by the following procedure until converge:

1. For each topic k, draw a distribution over words $\beta_k \sim \mathbf{Dir}(\eta)$
2. For each tweet t in T
   (a) Draw a vector of topic parameter $\theta_t \sim \mathbf{Dir}(\alpha)$
   (b) For each word n
      i. Draw a topic assignment $z_{t,n} \sim \mathbf{Mult}(\theta_t), z_{t,n} \in \{1, \ldots, K\}$

ii. Draw a word $w_{t,n} \sim \mathbf{Mult}(\beta_k), w_{t,n} \in \{1, \ldots, N\}$

After LDA variables estimated, we can infer topic distribution to any collections of tweets even it has not been seen by the LDA, as well as we can learn each topic top pertaining keywords.

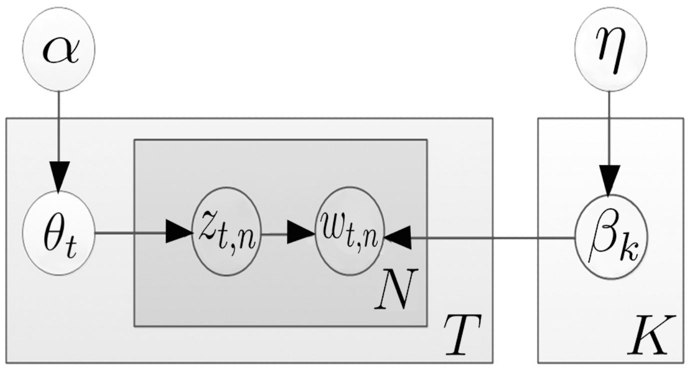


Figure 5.2: Plate, Nodes, Arrows Notation Representing for the LDA model.

5.3.7 Labeling Instance and Thresholding Determination

The framework categorizes spikes as: 1) **reactionary/ignored eastern side** spikes, 2) **reactionary/ignored western side** spikes, and identifies contentious phrasal frames driving the escalation from each political side. We detect **reactionary/ignored** spikes as follows:

- Utilize LDA topic distribution over all post tweets content to reveal latent topic distributions and its highest ranking keywords for each topic (default setting is 100 topics/bins)

- For each dimension and each political side, we identified corresponding spikes based on the spike density based.

- Next, each spike is labeled as "**reactionary**"/"**ignored**" based on the correspondence of its topics distributions to the topics distributions of the following spike from the other political side, Figure 6.5 (a) and (b). We used to measure Kullback-Leibler Distance between two distributions of **LDA** topics for the two spikes. Then, we employee the similarity in equation 5.4 for the LDA model.

5.4 Experiment

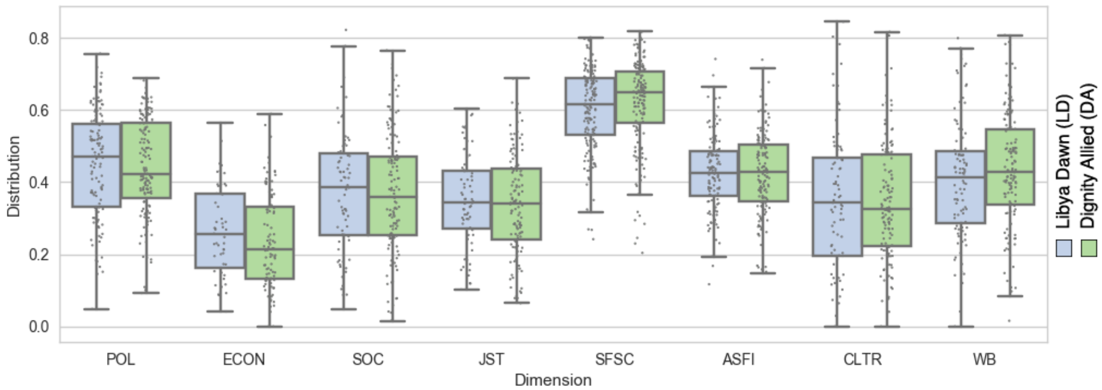
As preluded before, there are seven coalitions actively operating in cyberspace and on the ground, yet for this analysis, the two most belligerent factions were selected: Libya Dawn and Dignity Allied to uncover their communication strategy and understand their tactics in relation to escalated spikes. Tweet volume is updated daily, yet consecutive volume spikes constraints are utilized through a more granular time unit in seconds, since many attention models indicate peoples attention and reactions within social media are usually within the first two hours, but its reactionary effects can last much longer. Both studies in [21, 139] show that the average time to receive news events that are trending and are related to events is roughly 2-3 hours after the fact, and could last to 128 hours, after which a new phase of volume decaying starts.

After collecting the consequences spikes, the semi-supervised learning approach was leveraged by using similarity measures to consider distribution and define related threshold for instances labeling. The following three models were used: (1) Lexical (2) Word Embedding, and (3) the LDA. Figure aaa shows the distributions of the distances of the 8 dimensions, so thresholding values of labeling can be determined on the fly while the experiment was conducted.

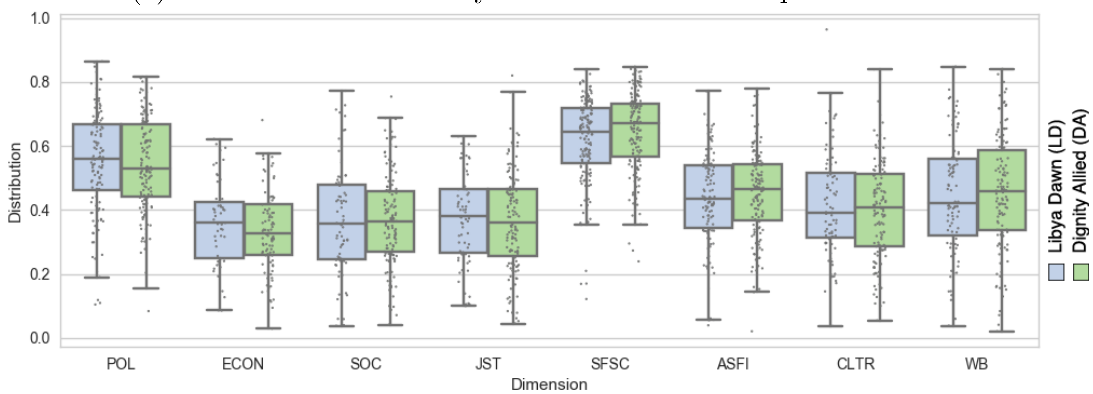
The window density-based per tweet spike was set to 16 tweets for Dignity Allied and to 8 to suit data volume proportions of both camps since the volume of Dignity Allied is approximately double that of Libya Dawns after examining both distributions related to the eight dimensions. Figure 4.3 shows distributions per model and per dimension as each dimension can be varied where right threshold value must be determined. Another observant, the SFSC dimension, is skewed to the top indicating a large number of tweets deliberating on security/insecurity topics, events, and concerns.

Next, after creating the context (i.e. availability of training data instances along with true labels) that obtained by the semi-supervised approach. A machine learning task was used to: (1) detect reactionary class (2) learn the most prominent frames that would elicit one camp reaction to an opposing camp for each dimension individually , eight instances of classifiers corresponds to the eight dimensions are trained separately to detect "**reactionary/ignored**" classes.

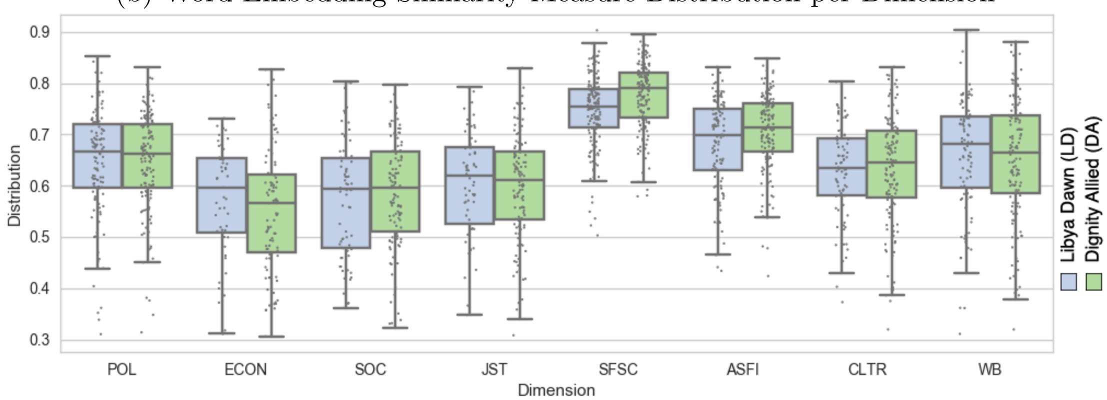
Next, a classifier was run on each dimension and the eight classifiers accuracies were reported for each model and per dimensions. All classifiers share the same data bigrams of tweets content and other topological network features extracted from the graph, yet the difference of data per models is the label of instances as it derived from similarity measures. The results, in Table 5.2, show that lexical based scored the lowest performance as it potentially did not capture the semantic-level pattern of the other models. The embedding model always scores higher than the lexicon and therefore this performance was set as a lower bound to optimize LDA parameters to constraint the LDA to reach higher values. After training and optimizing the LDA hyper-parameters bearing in mind the text embedding is highly competitive due to utilizing the neural networks as well as words order comes to accounts as if compared Bag of Words for the LDA. After training and utilizing the LDA, outperforming results were achieved on both the lexical and the embedding models when used as data instance labelers.



(a) Lexical Model Similarity Measure Distribution per Dimension



(b) Word Embedding Similarity Measure Distribution per Dimension



(c) LDA Model Similarity Measure Distribution per Dimension

Figure 5.3: 2D Example of Ordinal Data

Consecutive Spikes	Second Spike Type	Dimensions	Lexical Pr	Lexical Re	Lexical F1	Embedding Pr	Embedding Re	Embedding F1	LDA Pr	LDA Re	LDA F1
LD to DA	Escalation	POL	.83	.8	.81	.84	.88	.86	.91	.88	.9
		ECON	.9	.86	.88	1.	.95	.98	1.	.86	.92
		SOC	.92	.79	.85	.96	.83	.89	.96	.9	.93
		Justice	.82	.85	.84	.92	.85	.88	.92	.85	.88
		Security	.94	.8	.86	.84	.8	.82	.94	.82	.88
		Autonomy Sovereighty	.8	.58	.67	.82	.65	.72	.95	.75	.84
		Culture	.8	.85	.82	.93	.82	.87	.94	.91	.92
		Well Being	.74	.79	.76	.79	.72	.75	.86	.93	.89
	Ignored	POL	.8	.84	.82	.87	.84	.85	.88	.92	.9
		ECON	.86	.9	.88	.95	1.	.98	.88	1.	.93
		SOC	.82	.93	.87	.85	.97	.9	.9	.97	.93
		Justice	.85	.81	.83	.86	.93	.89	.86	.93	.89
		Security	.83	.95	.88	.81	.85	.83	.84	.95	.89
		AutonomySovereighty	.67	.85	.75	.71	.85	.77	.79	.96	.87
		Culture	.84	.79	.81	.84	.94	.89	.91	.94	.93
		Well Being	.86	.93	.89	.82	.65	.72	.91	.86	.89
DA to LD	Escalation	POL	.85	.89	.87	.81	.81	.81	.85	.82	.84
		ECON	.86	.82	.84	.89	.84	.86	.89	.87	.88
		SOC	.76	.76	.76	.85	.68	.76	.91	.86	.89
		Justice	.84	.78	.81	.81	.78	.79	.89	.8	.84
		Security	.83	.71	.77	.83	.69	.75	.92	.77	.84
		AutonomySovereighty	.75	.84	.79	.79	.72	.75	.74	.79	.76
		Culture	.8	.71	.75	.88	.73	.8	.91	.82	.87
		Well Being	.89	.84	.86	.74	.85	.79	.77	.72	.75
	Ignored	POL	.85	.68	.76	.81	.81	.81	.83	.86	.84
		ECON	.83	.87	.85	.85	.89	.87	.87	.89	.88
		SOC	.76	.76	.76	.73	.88	.8	.87	.92	.89
		Justice	.79	.86	.82	.78	.82	.8	.81	.9	.85
		Security	.75	.85	.8	.74	.85	.79	.81	.94	.87
		AutonomySovereighty	.82	.72	.77	.74	.81	.77	.77	.72	.75
		Culture	.74	.82	.78	.77	.9	.83	.84	.92	.88
		Avg	-	-	.81	-	-	.83	-	-	.87

Table 5.2: Classification Performance over the Three Models and for Each Social Cohesion Dimension. LDA Has Been Optimized, and It Outperforms Both Lexical Features and the Work Embedding Model after Optimization

5.5 Qualitative Analysis

Our trained classifiers models allow us to examine features importance weights and polarity (i.e., reactionary/ignore) and to utilize the learned weight to classify instances. Following two Tables summarize- after data cleaning- reactionary frames per camp and the ignored. Such an approach allows subject matter experts conclude communication strategy in a more systematic style than by looking into micro- level bundles of the word "bigram" to unveil a communication strategy. In Tables 5.3 and 5.4, frames that cause responses from the opposing camp and frames that are ignored, respectively. The bigrams are in Arabic, but a bundle of words was added in English to give some contextual meaning to serve the readers understanding of the frames employment in communication. Both camps agreed on fighting terrorism, and a high level of consensus in communication with regards to the importance of draining its resources was present. Yet, the political rivalry made many insurgencies exacerbate the situation in conflicting areas and areas of influence (i.e., tribal regions and oil fields). Many Libyans perceived GNA as representing the soft power of Muslim Brotherhood, meanwhile Libya Down is the armed branch of that soft power that used as leverage on the point of disagreements with the HoR and Dignity Allied Forces ignoring the suffering of Libyan people and Dignity Allied scarifies in the bloody war against terrorism and their supporters on the ground . Also, they (i.e. GNA) have been allegedly part of secret deal protections and large corruption networks benefiting insurgent groups that protect GNA for narrow poltitical gain, and DA influence places and oil fields that leading to financial depletion of national resources Libyan people fortune for narrow and short-term vision interests.They are also perceived as allies of insurgent militias, causing the unrest and the political turmoil. They also lack on-the-ground legitimacy and are cynically labeled as a Frigate Government that was previously operating from the Libyan coasts and was supported by foreign governments. Nonetheless, they are pushing to make themselves legitimate to all of Libya. Even though the lack of popular acceptance due to the Muslim Brotherhood agenda is viewed as anti-Arabism that is more inclined with Persian agenda aiming to dismantle the cohesion of all Arab nations. GNA is viewed as Muslim Brotherhood Group who are deceivingly calling for democracy, but in practice, they are indulged in political hypocrisies, and their agenda is to marginalize nationalist and become practicing cross-border ideologies that stands against the nation of Libya, which to a great extends aligned with Iranians goals. This view become more promenant after majority of counties in the Arab league, leading them to placed MBs on their counties' terrorists lists.

On the other hand, Dignity Allied and the House of Representatives are criticized by other camps for allegedly crippling the democratic transition and failing all votes to legitimize the GNA government by reasons of minimum ballots due to the lack of quorum, even though there was effort by the international community and the United Nations envoy. Dignity Allied has been recognized in fighting terrorist groups at all levels on the east side of Libya and the most north-west, even though the "stochastic operations" yielded mass casualties in their fights. Dignity Allied have been seen as a strong arm in conquering oil fields controlled by insurgent groups that used to pay a franchise to the GNA in a corruptive deal, which is a prerogative to all Libyans to decide which gave them high respect in their tribal area they rule, and each social dimensions have been explained in Tables 5.3 and 5.4 for easy-read.

This analysis shows that both camps and their followers seemingly and superficially agree on one goal: combating terrorism, yet they delegitimize one another and each faction has sought national aid in their political and on the ground fights, regardless of all consolidated efforts in restoring Libya after the Qathafi fallen regime.

LD phrasal frames eliciting response by the DA	Dimension	DA phrasal frames eliciting response by the LD
The only legitimate - The legitimacy of the Parliament - Amid Confusion - Election Crisis - Penal Code -Islamic Awakening - DA Military Rule - Hafar: New Qathafi's - One United Libya	**POL**	Lies and Forgeries - Election Fraud - Falsification of the fact - Falsifying Citizenship - Extorting citizens - Muslim Brotherhood Backers - Political Hypocrisy - Illegitimate "Sarraj" - Frigate Government - Federation is Imperative - Western Proxy Government - Not Emanating
Corruption with Oil Fields Protections Deals - Financial Corruption - Marginalization and Impoverishing - Marginalization of the South Region - Central Marginalization - Systematic Theft - Receiving Support from various Arab Countries not Supporting the proclaimed "Democratic Transition"	**ECON**	Corruptions - Wasting Resources on Claiming Islamic Militia in Fighting Again DA - Suspicious Financial Transaction to Islamist Groups - Not Investing in Real Army - Buying Security from Insurgent Groups - Receiving Financial Support from Qatar and Turkey
Penal Code - Society Bodies - Remnants of the Previous Regime - Intransigence of HoR to legitimize the GNA - Separatist Groups - Haftar Gangs Targeting Civilians - Mass Casualty Incidents on a Rise - Suffering of Displaced Persons - Subversive Agenda	**SOC**	Smuggler's Boats - Tribal and Regional Power - Complicity to Foreign Fighters - Complicity to Islamic militia - Failure to build a National Army - No National Doctrine – Complicity to Muslim Brotherhood - Pursuing Neutralization of Nationals and Veterans - Singling out Authority for MB - Fake Democracy - Supporting Colonial Agenda for the Sake of MB Future
Justice Act - Political Isolation Law - Justice and Litigation - Justice and Retribution - Restore Justice - Constitution Equality-Equality Principles - Justice of our Cause	**Justice**	Injustice and New Tyranny - Injustice Against Nationalist - Collusion With Islamist MB members in the Judicial System
Haftar's Allowing Terrorist to Move to GNA Controlled Regions - Restore Stability and Overthrown Haftar's - Libyan Dialogs with all Factions - Haftar Aiming to Rule	**Security**	Injustice and corruption - Injustice to people - Dirty Agenda - Muslim Brotherhood led Government - Secrete Control and Protection Deals with Militias
French Foreign intervention- Russian intervention - Egyptian intervention - External interference - Regional intervention - Veterans Interventions - One Army	**Autonomou Sovereignty**	Fighting Foriegner Millitia - US Interventions Supporting MB - UN envoy interventions - Interventions on HoR - Forcing GNA Legitimacy - Foreign Pressure - Tunisia Interventions
Inculcating Despotism - Fighting Civil Society - Fighting Political Islamists - Wstern Agendas - Westernizing Society	**Culture**	Colluding with Islamic Insurgent Groups - Propagating a Culture of Terror - Working under Wraps to Support Muslim Brotherhood in all Libyan regions
Democracy and Development - Standard of Transparency - Standard of Credibility - Sanctification Across All Society Spectrum - All Background Coexistence	**Well Being**	People Smugglers - Human Trafficking - Health Crises - War and Poverty - Selling Drugs for Operation Funds - Health Deterioration - People Displacement

Table 5.3: Contentions phrasal Frames Eliciting Response

This has led to a deterioration in the social cohesion and has weakened the country, all for accommodating the aspirations of all Libyans. By this work, we have been able to accomplish revealing sentiments of political factions' followers on various aspects of their national conflicts.

LD phrasal frames eliciting response by the DA	Dimension	DA phrasal frames eliciting response by the LD
Election Law- Council Election - Congress Election - The only legitimate - The legitimacy of the Parliament - Amid Confusion - Election Crisis - Penal Code -Islamic Awakening	**POL**	Worker Strikes - Continue Their Strike - Lies and Forgeries - Election Fraud - Falsification of the fact - Falsifying Citizenship - Extorting citizens - Muslim Brotherhood
Development Programs - Stability and Development - Equality and Rights - African Development - Sustainable Feature -	**ECON**	Financial Corruption - Marginalization and Impoverishing - Marginalization of the South Region - Central Marginalization - Systematic Theft
Muslim Community - All Community Activist - Coordinating Community - Society Bodies	**SOC**	Suffering of Displaced Persons - Smuggler's Boats - Tribal and Regional Power - No National Doctrine - Complicity to Islamic militia
Justice and Litigation - Justice and Retribution - Restore Justice - Constitution Equality- Equality Principles - Secretriat of Justice - Justice of our Cause - Justice from Qathaif's Backers	**Justice**	Injustice and marginalization - Injustice and tyranny - Injustice against Nationalist - Interconnected and Bonded
Stabilty and Security - Restore Stability - Stablility and Development - Libyan Dialogs with all factions - Skhirat dialogue - National cohesion	**Security**	Corrupted scene - people suffer prejudice - Prejudice against Libyans - MBs are protected by the government
Libya United - Protecting National Interest - Counter-Terrorism Cooperation - Libyan Must Be Notified with Foreign Forces	**Autonomou Sovereignty**	Libya United - Protecting National Interest - Counter-Terrorism Cooperation
Democracy Transition - Modern & Conservative Islamic Country -	**Culture**	Eradicating Terrorism - Concept of dependency to a political group (MB) - Pervasive Corruption Culture in the GNA - GNA Members LACK Culture Understanding - Lack Understanding Libyan Culture - Secular Trends Aligned with MB's Agendas
Spatial Development - Local Development - Democracy and Development - Prosperity - Standard of Transparency - Agricultural Jobs	**Well Being**	War and Poverty - Corruption and Drugs - Evil and Corruption - Principles Deterioration - Hunger Crisis

Table 5.4: Political Frames Not Necessitate a Response by Another Camp

5.6 Chapter Conclusion and Remarks

In this chapter, we showed how computational phrasal frame can be leveraged on phrases detection per dimension related to the social cohesion theory, and how to detect contentions per dimension. It has been proven that such an approach can yield a more precise analysis to be conducted by subject matter experts in social media networks. Three methods have been compared: Lexicon, Word Embedding and the LDA. These have shown that word embedding can be employed as a lower bound performance for the LDA optimization. The LDA was leveraged to explain each social cohesion dimension content and utilized its learned latent variables for a semi-supervised approach. This approach was used in labeling instances and allowing a learning model to predict the reactionary/contention phrasal frames. The experiments and model modification prove the proposed model can be extended with modifications to accommodate short messaging "microblogs" in social media as having successful results on longer textual corpus articles.

Chapter 6

COMMUNITY TEMPORAL CLUSTERS AND RANKING

6.1 Introduction

Capturing the value of the social media stream and deriving powerful insight of community characteristics through temporal dynamics of users shifts, issues discussed, and media shared are the essence of effective real-time comparative analysis. This chapter shows how batches of clusters are continually unfolding for each snapshot captured sequentially that necessitate expanding contents and profiling their characteristics. However, the whirl of voluminous data pose a significant challenge and even become beyond human capacity. Myriad blocks of content are automatically generated carrying propagated ideologies, resonated information where all would put an enormous burden on comparative analysts' shoulders. Analysts are believed to bring the most accurate summaries capturing the highest potential underlying information to ensure relevance and freshness without drift after they savor the unfolding spectacle in data, whereas their tasks are constrained by stipulated short timeframes. Community detections will be employed considering different features leveraged from Twitter to bring up those clusters for each snapshot. The snapshot batches are governed by the timeframe having a start and end time where the time do- main is slotted and sliced into consecutive batches. Then, the analysis can proceed with summarizing communities with the help of Augmented Artificial Intelligence (IA) that aims to enhance human summarizing to better capture issues and trends within most important contents, but not to substitute human involvement, after ranking contents based on various features. This chapter will propose a ranking method and statistical evaluation to show the significance of the model in ranking contents which in part can effectively help in understanding the presented text reach conclusions by the end of each snapshot of groups, dynamics, messaging and tactics.

6.2 Problem Statement

For a given number of community clusters, as consecutive snapshots, of tweets and their originated users. Comparative analysts and subject matter experts aim to better understand underlying information by sifting through clusters and aim to connect latent patterns by capturing significant temporal dynamics to help create meaning out of the presented clusters. Thus, a train set of ordinal y label was built, representing Liker scale value from (1 to 5) where 5 has the highest relevance, and 1 has the lowest. where will be employed to survey tool responses and can be generalized accurately would mimic comparative ranking responses. The rank can be obtained by answering some indicator questions, which are well-known in media analysis e.g. Odijk1 et al. [103].

It is important to consider an input space $A \subset \mathbb{R}^n$ n with the objective being represented by instances vector $a = (a_1, \ldots, a_n)^T \in \mathbb{R}$, where n denotes the number of features. Assume there is also an outcome space Y = r1, . . . , rq with ordered ranks$r_q \succ r_{q-1} \succ \cdots \succ r_1$. The symbol $\succ$ denotes the ordering between different ranks. Additionally, suppose an i.i.d. sample $S(a_i, y_i)_{i=1}^{l} \subset A \times Y$ is given. Assume a model space $H = \{h(.) : A \longmapsto Y\}$ of mapping from objects to rank. Function h induces an ordering on the input space.

6.3 Temporal Communities

The ranker model is proposes in an aimed to help experts swift through data community snapshots and automatically rank content to provide the most probable material for examining and summarizing . In the temporal community detection, a content and endorsement filtered connectivity model was applied to find community clusters of politically Arduent users in pure political communities. This model was proposed by [104] and was used with multiple consecutive snapshots for the Twitter stream, as shown in Figure 6.1 to keep track of user shifts. This would be based on the proposed algorithm by [104], taking $\mathbf{X_{u,w}}$, $\mathbf{W_{sim}}$, α, β and indicating cluster assignment for the W and U clusters that were solved by the optimizing objective function formulated in equation 6.1. The list of nomenclature is as follows:

- $\mathbf{X_{u,w}}$: user $\times$ word
- $\mathbf{W_{sim}}$ word similarity matrix
- α user connectivity regularization parameter
- β word similarity regularization similarity parameter
- U cluster assignment matrix of users, user $\times$ cluster
- W cluster assignment matrix of words
- $\mathbf{L}_U$ laplication matrix of user connectivity matrix
- $\mathbf{L}_{w_{sim}}$ laplication matrix of word similarity matrix

$$\begin{aligned} J_{U,W} = \|X_{uw} - UW^t\|_F^2 + \alpha Tr(U^t \mathbf{L}_U U) + \beta Tr(W^t \mathbf{L}_{w_{sim}} W) \\ \textbf{such that } U \geq 0, W \geq 0 \end{aligned} \tag{6.1}$$

Figure 6.1 charts the dynamic shift of users as result of the applied temporal community clustering. Additionally, a color code divergence of **GNA** sentiment was applied wherein the extreme red indicates opposing (i.e. supporting **HoR**), meanwhile the green implies full support of the GNA. Red-coded, on the other hand, does not necessarily imply violence or terrorism, but rather, it can be any general negative sentiment expressing the indignation toward GNA, Muslim Brotherhood Organization and their ties to militant groups, or pro-army located in the east, or others supporting only the Islamic State (e.g. Al-Qaeda and ISIS and their local arms in different regions). The algorithm is run either weekly or when some event/breakout occurs. The timing would also be governed by various factors such as the nature of data, the number of subject matter experts, and the computation power.

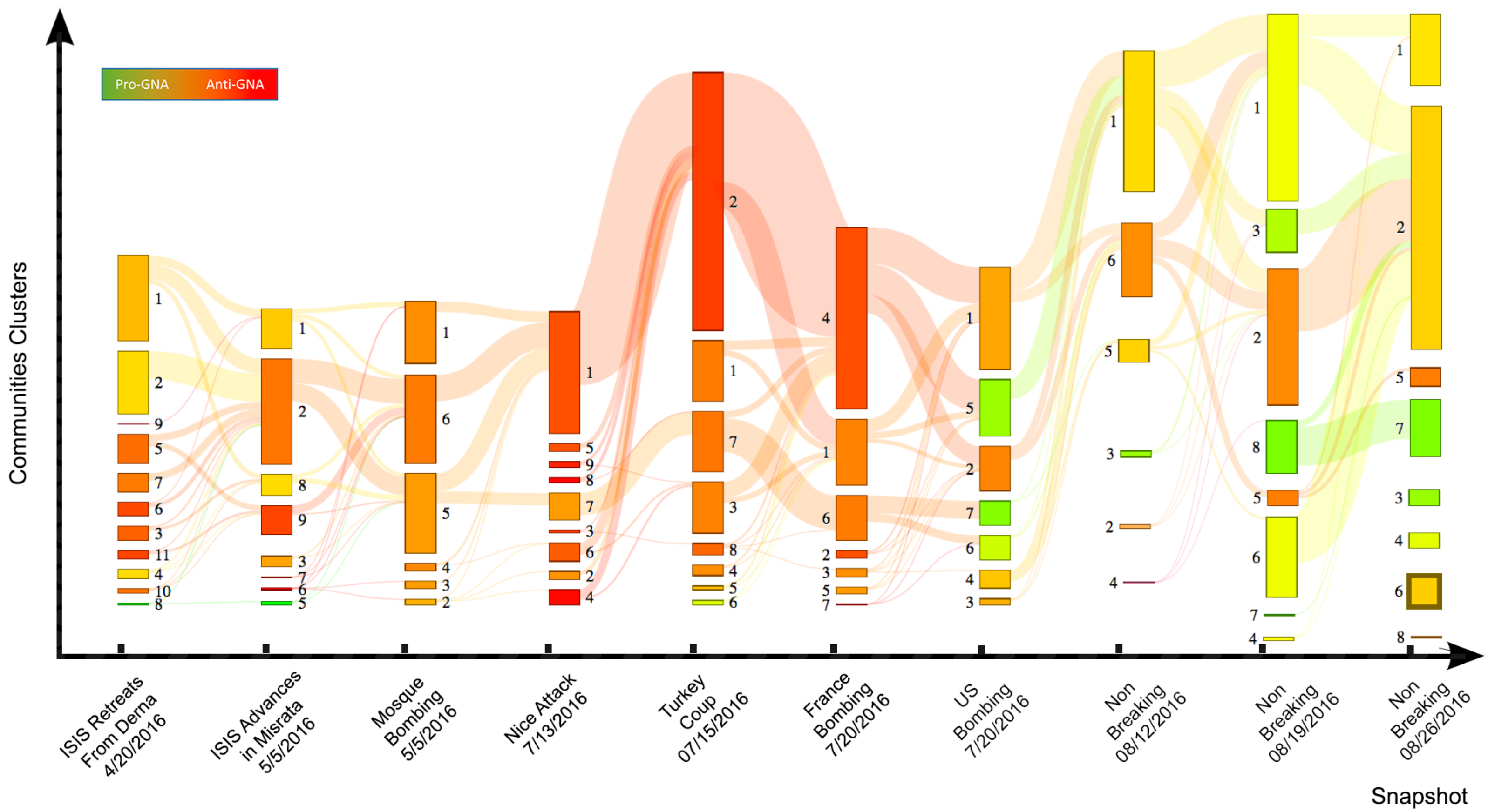


Figure 6.1: Key Events, Communities and Sentiment Trends.

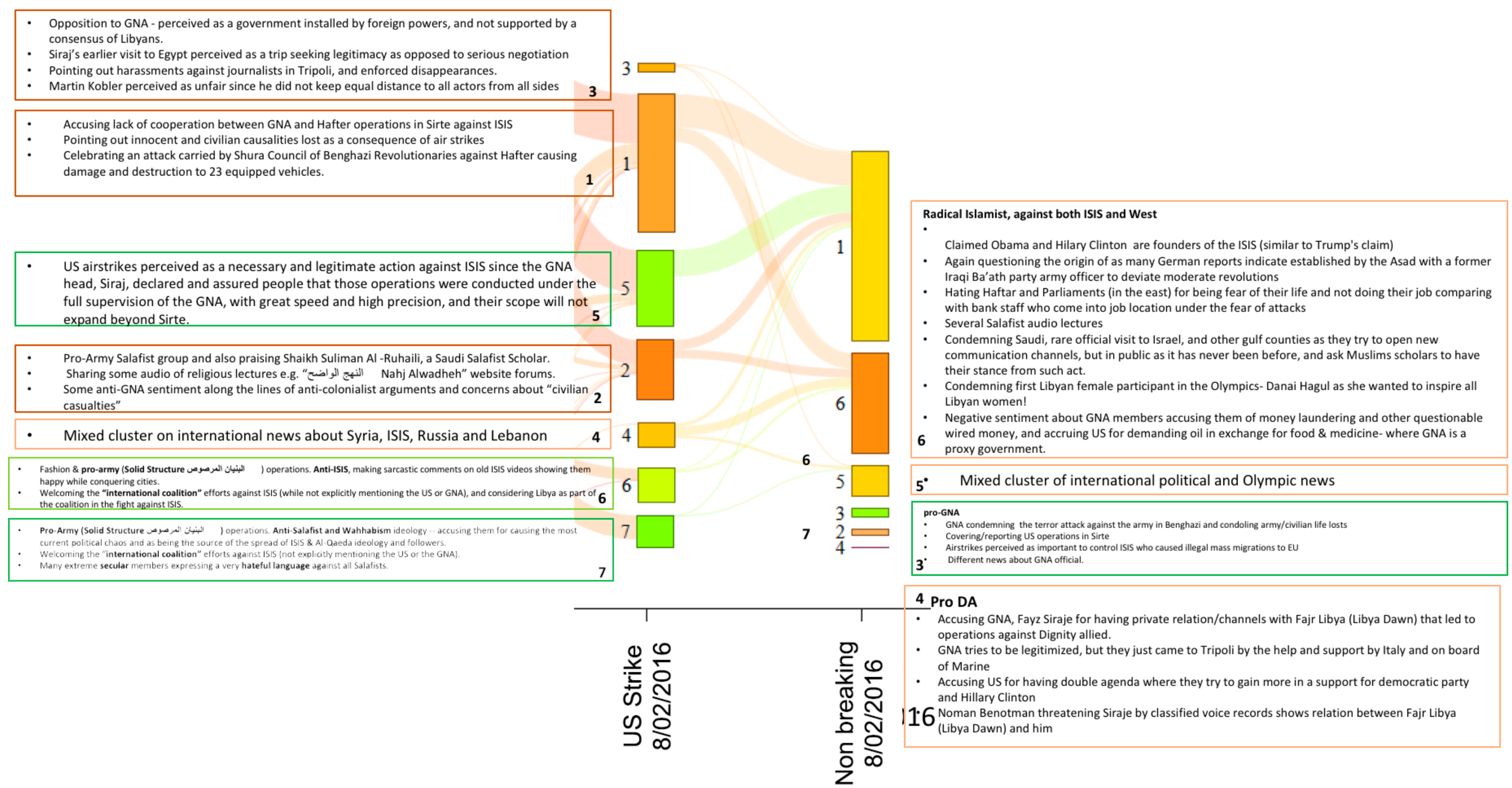


Figure 6.2: A Snippet of Key Summarization Showing a Shifts among Communities Communities

6.4 Feature Space Engineering

Feature Matrix A, where each row a_i represents a tweet, is the main input matrix for our ranker model. This matrix is a conflate of different heterogeneous features, but is related to tweets and it is believed it can identify most potential patterns underlying the data. A can be a concatenation of sub-matrices as follows: $A = [B|\Gamma|\Delta|E]$, where B represents the keywords from tweets (i.e. through text vectorization), Γ represents Page Rank Value of whose posting the tweet, Δ is PageRank "extension" average value (for a heterogeneous graph where vertices include hashtags, users, URL domains; explained more in section 6.4.1). E issues and ideologies are extracted from Chapter 4. Z are reactionary ideologies and issues detected in Chapter 4, and this can help to ideology keyword phrases from being filtered by TF-IDF scores in creating the VSM.

6.4.1 PageRank Extension

PageRank is a well-known algorithm used initially by Google as proposed by [105]. It is a variant of Eigenvector, and it was first applied by Google in an aim to rank webpages and optimize search engine algorithms. The basic difference between Eigenvector and PageRank is that PageRank divides its value of passed centrality by the number of outgoing edges instead of passing the whole centrality (as in Eigenvector). Passing the whole centrality is less desirable since not every high central user will be mentioned or retweeted by another high central user. It requires the graph to be directional, but the depletion factor is added to allow PageRank functions over unidirectional graphs to converge.

Later [140] introduced the notion of extending PageRank to effectively extend heterogeneous vertices for co-author networks as a bipartite graph where vertices are documents and authors. This approach inspired the author of this study to use the approach wherein the vertices are either users, hashtags, and domains. Eacha_i in the A matrix will have its PageRank score vector $\varDelta = (\delta_1, \delta_2, \ldots, \delta_n)$ as a column corresponding for each $Tweet_i$ where its value would be the maximum PageRank score of the set of *user*'s, set of *hashtag*'s, set of *domain*s presented in that $Tweet_i$.

6.5 Indicator Questions

The aim of the indicator questions is (a) to help understand the degree of tweet post relevance, which can effectively help in the summarization process, and (b) to build a context for learning models such that a proposed model can learn a mapping function from instances to ordinal labels. The five questions are sorted from highest to lowest to label a sample of selected tweets like the work by Odijk et al. [54]:

(5) If a posted tweet is expressing an ideology with political offering a prediction.

(4) If a posted tweet is expressing an ideology aligned with a political frame and shows a high degree of insight and explanation of political dynamics.

(3) If a posted tweet is posting updates/developing a piece of news/events but lacks an explanation.

(2) If a posted tweet is touches upon international news but not directly related to domestic matters.

(1) If a posted tweet containing frivolous content, such as sport news, prayers, meaningless social interactions outside of the political context.

6.6 Data Preparation, Training and Validation Process

Twitter is the main source of data for this experiment. Data are collected and the subject matter experts take a sample of tweets [2k tweets] and provide a rank number based on the indicator questions mentioned in the previous section. 2k tweets are then collected and the data is split into training and validation (80% and 20%, respectively). For the training data, the 10 fold cross-validation is employed to infer the class of each instance located in the test batch, and assess the ranker performance.

6.7 Ranking Evaluation

In this part, we aim to examine and sect a model that corresponds to the highest ranking performance judged by Kendall's τ_b which is a method indicating the quality of a ranker based on its resultant ranking labels. For the validation set comes as pair of $A = (a_1, a_2, \ldots a_N)^t$ and $Y = (y_1, y_2, \ldots y_N)^t$ as follows: $(a_1, y_1), (a_1, y_1), \ldots, (a_N, y_N),$. We will firstly apply the regression model $f : a \longmapsto y$. We will apply the model it on A to obtain $\hat{Y} = (\hat{y_1}, \hat{y_2}, \ldots \hat{y_N})^t$. Both Y, ground truth ranks and $\hat{Y}$ ranks obtained by the learner model will be used in computing the Kandell's τ_b.

6.7.1 Kendall's τ_b

Kendall rank coefficient ([26]) are mainly used to assess the ordinal associaltion between two quantities, namely Y and $\hat{Y}$. The τ_b coefficient. $\tau \in [-1, +1]$ where -1 indeicate the extereme discordance (e.g. if $\hat{Y}$ is an inverse of Y), and +1 indicates the extereme accordance (e.g. $\hat{Y}$ is an identical of Y), and any value near to 0 shows no association between Y and $\hat{Y}$. Kendall's τ_b can be computed as follows:

$$\tau_b = \frac{P - Q}{\sqrt{(P + Q + Y_0) \times ((P + Q + \hat{Y}_0)}} \tag{6.2}$$

The nomenclature is explained as follows:

- P: number of concordant pairs such that $[(y_i < y_j)$ **and** $(\hat{y}_i < \hat{y}_j)]$ **or** $[(y_i > y_j)$ **and** $(\hat{y}_i > \hat{y}_j)], \forall i, j \in [1, n], i \neq j$

- Q: number of disconcordant pairs which is the number of observed ranks below a particular rank; $[(y_i < y_j)$ **and** $(\hat{y}_i > \hat{y}_j)]$ **or** $[(y_i > y_j)$ **and** $(\hat{y}_i < \hat{y}_j)]$ $\forall i, j \in [1, n], i \neq j$, $\forall i, j \in [1, n], i \neq j$.

- Y_0: if observation tied with Y; $[(y_i == y_j)$ **and** $(\hat{y}_i > \hat{y}_j)]$ **or** $[(y_i == y_j)$ **and** $(\hat{y}_i < \hat{y}_j)]$, $\forall i, j \in [1, n], i \neq j$

- $\hat{Y}_o$: if observation tied with $\hat{Y}$; Y; $[(y_i > y_j)$ **and** $(\hat{y}_i == \hat{y}_j)]$ **or** $[(y_i < y_j)$ **and** $(\hat{y}_i == \hat{y}_j)]$, $\forall i, j \in [1, n], i \neq j$

- $(Y\hat{Y})_o$: ; **or** $[(y_i == y_j)$ **and** $(\hat{y}_i == \hat{y}_j)]$, $\forall i, j \in [1, n], i \neq j$

- N: number of possible pair. $N = \binom{n}{2}$ which is also can be captured by the sum of all quanities $N = P + Q + Y_0 + \hat{Y}_0 + (Y\hat{Y})_0$ since it all sets are mutually disjointed (exclusive) and they cover all possible pairs.

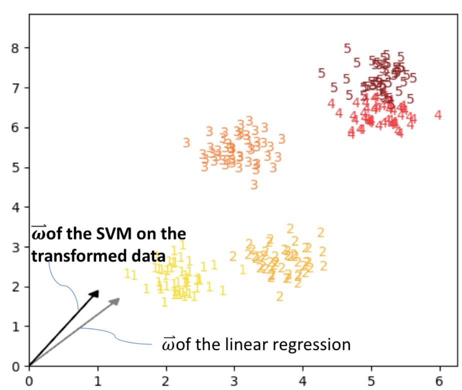
6.8 Baseline Model

This model was aimed to be simple and to serve as a baseline. General logistic regression was employed, treating each class as a nominal class (i.e. categorical where the order is not preserved). The training objectives is to minimize the error since the strategy used is one-verse-all (OVA), hence each instance acquires its learned label by means of a classifier of each label which provides the highest confidence score.

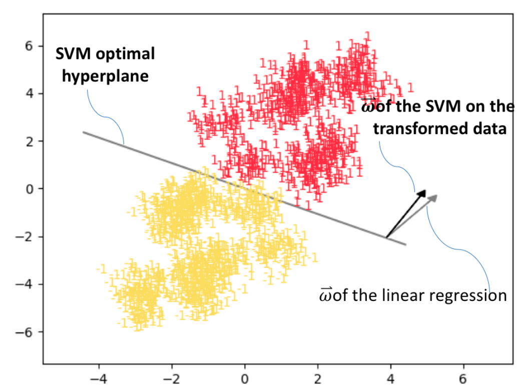
6.9 Ordinal Classifier

The previous model overlooked the ordinal nature of the label, wherein all classification mismatches are treated equally (e.g. true value is ‘1’ but the inferred value is ‘5’ and would be panelized equally if 2 is inferred), which is not quite precise in building a ranker model. Thus, the ordinal classifier can be as simple as an SVM classifier with the aim to infer the direction of predicted real value which corresponds to the ordinal value. Through bisecting prediction value into ordinal ranges of the prediction value, then through appropriate mapping, ordinal value can be computed which aims to mimic the original ordinal labels. This model has been proposed by Herbrich et al. [66] who introduced a learning task which requires transforming the input instances and their ordinal labels to different space of instances and labels that allow to simplify this problem and making it as a linear classification that can an SVM classifier be utilized with an optimal hyperplane and a vector w (orthogonal to the SVM hyperplane) by which the w is used to learn the right ordering direction as shown in the following figure of the 2-D numeric example.

$$\forall_{i,j \in N}(x_k, y_k) \leftarrow (x_i - x_j, sign(y_i - y_j)) \tag{6.3}$$



(a) Original Data Sample of Ranked Label Plotted on 2D



(b) Two Class Data Transformation and SVM is Utilized to Derive $\hat{w}$

Figure 6.3: 2D Example of Ordinal Data

Both vectors correspond to the linear regression, and the SVM are shown in Figure 6.5 for the sake of comparison. To assess the quality of ordering, we can use w of each and project samples then examine by which approach can yield to better model. The evaluation can be done by feeding both projected data instances (which came as real values) with ordinal labels. After computing Kendall's τ_b, 0.84 and 0.88 was found for the linear regression and SVM, respectively, which indicates a more accurate SVM Rank approach in identifying the right victor direction used to assign rank to each instance.

6.10 Gradient Boosting Decision Tree

The Gradient Boosting Decision Tree is a pointwise ranking approach vs. a pairwise classification approach as in the Ordinal Classifier that optimizes a regression-based method in an ensemble learning fashion decision tree [38]. It uses a relatively combined number of shallow trees like Random Forrest, but the addition of every tree is an attempt to correct previous errors of the previous stage errors by old trees. Errors are reduced as trees are added to the model.

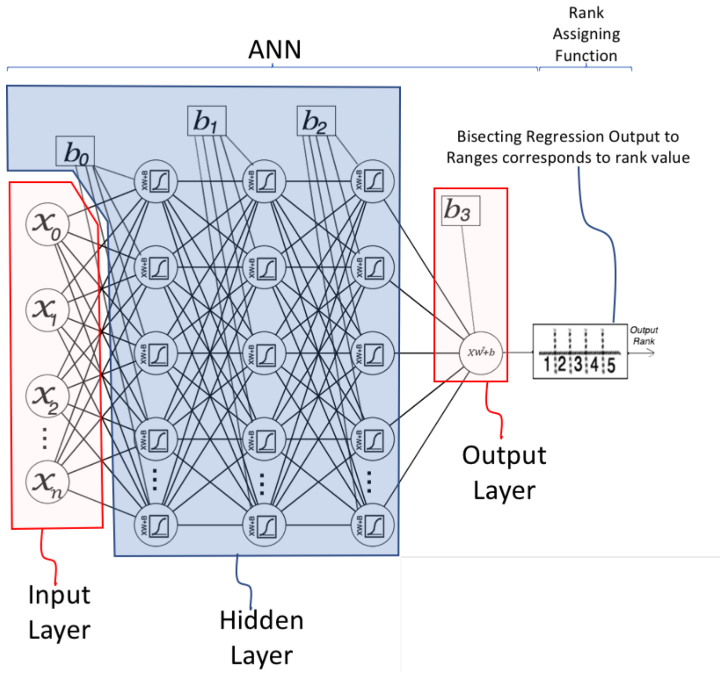
6.11 Artificial Neural Network Model

Artificial Neural Network Model (ANN Model) is similar to the previous model, yet it extends existing models to potentially suit more complex topology of given data and support the pointwise training approach. The ANN Model can reveal if the models would either suffice the data topology or there if a shortcoming is present which requires an ANN Model to improve the ranker significantly. An ANN model is considered as a graph of nodes and edges where nodes represent mathematical operations (i.e. multiplication with the of node weight then summed up with the bias followed by passing the results to the sigmoid kernel), while the edges represent the data traversing the network (either as original input data or an output of a proceeding layers). An optimization choice was made to suit this learning task and it was based on reducing error of the linear regression of final output when compared with the truth label ordinal values. The minimum square error ($MSE = \frac{1}{n}\sum_1^N (y - \hat{y})^2$) is used as cost function to the last layer, where y and $\hat{y}$ are true label and predicted, respectively. Therefore, the neural network will be trained and evaluated by the Kendalls τ_b and is compared to two other previous models.

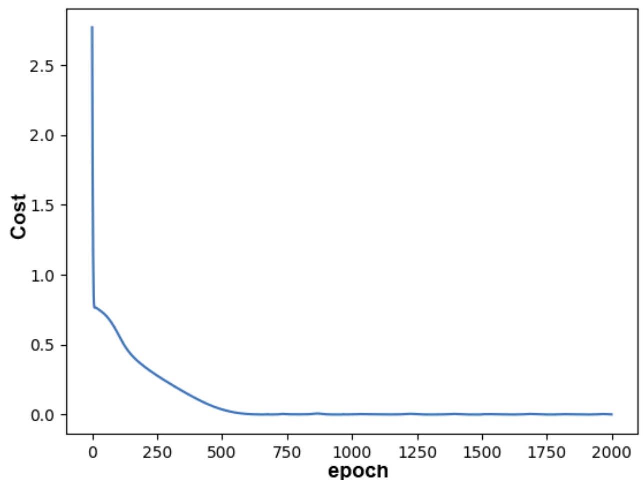
6.12 Experiment

After curating 5000 tweets and labelling them using the indicator questions, the labeled tweets were used with three learning models to examine their performance and show a robust model to help rank all tweets in all snapshots. The three models were benchmarked and set the Kendalls τ as the measure of which a method would result in the highest score. Additionally, for the sake of comparison, simple accuracy and correlation were included. The ranked dataset was split by the indicator question into two sets: (i) training and presenting 80% of the data and testing 20% (ii) training data used for training model and same data used for testing, and performance of each model are reported (i.e., accuracy, correlation, Kendalls τ). On the other hand, for the Testing set, all Training set to train models was used, but models are applied on the Testing set to report the performance. One observation made related to the performance of Training set and how it is always higher than the Testing due to unseen instances in the Testing dataset. Interestingly, when it comes to computing performance, accuracies do not reflect the true value of a model to rank, yet it can be noticed that the baseline model, which is an SVM, scored higher than SVM Rank due to the nature of optimization defined. However, the correlation is higher in SVM Rank indicating the most accurate optimization approach as the objective to primarily rank content (i.e. a higher correlation). However, ranking poses a subtle requirement which penalize inversion for instance when two rankers rank two elements says a, b, but each ranker oppositely give one element the opposite rank, the correlation may not capture this flaw especially when two values are relatively high or low, yet Kendalls τ takes this into account by being more to penalize more than a simple naive correlation approach. Among the three models, the ANN model outperforms the others with a Kendalls τ score of 0.76, which is high and acceptable to be used to rank all the tweets for each snapshot. Figure 6.4 shows how to apply rank for each tweet such that subject matter experts can retrieve contents corresponding to the highest-ranking scores to support their retentions, while exploring snapshot contents. The evaluation per snapshot has been completed and more significant patterns have emerged explaining the merging and the splitting regarding the looming issues associated to the new development and changes in the local and international arena. More tweets have been spotted showing a high level of (1) interpreting of current political complexities and riots on the ground (2) extrapolating of what future holds (3) Understanding each factions play cards and their limits of influence and funds on each snapshot within the political turmoil that affecting all Libyans at all levels. By all these contents Subject Matter Experts can reach a very careful read to snapshot with a deep understanding within a very short time.

Figure 6.4: Visualization Tool Charting Tweets after Content Ranking by Showing the Social Dimensions That Aforementioned in the Previous Chapter- the Tool Supports User Interaction Allowing Further Content Engagement for Further Understanding per Political Group, per Dimension, and per Cluster of a Snapshot



(a) ANN Model



(b) Cost Declination as Optimization Process Iterate over Epoch Size

Figure 6.5: ANN Structure and the Training Cost vs. Iteration Plot

	Model1:Baseline Nominal Classifier		Model2: SVM Rank		Model3: GBRank		Model4: ANN	
	Training	**Testing**	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
Accuracy	0.732	0.655	0.637	0.548	0.768	0.66	**0.894**	**0.769**
Correlation	0706	0.613	0.758	0.644	0.851	0.714	**0.951**	**0.794**
Kendall's τ	0.671	0.572	0.7	0.584	0.800	0.668	**0.924**	**0.765**

Table 6.1: Four Models Comparison in Terms of Accuracy, Correlation, and Kindall's τ, the Ann Model Outperforms All the Other Models, and Another Observation: The Baseline Model Is Higher in Accuracy If Compared with the RankSVM, but the Latter Relatively Scores Better in the Kindell's τ Score Which Is the Right Proper Ranking Measure Evaluation

6.13 Conclusion

Making meaning out of dynamic clusters based on relevance of the subject matter experts is essential in an environment fostering real-time and demanding high level of agility while SMEs try to summarize cluster content. This study proposes a ranker approach augmenting the co-clustering on each snapshot of Twitter feeds that take into consideration patterns presented in data as well as SME preference Making meaning out of dynamic clusters based on relevance of the subject matter experts is essential in an environment fostering real-time and demanding high level of agility while SMEs try to summarize cluster content. This study proposes a ranker approach augmenting the co-clustering on each snapshot of Twitter feeds that take into consideration patterns presented in data as well as SME preference

through leveraging the machine learning model. This notion can effectively bridge the gap between human limitation of not being exposed to all necessary information by alternatively ranking most relevant information from experts perspective that consequently would benefit them to reduce the volume of information needed to sift through and boost retention when analyzing contents. This study showed snapshots clusters and ranked their content with a Kendall's τ score of 0.765. For future work, I plans to supplement the framework with entity recognitions such that the score can reflect the importance of leaders' names and important places that are highly related to the conflict, aiding in more searching capability and more adaptive scoring techniques for the ANN model.

Chapter 7

FRAMING SHIFTS OF THE UKRAINE CONFLICT IN PRO-RUSSIAN NEWS MEDIA

7.1 Introduction

Analysts recognize that the Russian government uses information operations (IO) as a tactic in its strategic efforts to reclaim territory in former Soviet states (it's so-called "near-abroad"[114]) [1] . For example, in 2008 Russia sent troops into South Ossetia, Georgia in response to an attack on the semi-autonomous region by Georgian forces. The speed and decisiveness of the Russian invasion and their subsequent extension of the invasion into Georgia proper caught Western leaders by surprise. Russia had promoted ethnic conflict in Georgia to maintain influence there, [2] and provided extensive support to South Ossetian and Abkhazian separatists[72]. Russia also exchanged old Soviet passports for new Russian ones in both South Ossetia and Abkhazia [3] so-called "passportization"- creating a pretext for intervention to protect "Russian citizens," and to take de facto control. Less than six years later, the West was again surprised when Russia used the same techniques to support annexation of Crimea in Ukraine. Joint Chiefs Chairman General Martin Dempsey said of Vladimir Putin, "he's got a playbook that has worked for him now two or three times."[73]

[1] This work has been submitted and accepted [18]

[2] Archives of the CSCE, Georgia Files, Com. No. 408, Prague, Stockholm, 11 December 1992; Ibid, N.41, Prague, 2 February 1993; Bruce Clark, 'Russian Army blamed for Inflaming Georgian War,' The Times, 6 October 1992; Fiona Hill and Pamela Jewett, 'Back in the USSR: Russia's Intervention in the Internal Aairs of the Former Soviet Republics and the Implications for United States Policy toward Russia,' Cambridge, MA.: Harvard University JFK School of Government, Strengthening Democratic Institutions Project, January 1994.

What is in this playbook?

> Case officers for the intelligence community operate without official cover, [and] recruit sources and assess the battlefield. Then, small units of special operations forces sneak in, sometimes blending in with the populace, ready to make trouble. Then, special forces units that specialize in "information operations" designed to induce anxiety and outrage among local populations follow a strategy that comes from the top of the government. The idea is to generate genuine indigenous protest movements. Using these protest movements as evidence of "human rights violations," Russia intervenes [71].

It is widely believed that Russia aims to repeat this performance in other ethnically Russian areas, especially the Gaugazia region of Moldova[91]. The Baltics are also a potential target. Three years ago, a Russian Foreign Ministry official echoed playbook tactics when he warned that ethnic discrimination there "may have far-reaching, unfortunate consequences."[92]

If there is a playbook in operation, then in principle it should be possible to detect its IO signature, stimulated by Russian propaganda and other 'gray zone' activities, in mainstream media, to potentially provide early warning of another invasion in other near-abroad states. This chapter describes results of a proof-of-concept effort by the ASU's Center for Strategic Communication and Lockheed Martin Advanced Technology Laboratory. Our goal was to detect shifts in framing surrounding the 2014 annexation of Crimea using natural language processing of Russian propaganda articles and machine classifiers trained to recognize framing.

The corpus in the study was comprised of over 100,000 news articles from 372 news sources dated between 2010 and 2017. Our methods and contributions can be summarized as follows:

- We recruited a pair of area experts to classify top 200 news sources as either pro-Russian or other. We were able to train a classifier which achieved 90% F1-score to discriminate between propaganda vs. other articles.

- We worked with subject matter experts (SMEs) from the ASU Center for Strategic Communication (CSC) collaborated on this study to inductively develop a code book comprising five categories of Russian strategic frames used in the Ukraine. Four student coders were trained to map sentences in randomly selected articles to one (or none) of these framing categories. After multiple rounds of training, coders achieved a inter-coder reliability (a.k.a Krippendorff ratio) of $\alpha = 0.83$ [83], which was judged as acceptable.

- We used coded sentences to train a text classifier which achieved 77% F1-score in labeling unseen sentences with the correct frame (or "no frame").

- The propaganda and framing classifiers were used on the news corpus to produce a daily time series of framing density vectors for articles classified as Russian propaganda. We computed Jensen-Shannon [5] divergence between framing density vectors of consecutive days. Results show significant framing shifts exceeding a smaller peak of 2010, in November 2013, and sharply spiking and trending again in Dec 2013, three-four months ahead of Crimea's annexation by the Russian Federation – which took place between 20 February 2014 and 19 March 2014. The war has been ongoing in the Donbass region of Ukraine since 6 April 2014 until the present day.

The rest of the chapter is organized as follows. Section 6.2 presents a review of related works. Section 3 summarizes our data sources and approach. Sections 4 and 5 present the codebook of Russian strategic framing induced from propaganda articles and our sentence coding procedure. Sections 6 and 7 present text classifiers for frame detection, time series analysis of daily framing density vectors and significant framing shifts. Section 8 concludes the presentation with discussions and future work.

7.2 Related Work

Framing analysis has roots in mass media studies and several frameworks for assisting human identification and coding of frames were developed. Notable works include: Odijk et al. [54] where they developed a two-phase approach: (1) a systematic questionnaire for human coders to evaluate the nature (i.e. conflict, economic consequence, human interest, morality) and aspects of framing, (2) an ensemble of classifiers trained to detect frame presence in text using the coders questionnaire responses. Baumer et al. [25]compared performance effects of different types of features (i.e. lexical, grammatical and manual dictionary-based) for detecting frames in news. Their findings suggest that lexical n-gram features combined with grammatical part-of-speech (POS) tags result in significant improvements in frame detection. We also employed lexical frequent discriminative bi-grams alongside grammatical (subject, verb, object) based generalized triples [37] as features in our framework. Our experiments resulted in an accuracy of 41% average F1-score with bi-grams alone, and an average F1-score of 77% with combined features including bi-grams, generalized triples and other lexical features.

The temporal analyses of framing are also relevant since they can offer indications for detecting framing shifts. Several works were developed for spike detection in noisy time series data based on raw signal smoothing [126] and wavelet trans-forms [98] for different types of data (e.g. seismic analysis, disease epidemiology, and stock market prediction, etc.). Wend et al. [136] proposed an event detection framework in messages based on detecting correlated bursts of keywords that are expressed during events. To identify related keywords, they apply wavelet transformations on time series of keyword frequencies and measure cross-correlations between keywords and events. Next, they employ modularity-based graph clustering to detect keyword groups signaling events. In this research, we utilized Jensen-Shannon divergence [5] to measure the daily variations of framing densities in pro-Russian international news. We checked the overlaps of their framing shifts and trends over time with significant phases of the Ukraine crisis to draw our conclusions.

7.3 Approach

Our analysis is based on detecting strategic framing [49, 118] in news articles. Framing is accomplished when a choice of words, phrases, metaphors, images, and other rhetorical devices prefer a one side of possible interpretation of a set of facts, but intentionally discourage other interpretations- in purpose or unconsciously. A special case is adversarial framing, which "is typically competitive, fought between parties or ideological factions, and [where issues] are debated and framed in opposing terms."[40] A domestic example of adversarial framing is Republicans in the 1990s referring to the US estate tax as a "death tax"- connoting the long arm of the government taxing you even beyond the grave - while their political opponent Democrats referred to the same tax policy conventionally, as an "estate tax" - suggesting that only the super wealthy are subject to the tax.

Similar techniques are used by Russia with respect to the near abroad countries it threatens. One signature behavior is the framing of an ethnic issue as dealing with "human rights." In May 2014, the Russian Foreign Ministry released a white book detailing what it said were large-scale human rights violations in Ukraine [1], including discrimination against religious and ethnic minorities. In an earlier speech to the Russian Parliament, Vladimir Putin complained, "we hoped that Russian citizens and Russian speakers in Ukraine, especially its southeast and Crimea, would live in a friendly, democratic and civilized state that would protect their rights in line with the norms of international law. However, this is not how the situation developed." [2]

Framing is also undertaken by ethnic groups in the countries where Russian incursions are a threat. In 2012, a Latvian referendum rejected Russian as an official national language. Residents of Eastern regions where Russian is the primary language framed this act as a violation of rights. One such resident was quoted as saying: "[Latvian] society is divided into two classes - one half has full rights and the other half's rights are violated."[6]

Our approach, therefore, sought to identify and detect strategic framing before and after the 2014 invasion of Crimea. To do so we (i) collected mainstream media texts from Russian propaganda sources dealing with Ukrainian ethnic and political issues for the period between 2010 - 2017, (ii) inductively developed a set of framing categories, (iii) trained human coders to reliably identify sentences invoking these frames in sample texts, (iv) used these coded sentences to train machine classifiers to recognize all other framing instances in the corpus, (v) generated vectors representing the daily densities of these frames in news articles classified as propaganda, and (vi) conducted time-series analysis to identify shifts in framing densities and (vii) locate these shifts within significant phases of the Ukraine conflict.

7.3.1 News Corpus

This project was supported by Lockheed Martin Advanced Technology Laboratories and used news feeds extracted from Lockheed Martin's ICEWS system. ICEWS is a program of record in the U.S. Department of Defense used by component agencies to track conflict events. During its operation, ICEWS collects and archives English-language and translations of foreign language articles from mainstream media sources and websites worldwide. We queried the ICEWS database for articles between 2010 and 2017, which mentioned Ukraine, and further constrained this dataset to stories which contained keywords believed to be associated with Russian propaganda (i.e. anti-facist, discrimination, second-class citizens, etc.). This resulted in a news corpus containing 103,912 articles.

To focus our analysis on Russian propaganda sources, we recruited two area experts to classify the top 200 sources in our corpus (in terms of article frequency) as either pro-Russian or other. Next, we extracted bigrams and generalized concepts [37] from these sources and we trained a sparse logistic regression text classifier to discriminate between propaganda vs. other type of articles. A ten-fold cross-validation evaluation showed that the propaganda detection classifier has an average F1-score of 90% and an F1-score of 86% for the smaller Russian 'propaganda' category. We ran this classifier on the news corpus, yielding 30,845 texts classified as Russian propaganda. These texts formed the basis of our coding and framing analysis. Figure 7.1 shows the volumes of news as we can see three main phases, namely: before Russian military intervention, within and after. As this image may capture a very broad information, but this may lead to many questions as: (1) What frame category volumes is in the Russian propaganda, presented in documents and much deeper as sentence level. (2) Framing shifts and if exists? and what indications can be found in data that aligned in with action of military actions or invasion? Coming section will address these questions.

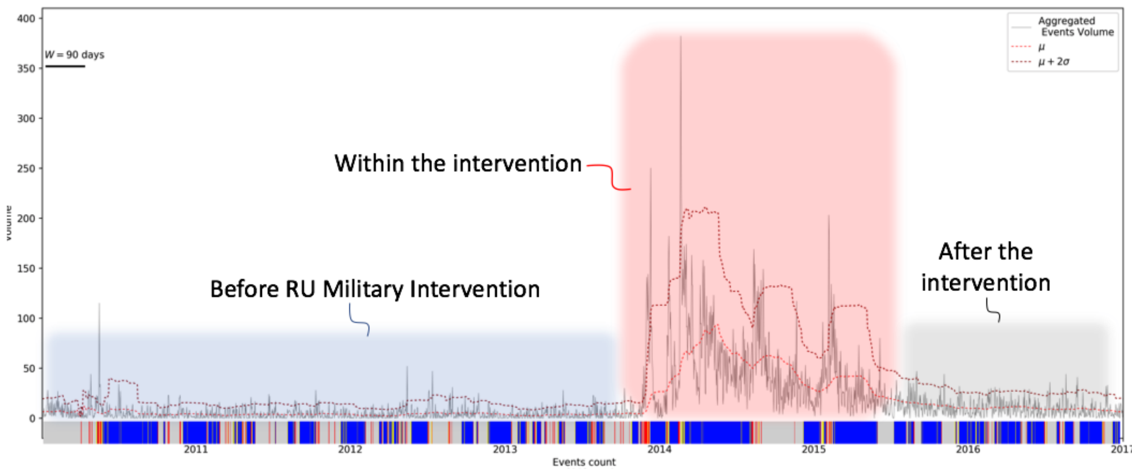


Figure 7.1: News Article Volume, Running Average, μ and $\mu + 2\sigma$ Are Plotted to Mark Break Out "Trending" Volumes. At the Bottom Band: Gray Indicating Idle State, Red: Breaking, Yellow: Accumulating, and Blue: Die Out 'Cascade'

7.4 Codebook

Using with the notion of the playbook described in the introduction, we randomly selected articles from our Russian propaganda sources with high counts of discriminative propaganda-related keywords. Two subject matter experts from ASU's Center for Strategic Communication (CSC) read these texts and identified the following five framing categories inductively:

Fascist vs. anti-fascist struggle (denoted by: fascist) . There are frequent accusations that leadership/society of a target country support "fascists" or "Nazis," and take actions to harass "anti-fascists" or hinder their efforts to protest and take other actions against the fascists. Essentially, the Nazis/fascists are the "bad guys" from the Russian point of view, and the anti-fascists are the "good guys." Almost any use of "Nazi," "fascist," or "anti-fascist" qualifies as framing, because it interprets the people involved and their actions as part of an ideological struggle between the two sides.

Discrimination against Russian minorities: (denoted by: discrim) This frame addresses discrimination against groups, usually ethnic groups; any such group having its rights trampled on, being marginalized or abused or similar affronts constitutes this frame. Russian information operations seek to convince members of the Russian speaking community in target countries that they are being victimized, discriminated against, and their rights are being violated. This might include references to general or human rights, or specific references to rights like voting, freedom of speech, and political participation. They also claim that there are efforts to stamp-out use of the Russian language, to suppress Russian culture, and to discriminate against Russian speakers in the job market and other domains. Lack of citizenship or denial of citizenship is a form of discrimination.

Assault on Soviet history (denoted by: history) Russian information operations seek to condemn the subversion or suppression of Soviet history. This can take the form of complaining about the removal of statues and memorials commemorating the Soviet role in World War II, changing names of Soviet-era streets and other geographical landmarks, or trying to change the historical narrative about the Union of Soviet Socialist Republics (USSR) and its role in former Soviet states.

Criticism of government (denoted by: gvmnt) Russian information operations seek to criticize the governments of target countries, in terms of functioning, procedures, and results (including economic results), as well as corruption among government officials. The frame implies that government is ineffective, not functioning properly, and acting in ways that are detrimental to good governance. The "government" includes legislative, executive and judicial branches at the national, provincial and municipal levels; it includes the police; it includes semi-synonymous terms like "the authorities". The frame applies when the national, provincial or municipal government of a target country is criticized (such as Ukraine, Latvia, Georgia, Lithuania, Estonia, Moldova, Poland, etc.)

Invasion of Crimea (denoted by: crimea) Russian information operations seek to justify and create support for their annexation of Crimea. This can involve discussions of sovereignty, discussion of the area's future, and statements supporting the annexation. The annexation is often framed as a moral imperative or a righteous act, and subsequent opposition by Ukraine, EU, and the international community are immoral, hypocritical, etc. Select this frame when the annexation of Crimea is clearly the context of some sort of justification, not when it could be the subject of the justification.

7.5 Finding related News Media Outlet related to Russian Propaganda

In this section, we need to explain our approach in determining and categorizing set of news outlet into two subsets:namely Russian propaganda vs. European propaganda. We have given set of news sources, and this task is crucial as other pipeline processes are very dependent on it. Some news outlet has very distinct affiliation to Russians propaganda, others might be such clear. To examine news sources, this entails subject matter experts to examine that meet both precision and agility to surface the most representative frames from right set of news outlets.

As an exploration step, we have an assumption that all news outlet discuss news items with latest events and related detailed articles, but at same time these outlets add their perspectives which likely attached with political agendas. To utilize this relation as a graph model to allow us separate Russian propaganda outlets from others (i.e. biased to European propaganda or neutral towards Russian conflicts with its neighbors)

relation one of approaches that can give some preliminary results is the bipartite network projection which serve both compression information and offer a quick summary.

Final labeled news source are reached and Table 7.1 shows some news source per category where the number of Russain Propaganda new Outlet is Russian propaganda P: 67, European Propaganda NP: 206, unkown: 1301

P news sources	NP news sources	Unknown news sources
Interfax	The British Broadcasting Corporation	OSC Summary
ITAR-TASS Information Telegraph Agency of Russia	Agence France-Presse	Channel One TV
Rossiya Segodnya International Information Agency	Thomson Reuters (Markets) LLC	Den Online
RIA Novosti	BBC Monitoring	USA Today Information Network
Interfax Information Services, B.V.	Dow Jones & Company, Inc.	Human Rights in Ukraine
Ukrainian National News Agency (Ukrinform)	AFP (North European Service)	MK Online

Table 7.1: Sample of News Outlets after Categorizing Them into the the Three Types

First we run NMF coclustering to indentify set of docuemts and set of bigrams that highly group with each other.

7.6 Frame Coding

Computer-aided techniques of frame coding essentially use two approaches: (I) dictionary/keyword lists based (e.g. [27]) or supervised learning approaches (e.g. [109]) trained with human coded sentences. In this project four student coders were trained to assign sentences in randomly selected propaganda texts to one (or none) of the five framing categories described above. Coders would first work independently, assigning each sentence to one (or none) of the coding categories. We would then

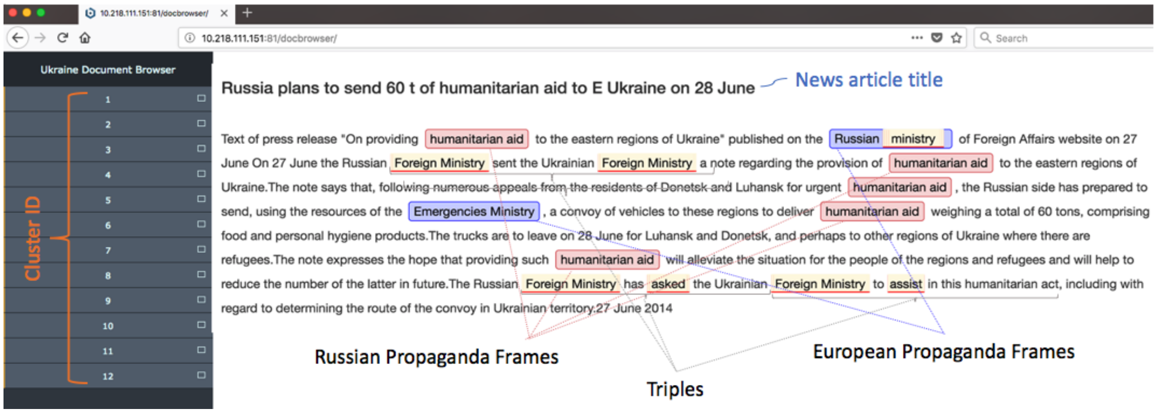
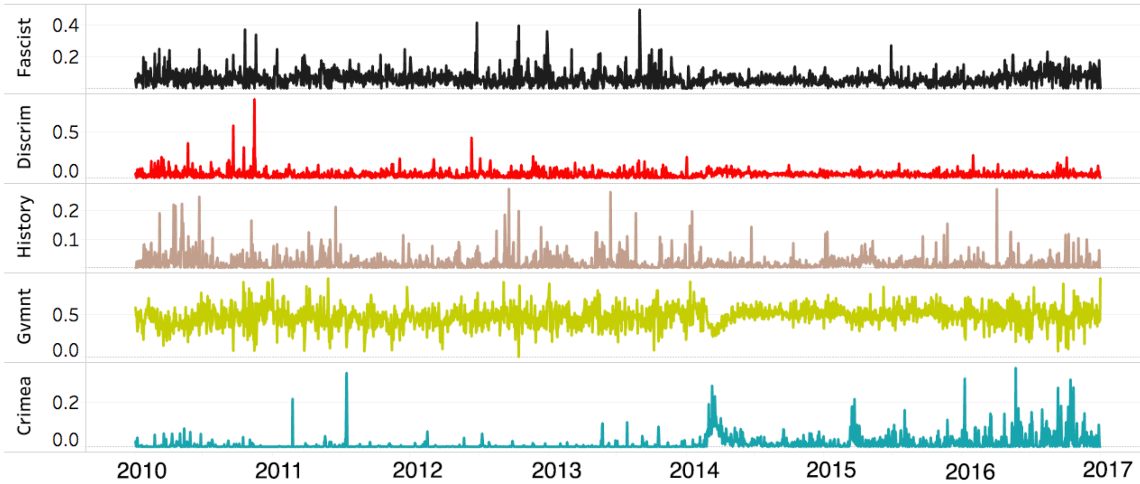


Figure 7.2: A Document Fetched from Database Showing How Frames and Generalized Triples Are Automatically Highlighted and Rendered. Reddish Color Indicating Pro-Russian and the Blueish Is Other

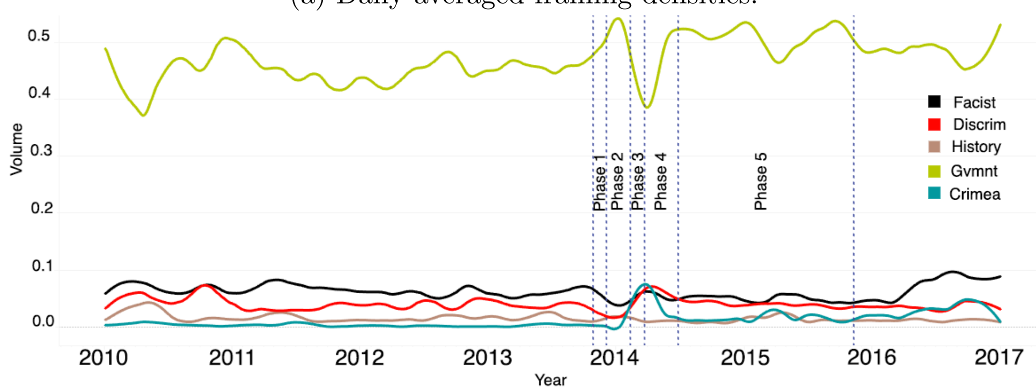
calculate reliability, and identify disagreements between coders. Coders would then discuss these disagreements as a group, and we would refine category definitions in the codebook as necessary. After seven rounds of training, coders achieved a inter-coder reliability (a.k.a Krippendorff ratio) of $\alpha = 0.83$ [83], which we judged acceptable. Subsequent coding was performed by two randomly assigned coders per text, who discussed and resolved disagreements to arrive at a final set of codes. They coded texts until we had a large enough set of coded sentences, where adding more coded sentences no longer significantly boosted the overall accuracy of the best text classifier model. The final number of coded sentences in each category was: *crimea*, 162; *discrim*, 196; *fascist*, 307; *gvmnt*, 334; *history*, 187, and those sentences were used as the labeled training dataset.

7.7 Frame Detection Model

We used coded sentences described above alongside a random collection of sentences that were not mapped to any framing category from coded articles to train five classifiers - one classifier for each frame category. We used one-vs.-all (OvA) strategy



(a) Daily averaged framing densities.



(b) Smoothed Daily Averaged Framing Densities.

Figure 7.3: Daily Averaged Framing and Smoothed Densities.

which involves training a single classifier per frame, with the samples of that frame as positive samples and all other samples as negatives. We extracted four sets of features from each sentence: keywords, frequent bigrams, whether the sentence contained a quote, and its matching generalized semantic triplets. Generalized semantic triplets (GST) are merged collections of subjects, verbs, and objects that co-occur together in similar contexts. The details of the GST features can be found in an earlier paper [16, 12]. We evaluated several text classifiers using ten-fold cross-validation. The best overall performance was obtained with a linear SVC (L1) classifier yielding the

Classifier	Frame														
	fascist			discrim			history			gvmnt			crimea		
	Pr	Re	F1	Pr	Re	F1	Pr	Re	F1	Pr	Re	F1	Pr	Re	F1
Ridge Classifier	.82	.68	.74	.78	.67	.72	.83	.62	.71	.73	.63	.68	.87	.8	.83
Perceptron	.78	.65	.71	.71	.77	.74	.77	.73	**.75**	.76	.62	.68	.84	.86	.85
Passive-Aggressive	.8	.65	.72	.79	.69	.74	.81	.67	.73	.75	.71	**.73**	.89	.8	.84
KNN	.6	.21	.31	.83	.08	.14	1	.01	.01	.47	.41	.44	.87	.08	.15
Random forest	.84	.55	.67	.8	.53	.63	.83	.52	.64	.87	.36	.51	.84	.81	.82
LinearSVC(L2)	.79	.68	.73	.76	.68	.72	.83	.67	.74	.74	.69	.72	.89	.78	.83
SGDClassifier(L2)	.8	.69	.74	.71	.69	.7	.79	.67	.73	.72	.64	.68	.88	.86	**.87**
LinearSVC(L1)	.79	.71	**.75**	.81	.71	**.76**	.8	.7	**.75**	.72	.74	**.73**	.85	.79	.82
SGDClassifier(L1)	.75	.65	.7	.73	.67	.7	.78	.72	**.75**	.7	.65	.68	.85	.82	.84
SGDClassifier(Elastic-Net)	.73	.65	.69	.75	.58	.66	.79	.71	.74	.78	.63	.7	.84	.83	.84
NearestCentroid	.45	.78	.57	.45	.7	.55	.64	.65	.64	.44	.87	.58	.79	.78	.79
MultinomialNB	.54	.66	.59	.47	.77	.59	.56	.91	.7	.69	.73	.71	.54	.92	.68
BernoulliNB	.5	.78	.61	.52	.83	.64	.61	.9	.73	.6	.82	.7	.65	.93	.76
LinearSVC(L1)	.77	.71	.74	.78	.68	.73	.82	.7	**.75**	.72	.72	.72	.87	.81	.84
GradientBoostingClassifier	.82	.65	.73	.8	.56	.66	.84	.6	.7	.81	.5	.62	.84	.83	.83

Table 7.2: Frame Detection Accuracies

following F1-scores: history, 74%; crimea, 87%; discrim, 76%; fascist, 75%; gvmnt, 73%; average, 77%. The rest of the results are shown in Table 7.2.

7.8 Time Series Analysis of Daily Framing Densities

The set of frame classifiers were applied to each sentence to produce real-valued confidence scores. The classifier which reported the highest confidence score was considered to be the dominant frame category for each sentence. We applied this technique to all sentences in each article one-by-one in order to produce a vector of framing density values for each article. These vectors were averaged daily to yield a vector of daily averaged frame densities shown in Figure 7.3. Since the time series were noisy, first we performed Gaussian smoothing, shown in Equation 7.1 and

Equation 7.2 (where σ, w are $2, 10$ respectively, acting as low-pass filter) to remove high frequency noise. The smoothed time series are shown in Figure 7.3. Next, in order to reveal framing shifts, we computed Jensen-Shannon [5] divergence, a statistical distance measure, between the daily framing density vectors of consecutive days. The resulting divergence plot is shown in Figure 7.4.

$$N(x; \mu = 0, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{-2\sigma^2}} \tag{7.1}$$

$$S(t) = \sum_{i=t-w/2}^{t+w/2} O(i)N(t-i) \tag{7.2}$$

Knowing that KL is the Kullback-Leibler divergence $KL(p; q) = p_i \ln \frac{p_i}{q_i}$, Jensen-Shannon divergence can be expressed in term of KL as follows

$$JS(v_1, v_2) = KL(v_1, \frac{v_1 + v_2}{2}) + KL(v_2, \frac{v_1 + v_2}{2}) \tag{7.3}$$

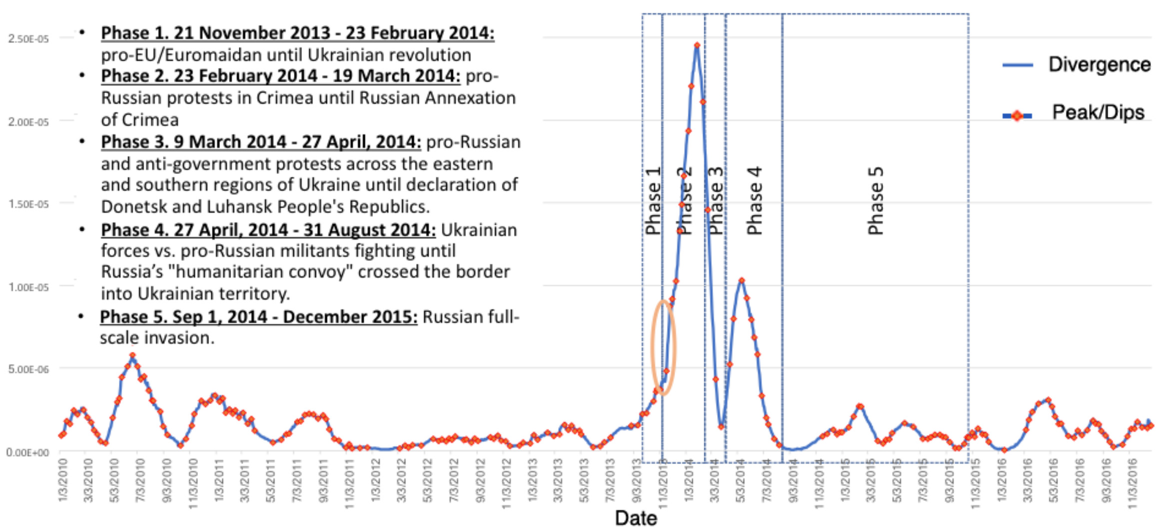


Figure 7.4: Daily Jensen-shannon Divergence-vertical Lines Demarcating the Significant Phases of the Ukraine Conflict Timeline Determined by the CSIS (CENTER FOR STRSATEGIC & INTERNATIONAL SUTDIES `http://ukraine.csis.org/`)

Prior to Phase 1, corresponding to the period between pro-EU Euromaidan protests until the Ukrainian revolution, divergence remains at relatively low levels, except for some small peaks during 2010 - 2011. As the pro-EU/Euromaidan protests begun in November 2013, the divergence signal begins to rise, exceeding all previous highs in November 2013, followed by a sharp rise in Dec 2013. Divergence increases sharply during the pro-Russian protests well into the midst of Phase 2 which terminates with the annexation of Crimea by the Russian Federation on March 19, 2014. Following that, divergence sharply falls to its baseline levels. During Phase 3, the signal spikes once again as pro-Russian and anti-government protests took place across the eastern and southern regions of Ukraine until the declaration of Donetsk and Luhansk People's Republics. The signal declines again in Phase 4 which marks the Ukrainian forces vs. pro-Russian militants fighting a war. The signal meets zero-line during the initial days of Phase 5 marking the Russian full-scale invasion which was framed as an "humanitarian convoy" crossing into the Ukrainian territory. Following that, the signal remains at its baseline levels with no more major breakouts.

7.9 Discussions and Future Work

A question arise: could Russian propaganda framing shifts forecast the onset of hostilities leading to an invasion? In the Ukraine case, the divergence signal's early rise, exceeding all previous highs in Nov. 2013 followed by the sharp rise in Dec 2013 provides a signal of interest three-four months ahead of Crimea's annexation. If the premise is accepted that information operations are intended to "soften-up" the target area and provide a pretext for active conflict, then shifts in strategic framing might provide an early warning before the onset of pro-Russian protests, militant action and invasion under the guise of an "humanitarian convoy".

Our future work involves various tasks. Since our classifiers achieved an average 77% F1-score only, we plan to experiment with additional syntactic and semantic (framenet, wordnet, verbnet, LIWC18) [3] features, and other features such as named entity types to improve performance.

Next, we believe it might be possible to automatically surface framing categories to help spot newly emerging framing categories. We aim to synthesize narrative graphs incorporating co-occurrence patterns [37] of discriminant bi-grams, their adverbs, adjectives, named entities (i.e. people, places, organizations and locations) and apply dynamic graph clustering algorithms [10] to detect newly emerging clusters for SME's attention. Our initial experiments indicate that we can surface expert induced framing categories developed in the Ukraine codebook with a Normalized Mutual Information (NMI) score of 56% and purity of 68%.

Finally, we plan to evaluate this framework in other historical contexts; such as the Transnistria War in November 1990 between Moldovan troops and pro-Transnistria forces supported by elements of the Russian Army and the Russo-Georgian War between Georgia, Russia and the Russian-backed self-proclaimed republics of South Ossetia and Abkhazia in August 2008.

[3] https://framenet.icsi.berkeley.edu/ https://wordnet.princeton.edu/ https://verbs.colorado.edu/verbnet/ https://liwc.wpengine.com/

Chapter 8

CONCLUSION AND FUTURE WORK

Social Computing is an area of computer science concerned with dynamics of communities and cultures created through computer-mediated social interaction. Various social media platforms, such as social network services and microblogging, enable users to come together and create social movements expressing their opinions on diverse sets of issues, events, complaints, grievances and goals. In this dissertation, the nature of disseminated information and the main players driving mobilization is the topic of this study. The interaction between users from varying ideologies is crucial for studying socio-politics and movements, the adversarial language among rival groups which polarize the conflict and the main players who influence the political landscape and the masses. These can be quite challenging to study as there is an enormous number of data and relational interactions reflecting the structure of societal allegiance. However, with the use of analytical and empirical methods guided by social computation, an examination is possible, and findings can be concluded. The impact of increasing social unrest and the promotion of hate speech creates a deceptive, fertile ground and charged environment, which triggers and exacerbates grievances for mass mobilization that aims to dismantle social cohesion in various nationalities. Such aggressive means can be achieved especially when societies suffer from severe inequality and economic distress, which may ignite political chaos and violence. Social computing uses experimental research design. It involves creating social conventions and contexts by collecting, representing, preprocessing and spreading of information. It is followed by the use of computational and machine learning approaches, such as community detection, classification, prediction, clustering, etc. In this dissertation, we determined how to find trending or breaking information at nascent stages through hashtags. Next, they search organizational accounts that act as rhetorical devices for political organizational narratives and propaganda. This work detects and studies contentious language rival groups dominating the political landscape, including phrases that elicit reactions between them. It also presents how users attention shifts according to events and significant developments, and what this reveals about users allegiances to political factions and their sentiments for various looming issues. Moreover, a visualization tool was proposed, which can enhance the ability of subject matter experts in understanding events and users shift in a complex political scene with high speed while avoiding the need to sift through heaps of social interactions contents. Finally, an empirical study was presented and examined how exogenous powers promote social unrest among minorities for political gain, expansionist ambitions and the fight over ideologies believed to threaten their hegemony.

In future work, the Information Volume Break-Out Detection chapter can be further enhanced by Long / Short Term Memory Neural Network due to the sequencing nature where a predictive model be used. Additionally, with regards to the Organizational Account Detection chapter, a simple structure of Neural Network can be employed for detection performance enhancement, whereas only a logistic regression (a single neuron of ANN for the sake of studying features polarity and its indication value) was used in this study. With regards to the Framing Shifts Detection of Conflicts chapter, Name Entity Recognition as feature complementing other bigrams and concepts. Another possible improvement is to tune the rules of the concept generator based on POS to help merge superficial word levels, indicating the same concepts, which potentially can enhance the frame detection classifier performance.

REFERENCES

[1] TASS: Russia - Russian Foreign Ministry presents White Book on human rights abuses in Ukraine. http://tass.com/russia/730463.

[2] Transcript: Putin says Russia will protect the rights of Russians abroad - The Washington Post. https://goo.gl/bacjU5.

[3] TSG IntelBrief: Russia's Passport Imperialism — The Soufan Group. http://www.soufangroup.com/tsg-intelbrief-russias-passport-imperialism/.

[4] Web 2.0 Summit 2010 - Co-produced by UBM TechWeb & O'Reilly Conferences, November 15 - 17, 2010, San Francisco, CA.

[5] Jensen-Shannon divergence and Hilbert space embedding. *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, 2004.

[6] Latvians reject Russian as official language — World news — The Guardian, 2012. https://goo.gl/wXR8K4.

[7] Debate — Definition of Debate by Merriam-Webster, 2016.

[8] debate — meaning of debate in Longman Dictionary of Contemporary English — LDOCE, 2016.

[9] Robert P Abelson. The structure of belief systems. *Computer models of thought and language*, pages 287–339, 1973.

[10] Riza Aktunc, Ismail Hakki Toroslu, Mert Ozer, and Hasan Davulcu. A dynamic modularity based community detection algorithm for large-scale networks: Dslm. In *ASONAM, IEEE/ACM*, 2015.

[11] Saud Alashri, Sultan Alzahrani, Lenka Bustikova, David Siroky, Hasan Davulcu, Decision Systems Engineering, and Global Studies. What Animates Political Debates ? Analyzing Ideological Perspectives in Online Debates between Opposing Parties. *in Proceedings of the ASE/IEEE International Conference on Social Computing (SocialCom-15)*, 2015.

[12] Saud Alashri, Sultan Alzahrani, Jiun-Yi Tsai, Steven R Corman, and Hasan Davulcu. "climate change" frames detection and categorization based on generalized concepts. *International Journal of Semantic Computing*, 10(02), 2016.

[13] Hana Alostad and Hasan Davulcu. Directional prediction of stock prices using breaking news on twitter. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2015 IEEE/WIC/ACM International Conference on*, volume 1, pages 523–530. IEEE, 2015.

[14] Mohamed Aly. Survey on multiclass classification methods. *Neural Netw*, pages 1–9, 2005.

[15] Sultan Alzahrani, Saud Alashri, Anvesh Reddy Koppela, Hasan Davulcu, and Ismail Toroslu. A network-based model for predicting hashtag breakouts in twitter. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, pages 3–12. Springer, 2015.

[16] Sultan Alzahrani, Betul Ceran, Saud Alashri, Scott W Ruston, Steven R Corman, and Hasan Davulcu. Story forms detection in text through concept-based co-clustering. In *SocialCom), 2016*. IEEE, 2016.

[17] Sultan Alzahrani, Chinmay Gore, Amin Salehi, and Hasan Davulcu. Finding organizational accounts based on structural and behavioral factors on twitter. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 2018.

[18] Sultan Alzahrani, Nyunsu Kim, Mert Ozer, Scott W. Ruston, Jason Schlachter, and Steve R. Corman. Framing shifts of the ukraine conflict in pro-russian news media. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 2018.

[19] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection *. *Statistics Surveys*, 4:40–79, 2010.

[20] Sitaram Asur, Bernardo A. Huberman, Gabor Szabo, and Chunyan Wang. Trends in Social Media: Persistence and Decay. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM'11)*, pages 434–437, 2011.

[21] Sitaram Asur, Bernardo A Huberman, Gabor Szabo, and Chunyan Wang. Trends in social media: persistence and decay. In *ICWSM*, 2011.

[22] Deborah L Bandalos and Meggen R Boehm-Kaufman. Four common misconceptions in exploratory factor analysis. *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*, pages 61–87, 2009.

[23] L. Baringhaus and C. Franz. On a new multivariate two-sample test. *Journal of Multivariate Analysis*, 2004.

[24] A Barrat, M Barthélemy, R Pastor-Satorras, and A Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, mar 2004.

[25] Eric P. S. Baumer, Elisha Elovic, Ying Crystal Qin, Francesca Polletta, and Geri K. Gay. Testing and Comparing Computational Approaches for Identifying the Language of Framing in Political News. *Acl*, pages 1472–1482, 2015.

[26] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188, 2001.

[27] Kenneth Benoit and Michael Laver. Estimating irish party policy positions using computer wordscoring: The 2002 election. *Irish Political Studies*, 2003.

[28] Brigitte Bigi. Using kullback-leibler distance for text categorization. In *European Conference on Information Retrieval*, pages 305–319. Springer, 2003.

[29] David M Blei, Blei@cs Berkeley Edu, Andrew Y Ng, Ang@cs Stanford Edu, Michael I Jordan, and Jordan@cs Berkeley Edu. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[30] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of mathematical sociology*, 2(1):113–120, 1972.

[31] Sergey Brin and Lawrence Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 56(18):3825–3833, 2012.

[32] Ethem F Can, Hüseyin Oktay, and R Manmatha. Predicting retweet count using visual cues. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 2013.

[33] Jaime G Carbonell. Politics: Automated ideological reasoning. *Cognitive Science*, 2(1):27–51, 1978.

[34] Jaime G Carbonell. Counterplanning: A strategy-based model of adversary planning in real-world situations. *Artificial Intelligence*, 16(3):295–329, 1981.

[35] Jaime Guillermo Carbonell. Learning by analogy: formulating and generalizing plans from past experience. 1982.

[36] Fabio Celli and Luca Rossi. The role of emotional stability in twitter conversations. In *Proceedings of the workshop on semantic analysis in social media*, pages 10–17. Association for Computational Linguistics, 2012.

[37] Betul Ceran, Nitesh Kedia, Steven R Corman, and Hasan Davulcu. Story detection using generalized concepts and relations. In *Advances in Social Networks Analysis and Mining (ASONAM), 2015 IEEE/ACM International Conference on*, 2015.

[38] Olivier Chapelle and Yi Chang. Yahoo! learning to rank challenge overview. In *Proceedings of the Learning to Rank Challenge*, pages 1–24, 2011.

[39] Justin Cheng, Lada Adamic, P. Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. Can cascades be predicted? *Proceedings of the 23rd international conference on World wide web*, pages 925–936, 2014.

[40] Dennis Chong and James N Druckman. A theory of framing and opinion formation in competitive elite environments. *Journal of Communication*, 57(1), 2007.

[41] C-K Chu and James Stephen Marron. Comparison of two bandwidth selectors with dependent errors. *The Annals of Statistics*, pages 1906–1918, 1991.

[42] Claudio Cioffi-Revilla. Computational social science. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):259–271, 2010.

[43] Guilherme Ferraz De Arruda, André Luiz Barbieri, Pablo Mart??n Rodr??guez, Francisco Aparecido Rodrigues, Yamir Moreno, and Luciano da Fontoura Costa. Role of centrality for the identification of influential spreaders in complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 90(3), apr 2014.

[44] Lalindra De Silva and Ellen Riloff. User type classification of tweets with implications for event recognition. *ACL 2014*, page 98, 2014.

[45] Maintainer Alex Deckmyn. Package mapdata. 2016.

[46] Jennifer Earl, Andrew Martin, John D McCarthy, and Sarah A Soule. The Use of Newspaper Data in the Study of Collective Action. *Annual Review of Sociology*, 30(1):65–80, 2004.

[47] J. McCorriston el al. Organizations are users too: Characterizing and detecting the presence of organizations on twitter. In *ICWSM*, 2015.

[48] Ahmed El-Kishky, Yanglei Song, Chi Wang, Clare R Voss, and Jiawei Han. Scalable topical phrase mining from text corpora. *Proceedings of the VLDB Endowment*, 8(3):305–316, 2014.

[49] Robert M Entman. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4), 1993.

[50] Conover et al. Predicting the political alignment of twitter users. In *Privacy, Security, Risk and Trust (PASSAT)*, pages 192–199. IEEE, 2011.

[51] D. Rao et al. Classifying latent user attributes in twitter. In *Proceedings of the 2nd international workshop on Search and mining user-generated contents*. ACM, 2010.

[52] J. Burger et al. Discriminating gender on twitter. In *EMNLP2011*, 2011.

[53] O. Varol et al. Online human-bot interactions: Detection, estimation, and characterization. *arXiv preprint arXiv:1703.03107*, 2017.

[54] Odijk et al. Automatic thematic content analysis: Finding frames in news. In *International Conference on Social Informatics*. Springer, 2013.

[55] S. Wu et al. Who says what to whom on twitter. In *Proceedings of the 20th international conference on World wide web*. ACM, 2011.

[56] Sunghwan Mac et al. Distinguishing individuals from organisations on twitter. In *WWW2017*, 2017.

[57] Yasugi et al. Use of twitter as an instrument for disseminating public information in providing public goods and roles of e-government: Evidence from japanese prefectures. *International Journal of Engineering and Innovative Technology*, 2013.

[58] Linton C. Freeman. A Set of Measures of Centrality Based on Betweenness, mar 1977.

[59] Sravan Kumar Garipalli. Visualization tool for islamic radical and counter radical movements and their online followers in south east asia. Master's thesis, Arizona State University, 2015.

[60] Kristian Skrede Gleditsch. Civil war from a transnational perspective kristian skrede gleditsch, university of essex & peace research institute oslo (prio). *Oxford University Press*, pages 1–20, 2017.

[61] Lukasz Golab and M Tamer Özsu. Issues in Data Stream Management. *SIGMOD Rec.*, 32(2):5–14, 2003.

[62] Rebecca Goolsby. On Cybersecurity, Crowdsourcing and Social Cyber-Attack. *Wilso Center, policy memo series*, 1:1–8, 2012.

[63] Chinmay Gore. Detecting Organizational Accounts from Twitter Based on Network and Behavioral Factors. Master's thesis, Arizona State University, 2017.

[64] Tom Gruber. Collective Knowledge Systems: Where the Social Web meets the Semantic Web. *Semantic Web and Web 2.0*, 6(1):4–13, 2008.

[65] Pedro Henrique Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. A measure of polarization on social media networks based on community boundaries. In *ICWSM*, 2013.

[66] Ralf Herbrich, Thore Graepel, and Klaus Obermayer. Support vector learning for ordinal regression. 1999.

[67] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[68] Thomas Hofmann. Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 289–296. Morgan Kaufmann Publishers Inc., 1999.

[69] Bernardo A Huberman, Daniel M Romero, and Fang Wu. Social networks that matter: Twitter under the microscope. *Available at SSRN 1313405*, 2008.

[70] Internet World Stats. Bangladesh Internet Usage and Telecommunications Reports.

[71] Jacob Resneck. In tiny Moldova, Russia is repeating its Ukraine playbook — Public Radio International. https://goo.gl/PQR5XA.

[72] James Graham. Russia and Ethnic Conflict in Georgia - On This Day. https://www.onthisday.com/russia/georgia.php.

[73] James Kifield. How to Prevent War with Russia - POLITICO Magazine. https://goo.gl/ocLo55.

[74] M Sharon Jeannotte. Social cohesion: Insights from canadian research. *Gatineau: Department of Canadian Heritage.(2000), Social cohesion around the world: An international comparison of definitions and issues, Strategic Research and Analysis, Ottawa: Canadian Heritage*, 2003.

[75] Sarvnaz Karimi, Jie Yin, and Jiri Baum. Evaluation Methods for Statistically Dependent Text. *Computational Linguistics*, 41(3):539–548, 2015.

[76] Yoshikiyo Kato, Sadao Kurohashi, Kentaro Inui, Robert Malouf, and Tony Mullen. Taking sides: User classification for informal online political discourse. *Internet Research*, 18(2):177–190, 2008.

[77] Elihu Katz. The two-step flow of communication: An up-to-date report on an hypothesis. *Public opinion quarterly*, 1957.

[78] Irwin King, Jiexing Li, and Kam Tong Chan. A brief survey of computational approaches in Social Computing. *Proceedings of the International Joint Conference on Neural Networks*, pages 1625–1632, 2009.

[79] Breanna V Kingsley. The Effects that War Has on Children and Child Soldiers. page 45, 2017.

[80] Matthew G Kirschenbaum. The remaking of reading: Data mining and the digital humanities. In *The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, Baltimore, MD*. Citeseer, 2007.

[81] Konstantin Klemm and Victor M Eguiluz. Growing scale-free networks with small-world behavior. *Physical Review E*.

[82] Paul H Klingbiel. Machine-aided indexing of technical literature. *Information Storage and Retrieval*, 9(2):79–84, 1973.

[83] K. Krippendorff. *Content analysis: An introduction to its methodology*. Sage, 2004.

[84] Stephan Kyburz. Ethnicity, Inequality, and Civil War. 2010.

[85] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1188–1196, 2014. `http://proceedings.mlr.press/v32/le14.pdf`.

[86] Chenliang Li, Aixin Sun, and Anwitaman Datta. Twevent: Segment-based Event Detection from Tweets. *CIKM*, pages 155–164, 2012.

[87] Kar Wai Lim and Wray Buntine. Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 1319–1328. ACM, 2014.

[88] Fei Liu, Feifan Liu, and Yang Liu. Automatic keyword extraction for the meeting corpus using supervised approach and bigram expansion. In *Spoken Language Technology Workshop, 2008. SLT 2008. IEEE*, pages 181–184. IEEE, 2008.

[89] Charles G Lord, Lee Ross, and Mark R Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11):2098, 1979.

[90] Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. Home location identification of twitter users. *ACM Transactions on Intelligent Systems and Technology*, 2014.

[91] Marc Ambinder. Russia masters the art of clandestine warfare against Ukraine. https://goo.gl/dG3EzD.

[92] Max Seddon. Russia Warns Of "Unfortunate Consequences" Over Ethnic Tension In Baltic States. https://goo.gl/rfA39V.

[93] Ulises A. Mejias and Nikolai E. Vokuev. Disinformation and the media: the case of Russia and Ukraine. *Media, Culture and Society*, 39(7):1027–1042, 2017.

[94] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013. `https://arxiv.org/pdf/1301.3781.pdf`.

[95] Burt L Monroe, Michael P Colaresi, and Kevin M Quinn. Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403, 2008.

[96] Diana C. Mutz and Lori Young. Communication and public opinion: Plus ÇA change? *Public Opinion Quarterly*, 75(5 SPEC. ISSUE):1018–1044, 2011.

[97] B. Nacos. Politics and the twitter revolution: How tweets influence the relationship between political leaders and the public. *Political Science Quarterly*, 2013.

[98] Zoran Nenadic and Joel W Burdick. Spike Detection Using the Continuous Wavelet Transform. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, 2005.

[99] M. E J Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, 2005.

[100] Dong-Phuong Nguyen, Rilana Gravel, Rudolf Berend Trieschnigg, and Theo Meder. " how old do you think i am?" a study of language and age in twitter. 2013.

[101] James Norriss. Othering 101: What Is Othering? — There Are No Others. Technical report, 2011. https://therearenoothers.wordpress.com/2011/12/28/othering-101-what-is-othering/.

[102] Joseph S Nye. Soft power. *Foreign policy*, (80):153–171, 1990.

[103] Daan Odijk, Björn Burscher, Rens Vliegenthart, and Maarten De Rijke. Automatic thematic content analysis: Finding frames in news. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8238 LNCS:333–345, 2013.

[104] Mert Ozer, Nyunsu Kim, and Hasan Davulcu. Community detection in political twitter networks using nonnegative matrix factorization methods. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*, pages 81–88. IEEE, 2016.

[105] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[106] Kishore Papineni. Why inverse document frequency? In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*, pages 1–8. Association for Computational Linguistics, 2001.

[107] Karl Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(1):559–572, nov 1901.

[108] F Pukelsheim. The three sigma rule. *The American Statistician*, 48(2):88–91, 1994.

[109] Kevin M Quinn, Burt L Monroe, Michael Colaresi, Michael H Crespin, and Dragomir R Radev. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, 54(1), 2010.

[110] Paul Resnick, Samuel Carton, Souneil Park, Yuncheng Shen, and Nicole Zeffer. Rumorlens: A system for analyzing the impact of rumors and corrections in social media. In *Proc. Computational Journalism Conference*, 2014.

[111] Ryan Rifkin and Aldebaro Klautau. In defense of one-vs-all classification. *Journal of machine learning research*, 5(Jan):101–141, 2004.

[112] Suman Deb Roy, Gilad Lotan, and Wenjun Kevin Zeng. The attention automaton: Sensing collective user interests in social network communities. *IEEE Transactions on Network Science and Engineering*, 2(1):40–52, 2015.

[113] Gert Sabidussi. The centrality index of a graph. *Psychometrika*, 1966.

[114] W. Safire. ON LANGUAGE; The Near Abroad. *The New York Times*, may 1994.

[115] Gerard Salton and Christopher Buckley. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5):513–523, 1988.

[116] David Savage, Xiuzhen Zhang, Xinghuo Yu, Pauline Chou, and Qingmai Wang. Anomaly detection in online social networks. *Social Networks*, 2014.

[117] Roger C Schank and Robert Abelson. Scripts, goals, plans, and understanding, 1977.

[118] Dietram A Scheufele and David Tewksbury. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication*, 2007.

[119] Mario Schmidt. The sankey diagram in energy and material flow management. *Journal of industrial ecology*, 12(1):82–94, 2008.

[120] Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1):1–47, 2002.

[121] Alexandra Segerberg and W. Lance Bennett. Social media and the organization of collective action: Using twitter to explore the ecologies of two climate change protests. *Communication Review*, 14(3):197–215, 2011.

[122] Stephen B Seidman. Network structure and minimum degree. *Social networks*, 1983.

[123] Robert P. Sheridan. Time-split cross-validation as a method for estimating the goodness of prospective prediction. *Journal of Chemical Information and Modeling*, 53(4):783–790, 2013.

[124] Paul M. Sniderman and Sean M. Theriault. The Structure of Polical Argument and the Logic of Issue Framing. In *Studies in Public Opinion: Attitudes, Nonattitudes, Measurement Error, and Change*, volume 3, pages 133–65. 2004.

[125] Michael Stonebraker and Stan Zdonik. The 8 Requirements of Real-Time Stream Processing. 34(4):42–47, 2005.

[126] B. D. Tapley, Srinivas Bettadpur, John C Ries, Paul F Thompson, and Michael M Watkins. GRACE Measurements of Mass Variability in the Earth System. *Science*, 2004.

[127] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[128] TNS. Arab social media report first report. 2015.

[129] I. Tomek. Two modifications of cnn. *IEEE Trans. Systems, Man and Cybernetics*, 1976.

[130] Jean M Twenge and W Keith Campbell. *The narcissism epidemic: Living in the age of entitlement*. Simon and Schuster, 2009.

[131] Thijs Van de Graaf and Jeff D. Colgan. Russian gas games or well-oiled conflict? Energy security and the 2014 Ukraine crisis. *Energy Research and Social Science*, 24:59–64, 2017.

[132] Graham Vickery and Sacha Wunsch-Vincent. *Participative web and user-created content: Web 2.0 wikis and social networking*. Organization for Economic Cooperation and Development (OECD), 2007.

[133] Claudia Wagner and Markus Strohmaier. The wisdom in tweetonomies: Acquiring latent conceptual structures from social awareness streams. In *Proceedings of the 3rd International Semantic Search Workshop*, page 6. ACM, 2010.

[134] Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1031–1040. ACM, 2011.

[135] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 1998.

[136] Jianshu Weng, Yuxia Yao, E Leonardi, and F Lee. Event Detection in Twitter. *Development*, pages 401–408, 2011.

[137] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. Virality prediction and community structure in social networks. *Scientific reports*, 3:2522, 2013.

[138] Yingcai Wu, Shixia Liu, Kai Yan, Mengchen Liu, and Fangzhao Wu. Opinionflow: Visual analysis of opinion diffusion on social media. *IEEE transactions on visualization and computer graphics*, 20(12):1763–1772, 2014.

[139] Jaewon Yang and Jure Leskovec. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 177–186. ACM, 2011.

[140] Ding Zhou, Sergey A Orshanskiy, Hongyuan Zha, and C Lee Giles. Co-ranking authors and documents in a heterogeneous network. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 739–744. IEEE, 2007.

[141] Ali Zifan. Libyan Civil War Wikimedia Commons. Technical report, 2016.