Analyzing the Impact of Cultural Factors on Happiness Levels in Arabic Language Tweets

Parisa Suchdev
University of Vermont

Analyzing the Impact of Cultural Factors on Happiness Levels in Arabic Language Tweets

A Thesis Presented

by

Parisa Suchdev

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Computer Science

August, 2023

Defense Date: July 18th, 2023
Thesis Examination Committee:

Chris Danforth, Ph.D., Advisor
Peter S. Dodds, Ph.D., Advisor
Pablo S. Bose, Ph.D., Chairperson
Donna M. Rizzo, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College

Abstract

Culture is a fundamental force shaping our view of the world. Filtered through the stories we share on social media, our collective behavior both reflects and amplifies cultural impacts. The present study seeks to describe the effect of cultural factors, such as religion, on happiness scores in Arabic language tweets from January 2010 to June 2023. Our methodology involves using present tools called Hedonometer (https://hedonometer.org/) to study happy and sad events and StoryWrangler (https://storywrangling.org/) to study the usage of keywords related to those events. Our findings reveal a notable pattern of Twitter happiness declining following the start of the Arab Spring in 2010, and this decline persisted until around 2013. This pattern reflects the long-lasting impact of significant Arab Spring events on the overall happiness score on Twitter. Arabic culture seems to be heavily influenced by religion, with religion often being the driving force behind happy events, while acts of violence and conflict related to political and religious injustice are associated with sad events.

Acknowledgements

It was a dream come true for me to come to the United States for higher education. That is due to a number of people. I came here as a confused, scared girl who simply wanted to do something but didn't know what. UVM and its people have made me a strong, confident woman with the ability to think in ways I never imagined. It has given me a purpose that is not solely for myself. It has helped me grow and become a better person, both emotionally and mentally.

I'd like to thank my family - Mama, Papa, my sister, and my brother - for being pillars of strength and unwavering support throughout my journey. Your love and encouragement have guided me and shaped me into the person I am today.

I would like to express my heartfelt gratitude to my esteemed advisors, Chris Danforth and Peter S. Dodds, for their invaluable mentoring and kindness. Your advice has been invaluable in shaping my research and academic endeavors. I am also grateful to Michael Arnold, a diligent PhD student, whose generous support and countless hours of assistance were critical to the success of my thesis.

Juniper Lovato, you have been thoughtful, wise, and supportive. I'm glad I met you because you provided me with my first research experience.

I'd like to thank my cousin Rakesh in particular, whose constant encouragement and constant trust in me has been a driving force. Your constant presence, always a phone call away, has been a true blessing.

My friends, your constant encouragement and shared laughter have made my academic journey more enjoyable.

A heartfelt thank you to the professors who have imparted knowledge and wisdom during my master's degree, specially Professor Byung Lee, whose guidance in the field of computer science has been invaluable.

Lastly, I am grateful to the University of Vermont for providing me a home away from home. UVM's nurturing environment has played a significant role in my academic growth.

Table of Contents

List of Figures

Chapter 1

Introduction

Communication is as old as humankind, from the early days when a few people gathered around campfires to share stories, emotions, and beliefs, to now, when individual thoughts have the power to spread to millions of people in matter of seconds through social media. Since the beginning of human civilization, the mechanism of information sharing has evolved and laid the foundation for culture, including its norms, values, and traditions [1]. It is natural for humans to create methods to communicate [1]. As oral communication began to be encoded into writing, the thoughts of a small group of people were recorded. From cave art to papyrus, from the printing press to the internet, the medium has evolved to the point now where the expressed thoughts of billions of everyday people are being recorded electronically. Social platforms such as Twitter allow us to express opinions, while simultaneously keeping digital records, providing us with the opportunity to explore unfiltered opinions and emotions at the population scale.

Digitized opinions and emotions offer a great deal of research opportunities, including our present focus of analyzing the impact of cultural events have on tweets shared by people who speak different languages from different regions. The Arabic language is ranked among the top 10 on Twitter in terms of The Arabic language is ranked among the top 10 on Twitter in terms of the number of tweets [2], and currently there are a limited number of studies focused on social media analysis of the Arab world [3]. Although there are several studies that focus on linguistic analysis or sentiment analysis of Arabic, for example, Al Shehii et al. [3] focused on cross-linguistic analysis of tweets by evaluating happiness patterns in English and Arabic originating in the UAE, in another paper, Alshaabi et al. [2] focused on the increasing fraction of retweets in more than 150 languages without particularly focusing on Arabic. However, to our knowledge, no studies have assessed overall trends in the Arabic language.

Therefore, the objective of this paper is to understand the effect of various cultural aspects, such as religion, traditions, and beliefs, on the happiness of tweets in the Arabic language.

On Twitter, individuals looking to contribute can post something original (tweet), re-post existing posts (retweet), quote existing posts while adding commentary (quote tweet), reply to existing posts (reply), or simply endorse a message (like). Users can select which language to be the default for reading, and translate messages in other languages with a 'translate tweet' button in over 100 commonly spoken languages [2]. Using a random 10% of all public tweets between 2009 and 2020, which included more than 118 billion tweets, including more than 100 different languages [2], 7 billion of those tweets were sent in Arabic. This access to linguistic data provides the opportunity to study psychological and sociocultural variables in a systematic way to produce useful information [3].

As a natural language, Arabic is fascinating due to its rich cultural and literary history [4, 5], and challenging due to its complicated structure [4–6]. Arabic could be compared to all other languages, including English, in terms of loan words and grammar; however, due to its distinctiveness, it is important to comprehend the complexity of the data [5]. Consequently, Natural Language Processing (NLP) has made significant progress analyzing social media [2, 7–9]. This progress has also evident in the field of sentiment analysis [2, 10–13].

Sentiment analysis, in the simplest definition, is the attempt to characterize emotions using computational methods within a specific context. A three-dimensional model called PAD (Pleasure, Arousal, Dominance) is presented by Mehrabian and Russell [14] to describe the complexity of emotion. The evidence presented in this article defines these dimensions as bipolar because each of them can vary from one extreme to the other, such as the pleasure-displeasure dimension, high-low (arousal), and dominance-submissiveness can range from pain-happiness, sleep-alertness, and helplessness-control, respectively. The authors state that a person can be placed on any region of these dimensions because one is considered to be in emotional state all the time. In another paper [3], called Arabia Felix 2.0, these states are defined as dynamic and are influenced by the range of several internal and external factors that may include weather, social interactions, or subjective perspectives. Currently, people have the means to express their emotional states directly or indirectly through social media platforms, allowing researchers to investigate how different affective states change with time, place, and the particular object of the attitude under study [3]. To examine how people's emotions change in response to topics such as religion, political candidates, or smoking, social media data are used, leading to the development of different sophisticated sentiment analysis techniques [3].

One of the techniques of sentiment analysis is to calculate the happiness score by quantifying the level of happiness expressed within a text. Plato believed that the achievement of eudaimonia, or flourishing, defines the ultimate goal of human life [15, 16], while Bentham and Stuart Mill believed that the collective happiness of society is the motive of life [16, 17]. The term happiness is subjective and the definitions may vary according to the culture, but humans generally desire well-being and avoid suffering in their true behavior [16, 18–21].

To quantify the 'happiness', we attempt to follow method introduced by Dodds and Danforth [16] for a hedonometer based on a single dimension of pleasure-displeasure (happy-sad) valence by using semantic differential technique introduced by Osgood, et al. [16, 22, 23]. In their 2015 paper, the authors have selected 10 different languages to construct the 24 corpora of text data from various sources, including using books, news outlets, social networks, movie subtitles, and music lyrics [23]. First, they measured the relative importance of words by calculating the frequency of their appearance in a variety of texts. Second, they surveyed native speakers of each language to rate how they felt in response to words in isolation on a scale of 1-9, ranging from saddest to happiest. Fifty individual ratings were averaged for each word, resulting in the LabMT dataset (Language assessment by Mechanical Turk). With happiness scores for the most frequent 10,000 words in each language, the authors calculated the happiness score of several texts, e.g., books by a specific author or news articles on a specific date. To do this, they calculate an average by summing the happiness scores for all words present in text, and divide by the total number of words (Equation 1.1). The resulting average was treated as a proxy for the emotional temperature of the bag of words being considered.

$$H = \frac{1}{N} \sum_{i=1}^{N} w_i s_i \tag{1.1}$$

where N is the total number of words in the text, w_i is the frequency of the ith word in the text, s_i is the LabMT happiness score of the ith word (on a scale of 1 to 9) and H is the happiness score calculated for the text.

While equation 1.1 offers a straightforward method for calculating a text's overall happiness score, it may not accurately reflect the text's emotional content because each word is assigned the same weight in the tweet independent of its frequency. As an improvement, equation 1.2 [10] was introduced, the happiness scores for each word are weighted according to their frequency in a given text. As a result, words that occur more frequently in the text are therefore given more weight.

$$h_{\text{avg}}(T) = \frac{\sum_{i=1}^{N} h_{\text{avg}}(w_i) f_i}{\sum_{i=1}^{N} f_i} = \sum_{i=1}^{N} h_{\text{avg}}(w_i) p_i \tag{1.2}$$

where $h_{\text{avg}}(T)$ is the happiness score for a particular text T, $h_{\text{avg}}(w_i)$ is the average happiness score for a word w_i, f_i is the frequency of the word w_i in the text T, and p_i is the corresponding normalized frequency.

By utilizing this equation 1.2, the authors present a tool called a hedonometer that can be used to measure and track the happiness level of tweets in real time. It can help to examine happiness trends and timelines in several languages, including Arabic. Clicking on the happiness trends also generates word shifts, which are discussed in more detail in 2.2. These word shifts highlight the specific words that contribute to a particular happiness score. To further analyze this, we employ Twitter's advanced search function to find tweets containing those keywords on the specific dates within the timeline. In this analysis, a human evaluator plays a crucial role in identifying the events or circumstances that led to the generation of that particular score. Additionally, any noticeable spikes in the data are identified and labeled by the human evaluator to provide context. Using this tool, researchers compared the levels of happiness among various demographic groups as well as temporal patterns of happiness across various locations and events. For example, researchers have used it to study discussion of breast cancer [24], exercise [25], and police violence [26] on Twitter.

The underlying assumption of the hedonometer is that the frequency of happy words reflects the general happiness of people using Twitter at a given time and is based on happy-sad valence [16]. In this study, we use this tool to study the trends of events that impacted the happiness level of tweets in Arabic as it uses 10% sample of tweets from 2008 up to the present time.

Consequently, to study volumes and trends of keywords, we use Alsahaabi et al [2] who describe StoryWrangler, a research tool that enables scholars to analyse enormous amounts of Twitter data to observe patterns in sociolinguistic, cultural, socioeconomic, and political trends over time. It employs natural language processing and unsupervised learning to classify tweets into coherent narratives or "stories." These stories are then displayed on a timeline and evaluated to discover trends and patterns in the data. Researchers can also filter and search the data based on various criteria, including keywords or user demographics, to produce customised subsets of data for analysis. It uses a 10% sample of tweets from 2008 up to the present time..

Therefore, the aim of this paper is to explore the impact of culture on all Arabic tweets by evaluating the 10% of the public tweets available from 2010 to 2023 through hedonometer. We look at this tool in two following approaches:

1. analyze the overall timeline of trends in the happiness levels of tweets.
2. determine the major cultural changes, events, and norms that affect the happiness levels of users.

This paper poses several research questions:

1. What was the impact of the Arab Spring (series of widespread protests and uprisings across the Arab world) [27] on the happiness levels of the tweets?
2. What are the common characteristics of Arabic tweets that contribute to happy events?
3. What are the common characteristics of Arabic tweets that contribute to sad events?
4. How do religious beliefs impact the happiness of the tweets?

It can be difficult to use social media sites like Twitter as a window into the outside world because they frequently overrepresent some demographic groups while underrepresenting others, leading to skewed data. Additionally, self-presentation bias may come into play as users may curate or idealize their lives on social media, and pretend to feel certain emotions like happiness to conform to social standards. Hence, interpreting happiness scores based on social media requires caution.

Chapter 2

Material and Methods

2.1 Material

We used the Hedonometer tool to analyze the overall trends and StoryWrangler tool to analyze volume of keywords and languages.

Figure 2.1 illustrates the daily number of Arabic speakers (unique twitter IDs) found in the Decahose sample from 2010 - 2023. The number rises exponentially during the Arab Spring of 2010, then linearly from roughly 100,000 to 1 million between 2012 and 2014. Nearly a million daily Arabic tweets are observed from 2014-2017, after which the count oscillates while slowly decaying to present levels of 1/2 million in 2023.

Figure 2.2 plots longitude and longitude coming from less than 1% of geo-tagged tweets highlighting the location of Arabic tweets and Figure 2.3 shows the top 15% of countries from which Arabic tweets originate, with Saudi Arabia, Kuwait, Egypt, UAE, and the USA ranking among the top five.

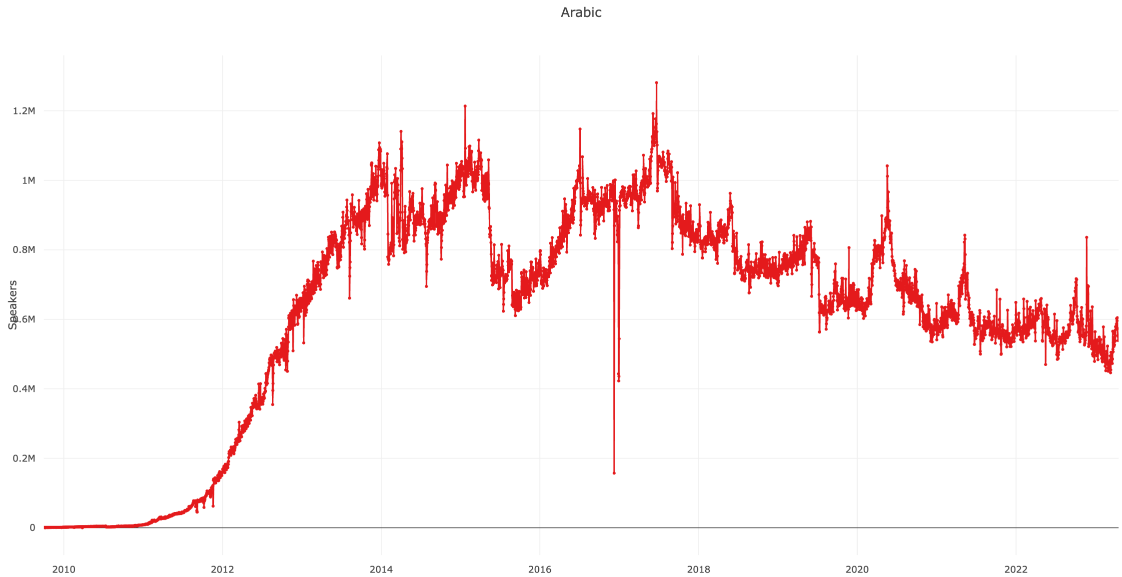


Figure 2.1: StoryWrangler: Time-series of daily speakers posting Arabic tweets of from 2010 - 2023, data from http://storywrangling.org

2.2 Methods

2.2.1 Overall Arabic Trends ft Hedonometer:

The instrument hosted on http://hedonometer.org visualizes a timeline of tweets based on their daily average happiness score, with higher spikes indicating happier tweets and lower spikes indicating sadder tweets. The website also provides the option to generate timelines for specific time periods and highlight trends on specific days of the week. We used this tool to identify events that contributed to the happy-sad valence of the tweets. Figure 2.4 illustrates the Hedonometer timeline, with the highest spikes representing the happiest days and the lowest spikes representing the saddest days. The seven differently colored circles at the top of each spike represent

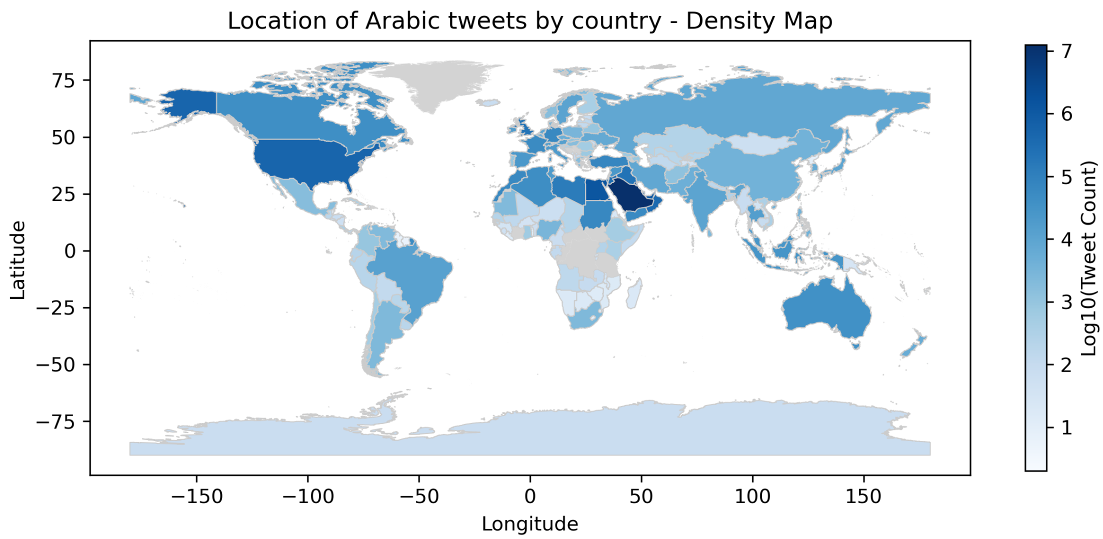


Figure 2.2: Where the tweets are coming from

each day of the week.

To determine which events were responsible for the spikes in the timeline, we clicked on each spike and reviewed the word-shift, as seen in Figure 2.5. By analyzing the word-shift, we could see which keywords contributed to the happiness and sadness of the tweets on that particular event. To analyze those events further, we used the keywords and dates of that particular timeline to run an advanced search on twitter to see what those tweets discussed.

Labelling spikes on the timeline

Although the website generates the Hedonometer timeline for Arabic tweets, it does not label the spikes with event names and information. While the keywords that caused high or low happiness scores were visible, the events that triggered the use of

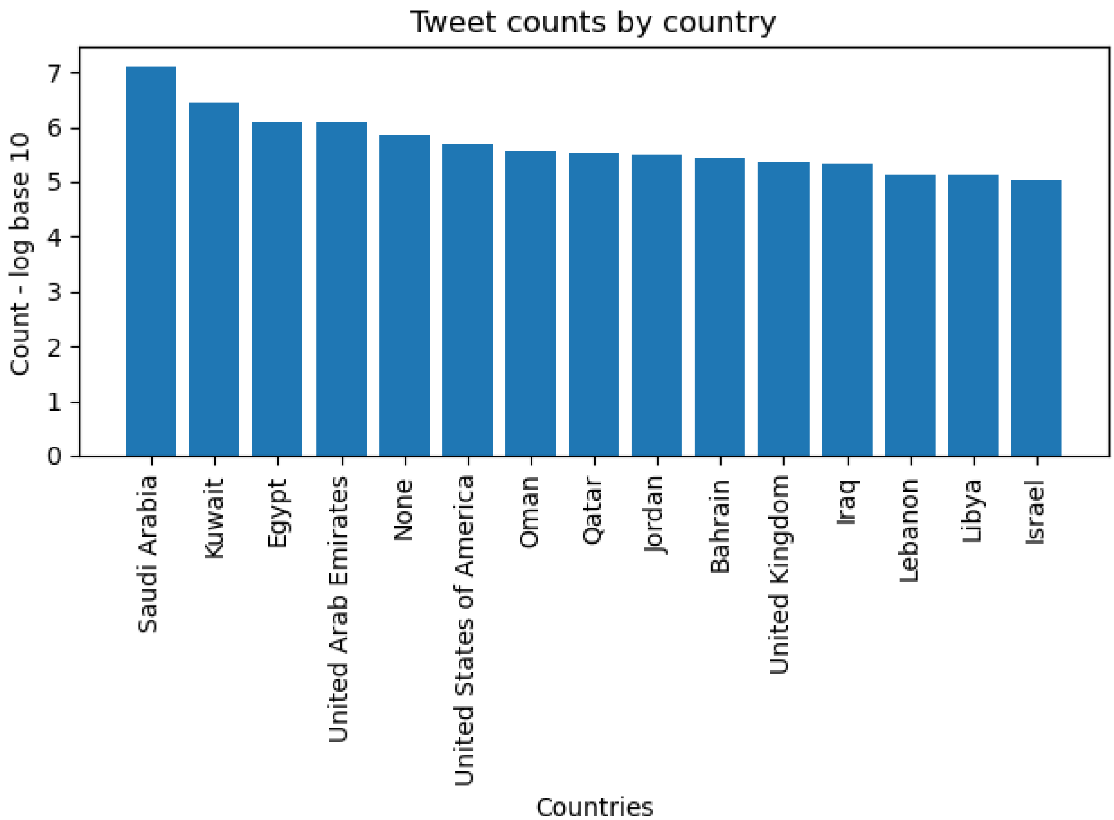


Figure 2.3: Top 15 countries of from 1% geolocated tweets

those keywords were not. As events are closely tied to human emotions and provide insight into culture, it was imperative to study them for this research. To label the major spikes in the data, we manually searched for tweets containing the keywords identified in the word-shifts and determined the events that prompted people to use happy or sad words on specific days of the timeline. This approach enabled us to identify the events that influenced happiness levels and to explore their impact.

Figure 2.5 shows the labelled happy event called Eid-ul-Adha which is widely celebrated by the Muslim community. The average happiness score for this event was 6.17, and the increase in positive words such as 'God' and 'Eid' and the decrease in

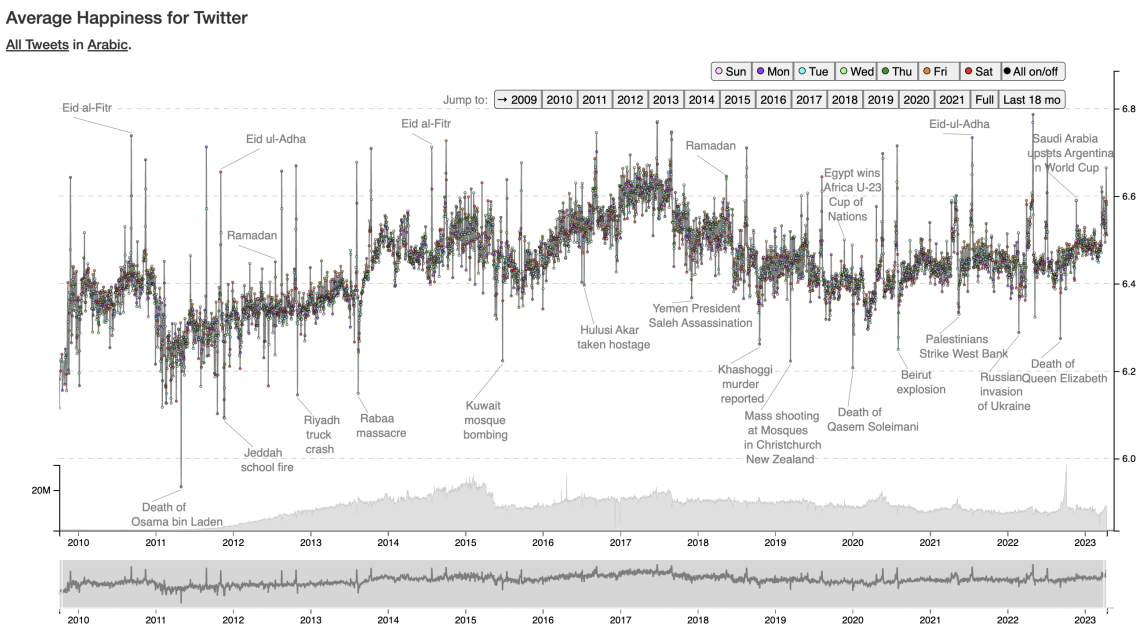


Figure 2.4: Hedonometer: estimated sentiment for a random 10% of Arabic tweets from 2010-2023

the negative words such as 'Fire' contributed to its high happiness score.

2.2.2 Keywords & languages ft StoryWrangler:

This enables users to gain insight into the number of tweets, retweets, speakers from a certain language as shown in 2.1, or the use of keywords in tweets in several languages over time as shown in 3.15. As previously mentioned, we use the hedonometer to examine the event influencing high and low happiness scores. Subsequently, we use wordshifts of these events to identify the keywords contributing to specific happiness scores and then to understand the evolving patterns of these keywords over time, we leverage the visualizations generated through storywrangler.

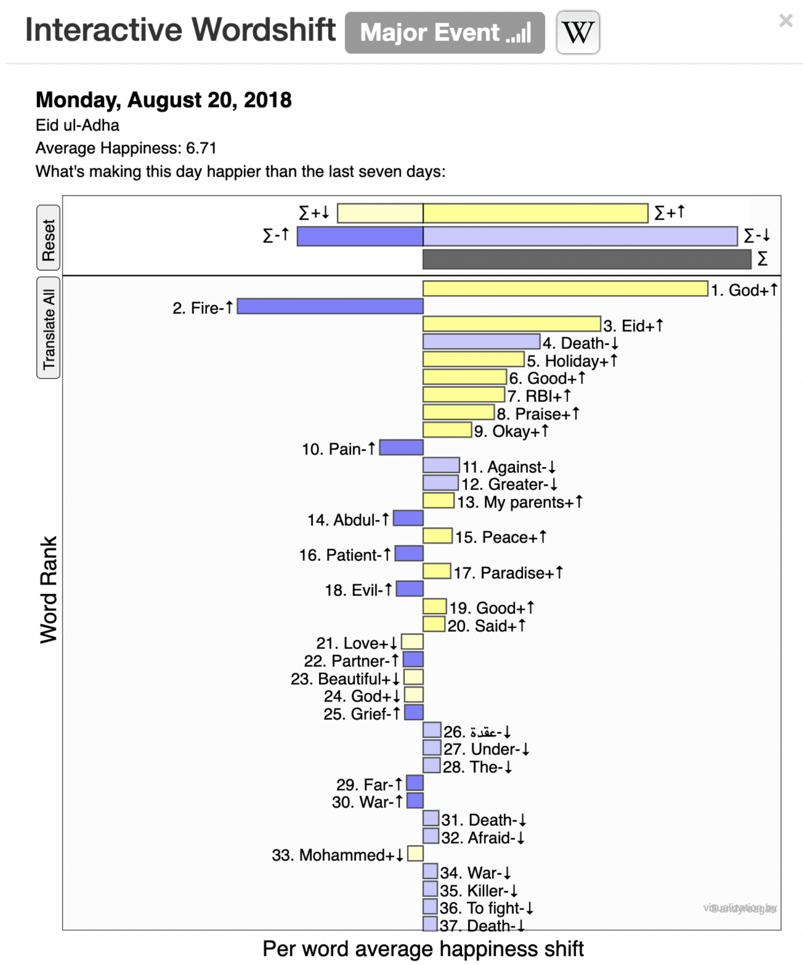


Figure 2.5: Hedonometer: Word-shift of Eid-ul-Adha on August 20, 2018

Chapter 3

Results

In this section, we discuss each research question in depth in correspondence with results found.

Figure 2.4 presents the overall picture of Arabic happiness trends captured by the Hedonometer. There are a few noticeable trends that are worth discussing in this chapter. First, according to Figure 3.1, there appears to be a decline in the happiness score at the end of 2010 coinciding with when the Arab Spring [27] started and it takes until late 2013 for it to return to its normal level, as shown in Figure 3.2. The precise end date of the Arab Spring is not generally acknowledged because it involved numerous protests, uprisings, and political changes in numerous Arab countries, each with its own timeline of events. Some believe that the Arab Spring came to an end with the dissolution of significant protest movements or the removal of particular governments, but others believe that it is still a process that will have long-term effects. Second, in addition to a few sports events, there appear to be recurring annual trends of different Eids and Ramadan throughout the time series that contribute to high spikes. Third, it appears that the trends driving the low spikes are derived by some sort of violence or tragedy.

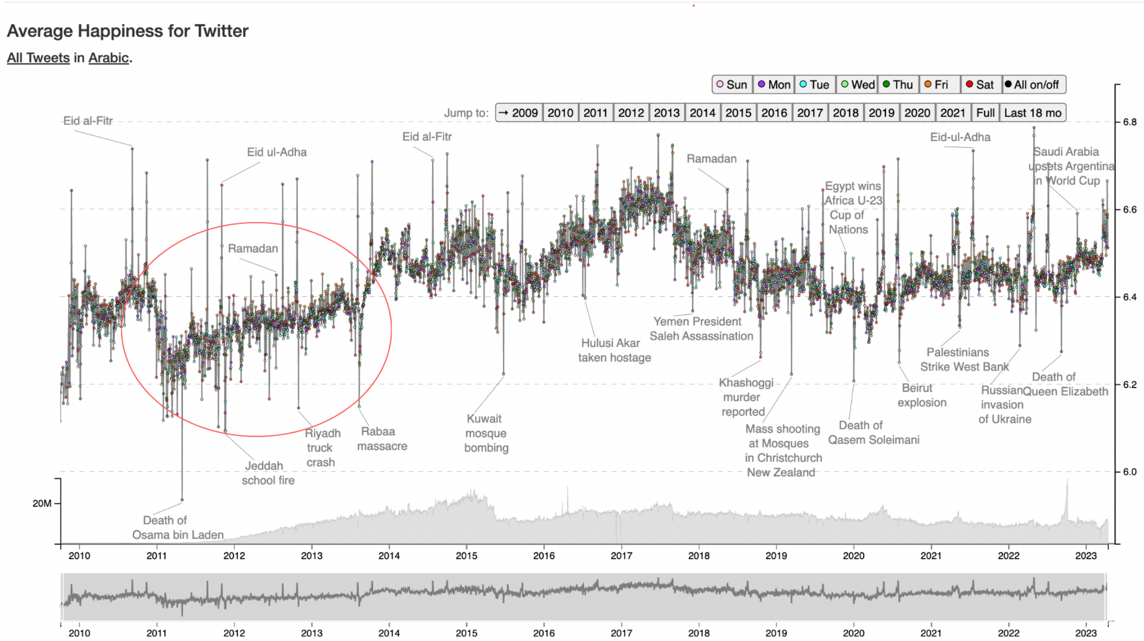


Figure 3.1: Hedonometer: drop in happiness score during Arab Spring

The Arab Spring: It all began with the "jasmine revolution" in December 17, 2010 that initiated the Arab Spring in Sidi Bouzid, Tunisia [28] after Mohamed Bouazizi burned himself outside the government office because the police seized his scales humiliatingly when he was operating as a street vendor without the required permit [27, 29]. The paper "Tear Gas and Twitter" states that the protests in Tunisia were fueled by digitally savvy activists and widespread use of social media, especially Facebook, which was quickly acquired by users after the Arabic language was launched in 2009 [27]. Defiant images of city Sidi Bouzid protesters circulated rapidly, aided by Al Jazeera's news channel broadcasting of social media videos [27]. Despite police violence, the movement grew uncontrollable, ultimately leading to the exile of President Ben Ali. The fire that sparked in Tunisia spread to Egypt on January 25,

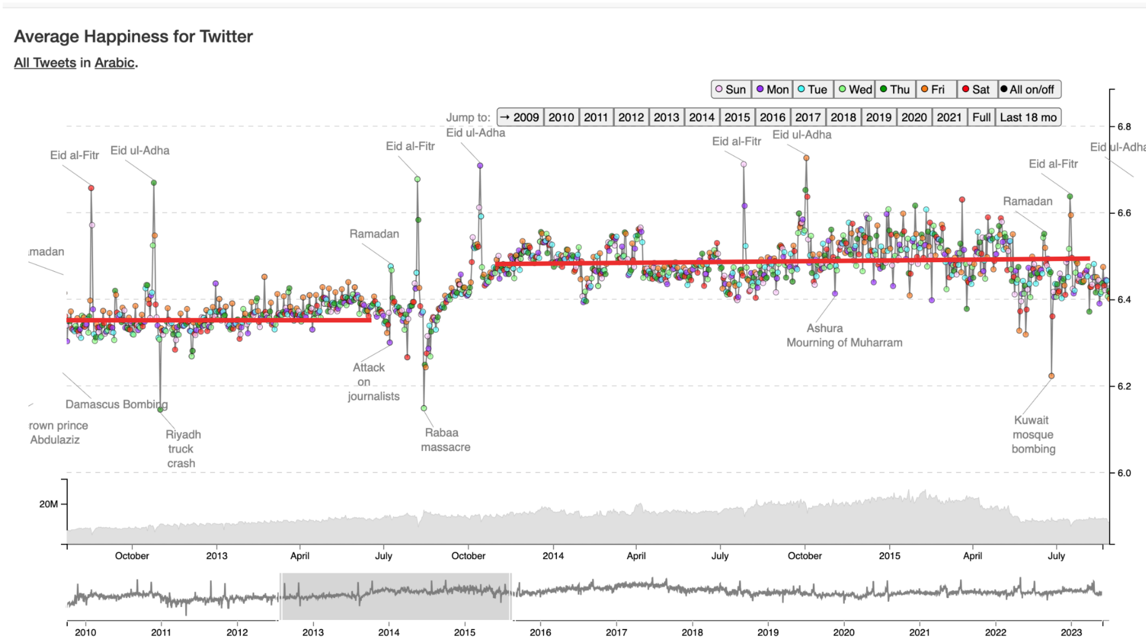


Figure 3.2: Hedonometer: Comparision of average levels

2011, with the Twitter hashtag #jan25 symbolizing the subsequent Tahrir uprising. As a result, then President Hosni Mubarak resigned from his post. [27]. Popular uprisings successfully overthrew dictators in Tunisia, Egypt, and Libya. In addition, ongoing protests in Yemen and Syria continue to pose significant threats to their respective dictators [30]. Bahrain had witnessed massive political protests, which had received substantial violence from the authorities, while Morocco had also experienced large-scale protests [30]. Therefore, Arab Spring has different conclusion in different countries.

Before delving into the happiness analysis of Arabic tweets for the Arab Spring event, we must first consider if Twitter truly offers a window into this complicated turmoil. According to the paper [27], Twitter played a significant role in the Arab Spring, but it was not the only platform for communication and activism during that time. Before the protests, platforms like Facebook and blogs were also popular in Tunisia, allowing for political discussions and resisting government censorship. In Egypt, a minority had internet access, enabling online political engagement. Moreover, this paper [27] also discusses that digital connectivity extended beyond the online population through offline interactions. While Twitter was crucial, it's difficult to determine which platform was more active in which country, due to factors like government censorship and cultural preferences. Therefore, Twitter is not the sole window to understand the complex Arab Spring uprisings but its use of hashtags and access to public tweets provide valuable opportunity to look into the happiness scores.

1. **What was the impact of the Arab Spring [27] on the happiness levels of the tweets?**

   We analyzed Arabic tweets on Twitter using hedonometer.org and found that there was a continuous decline in the overall happiness score of these tweets from late 2010 until September 2013. The tweets discussed numerous sad events, protests, terrorist attacks, and deaths in different countries during those years. Although some countries experienced victories, there are still ongoing struggles in countries like Syria and Yemen. The hedonometer.org analysis also revealed a increase in happiness during events such as Eid, where tweets mentioning words like "God," "Eid," and "Holiday" dominated the conversation.

   Sad Arab Spring Events: We will present notable occurrences of low happiness scores observed on hedonometer.org during Arab Spring as depicted in Figure 3.3. Various events associated with country Egypt, particularly the violent incidents in Tahrir Square (Figure 3.4) and the trial of Hosni Mubarak

(Figure 3.5) are highlighted. Additionally, there were several events related to Libya, with particular emphasis on occurrences referred to as "Qaradawi Issues Gaddafi Fatwa" (Figure 3.6) and the "killing of Gaddafi" (Figure 3.7).

Regarding the violence in Tahrir Square (Figure 3.4), on February 2, 2011, a decline in the happiness score was observed on hedonometer.org. This decline was attributed to the use of negative words such as "wars," "bombs," "blood," and "dead," which were identified through a Twitter search using specific keywords. During that time frame, discussions about violence in Tahrir Square were taking place as thousands of protesters gathered, demanding the ouster of President Hosni Mubarak. However, an interesting observation was made regarding the word "Egypt." used in tweets (Figure 3.4). Although it was associated with negative incidents in this particular context, the word itself had a positive impact on the happiness score. This could be because users who rated the words on the Hedonometer tool may have considered the country name to have a positive connotation. This limitation of the tool highlights the challenge of interpreting emotional impact accurately.

Additionally, the analysis of other events followed a similar approach of looking at keywords and running a search on Twitter. Two events labeled "Qaradawi Issues Gaddafi Fatwa" (Figure 3.6) and "Killing of Muammar Gaddafi" (Figure 3.7) were classified as sad events in Figure 3.3 due to the significant usage of words like "Blood" and "Killing." These events suggest a potential correlation between the fatwa (a religious ruling) issued by Yusuf al-Qaradawi, which called for the killing of Muammar Gaddafi, and the subsequent events that ultimately led to Gaddafi's death.

Happy Arab Spring Events: Other than that, there were also few events noticed to have high happiness score, specifically, widely celebrated Eid-al-Adha (Figure 3.9) and Eid-al-Fitr (Figure 3.8) consistently during both years of Arab Spring despite the chaos. Words like "God," "Eid," and "Holiday" contribute to high happiness score irrespective of the order of these words because it varies throughout the years but God and word Eid remain at the top consistently when compared to other words. There were not many major happy events found during that time.

Arab Spring aftermath: We also examine the year 2013 following the end of the Arab Spring to understand the events that reflected a delay in the happiness score going back to normal and occurrences of events. We look at major sad events captured by hedonometer.org such as killing of a Somali broadcast journalist in Puntland, and the Rabaa massacre, occurred after the end of Arab Spring. It's worth noting that the mentioned events are not directly related to the Arab Spring in the sense that they were not the primary catalysts for the uprisings or directly linked to the protests themselves. However, they are events that occurred within the broader context of political instability and unrest that characterized the period. The killing of the Somali broadcast journalist in Puntland in 2013 was a separate incident that occurred in Somalia and is not directly connected to the Arab Spring, although Somalia had experienced political instability and conflict during that time. The Rabaa massacre in August 2013 refers to the violent dispersal of a sit-in protest in Cairo, Egypt. The protest was organized by supporters of ousted Egyptian President Mohamed Morsi, who was deposed following mass protests in the country. It was a significant event in the aftermath of the Arab Spring, reflecting the deep divisions and political instability that Egypt faced during that period. In Figure 3.2, it can be seen that happiness score begins to regularize after Rabaa massacre as the use of negative words such as "Killing", "Arrested" and "Prison" declines on a random day after this event as shown in Figure 3.10, but the sudden rise in happiness score reflected by Eid-al-adha regularize the happiness levels as the use of words such as "God", "Rain", "Sun" begin to increase after this Eid on a random day as shown in Figure 3.11.

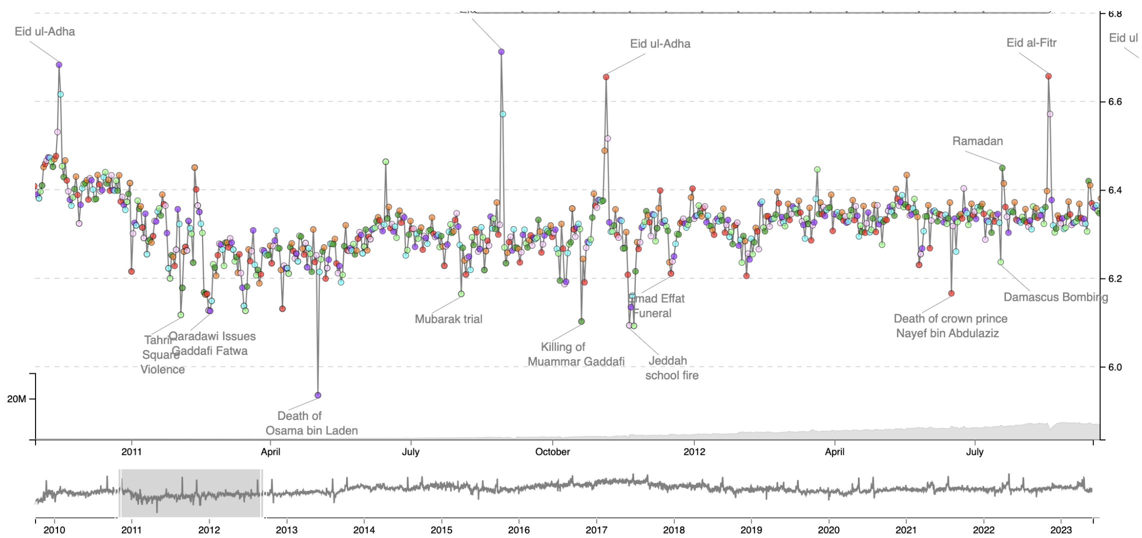


Figure 3.3: Hedonometer: The happiness score during the Arab Spring from late 2010 to late 2012

2. **What are the common characteristics of Arabic tweets that contribute to happy events?**

   To locate the happy events on the Hedonometer, we observe the high spikes from 2010 to 2023. By examining the complete hedonometer data, we can identify

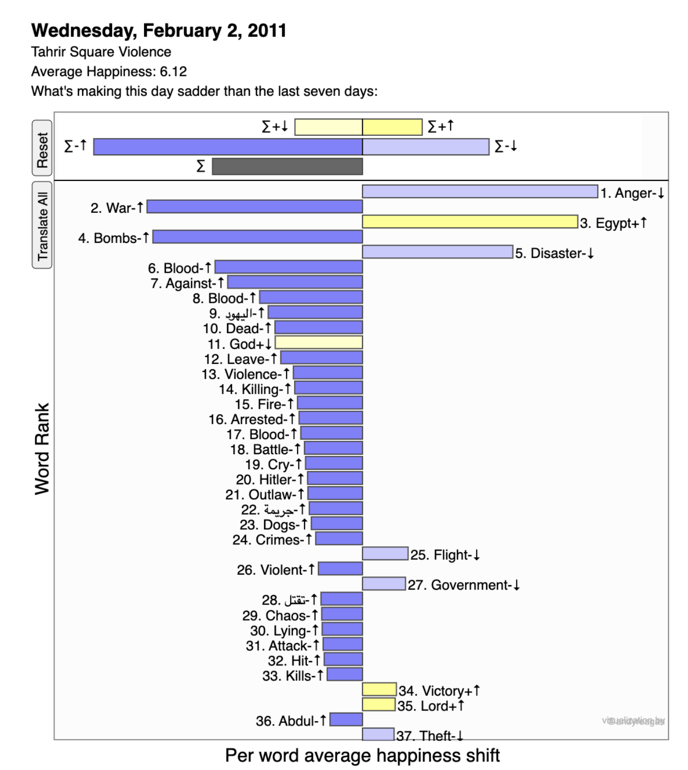


Figure 3.4: Hedonometer: Tahrir Squate Voilence

consistent occurrences labeled as Eid-al-Fitr, Eid-al-Adha, Ramadan, and a few sports events, as depicted in Figure 2.4. The words "God," "Eid," "Holiday," and "Prayer" are commonly used during both Eid celebrations, while the use of "God" at the start of Ramadan intensifies, leading to a higher happiness score. In contrast, sports events exhibit words like "Win," "Victory," and "Champion" as frequent terms. Unlike religious festivals, sports events are not consistently observed throughout the years. We also see that Friday has higher happiness score among all other days of week as shown in Figure 3.12.

The Story Wrangler tool allows us to analyze the usage of specific keywords over the years. We can examine the trends and observe how the rank of these

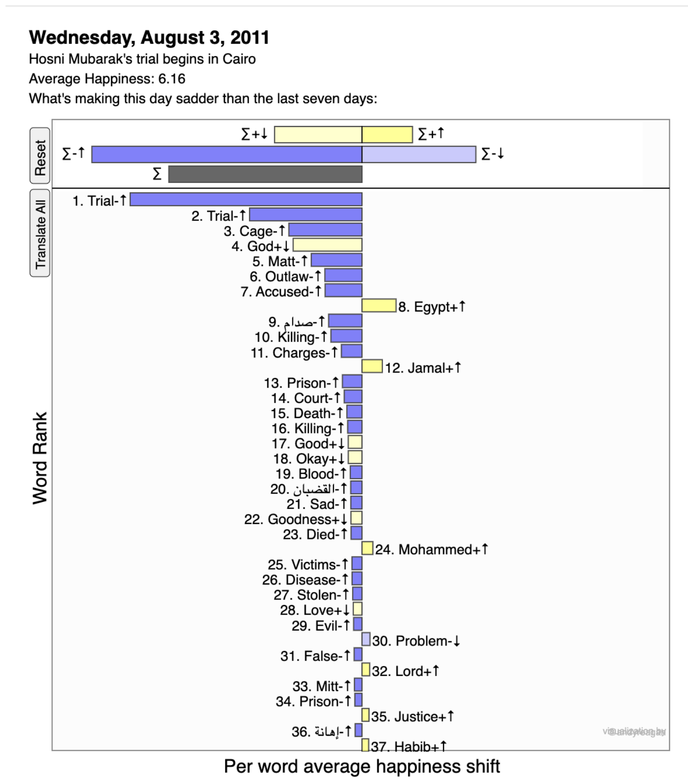


Figure 3.5: Hedonometer: Hosni Mubarak's Trial

keywords changes, indicating their usage frequency. The rank is measured on a logarithmic scale, where a high rank suggests low usage and a low rank suggests high usage of the keyword.

We being with the keyword "Eid" (عيد) and validate that the rank decreases as its usage increases every year for both Eid-al-Fitr and Eid-al-Adha consistently. Additionally, we notice that Eid-al-Adha ranks lower than Eid-al-Fitr till year 2016, indicating that it is more popular in the Arab language but then after that Eid-al-Adha ranks higher than Eid-al-Fitr as shown in Figure 3.13. There's no apparent reason behind that..

Interestingly, we also come across two other consistent Eids named Eid-al-Hub,

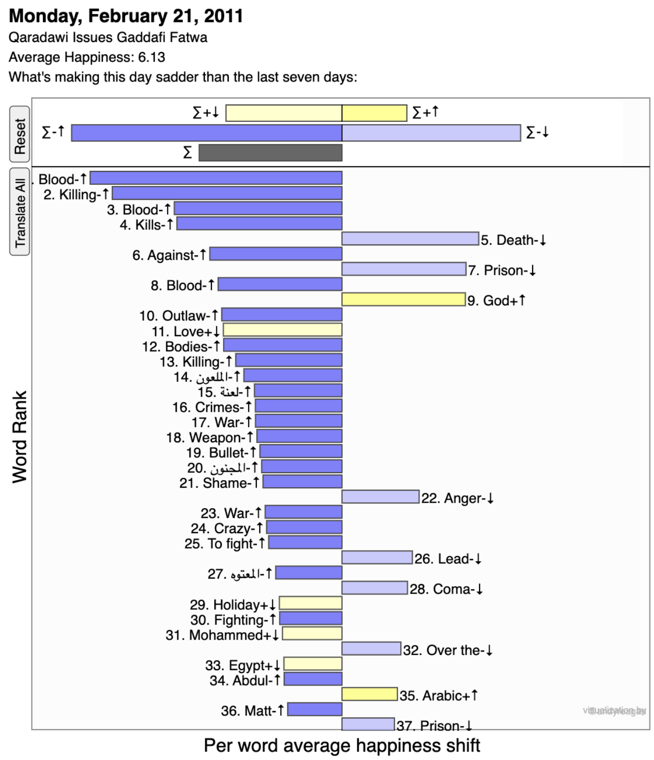


Figure 3.6: Hedonometer: Qaradawi Issues Gaddafi Fatwa

which is celebrated on February 14th and is considered equivalent to the Arabic Festival of Love, and Eid-al-Umm, celebrated on March 21st, equivalent to Mother's Day. The paper by Kreil el at., talks about how Festival of Love in Egypt and Tunisia is conceived. The authors suggest that, while the Festival of Love is celebrated in some countries, such as Egypt and Tunisia, there are arguments against it based on religious reasons. Some people explicitly oppose the Valentine's Day celebration, seeing it as an adaptation of an infidel. They believe that Islam acknowledges just two Eids: the one marking the end of Ramadan (Eid-al-Fitr) and the one remembering Abraham's sacrifice (Eid-al-Adha). Celebrating love on February 14th, a holiday that encourages prohibited

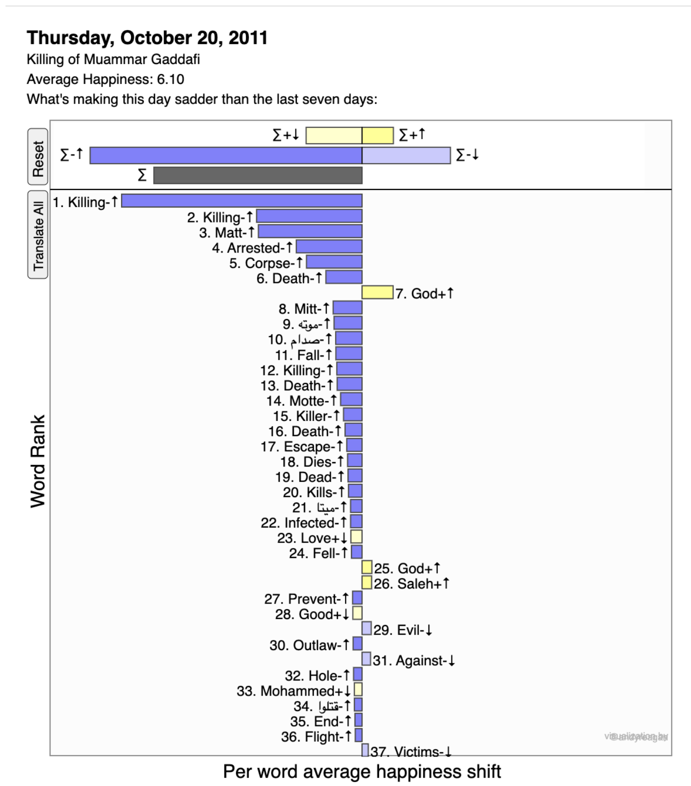


Figure 3.7: Hedonometer: Killing of Muammar Gaddafi

relationships is considered as conflicting with Islamic values. They further state that some campaigners have attempted to spread the hashtag "I am Muslim, I do not celebrate Valentine's Day." [31]. However, it's worth noting that while celebrating Eid-al-Hub is banned in Saudi Arabia, Mother's Day and Father's Day are still celebrated [32], even though only the two main Eids are supposed to be observed.

These lesser-known Eids are not highlighted on Hedonometer.org like the other Eids, but the usage of the term "Eid" in the Arabic language still increases, as shown in Figure 3.14.

We also search the word Friday (جمعة) and the rank suggest the usage increases

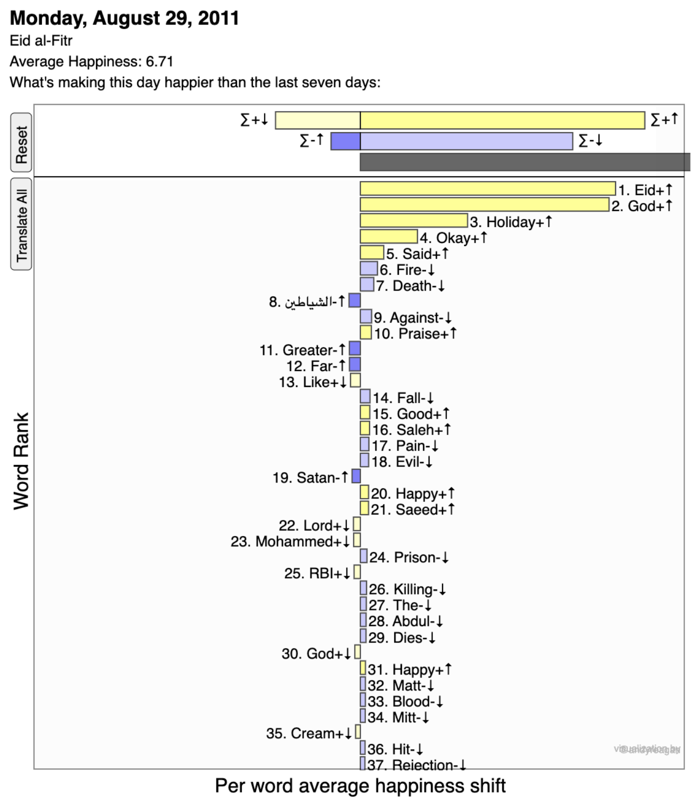


Figure 3.8: Hedonometer: Eid-al-fitr 2011

every week as shown in Figure 3.15 indicating the significance of Friday Prayer.

3. What are the common characteristics of Arabic tweets that contribute to sad events?

To locate the sad events on the Hedonometer, we observe the downward spikes from 2010 to 2023. By examining the complete hedonometer data, we can identify consistency between the sad events as they are mostly reflected by events of death and violence such as protests, terrorism and accidents. Additionally, there was a notable melancholic event associated with religion known as Ashura,

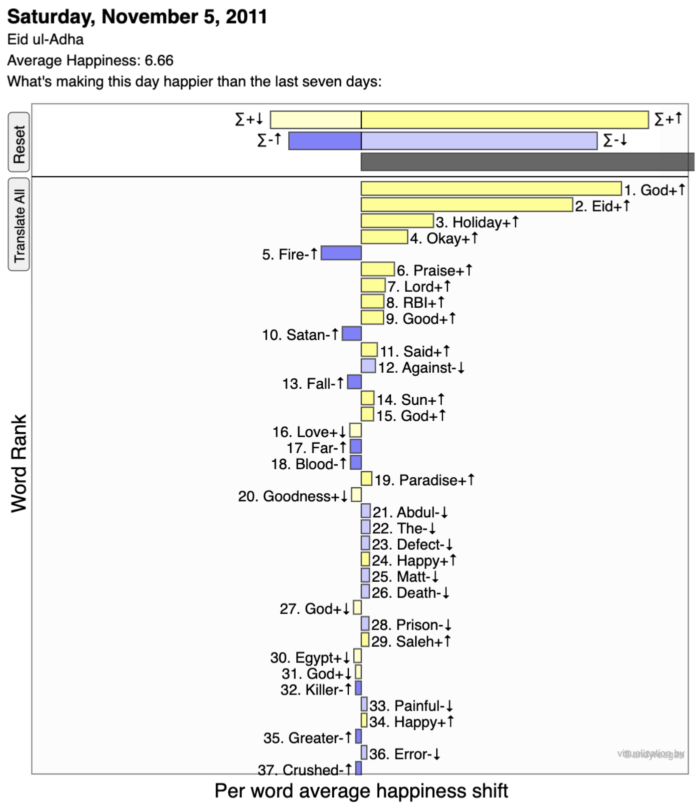


Figure 3.9: Hedonometer: Eid-al-adha 2011

which commemorates the death of Husayn ibn Ali. In cases where the death occurs naturally or through non-violent circumstances, we observe an increase in the usage of words like "pain" and "death." However, when the death is a result of a violent incident, the use of words such as "killing" and "death" becomes more prevalent. In the context of a terrorist attack, words like "killing", "bombing", "blood", "explosion", and "victim" also experience an upsurge. The Ashura event evokes the mention of words like "To die", "Evil", and "Fire".

4. **How do religious beliefs impact the happiness of the tweets**?

Religious festivals consistently generate a higher number of happy tweets, sur-

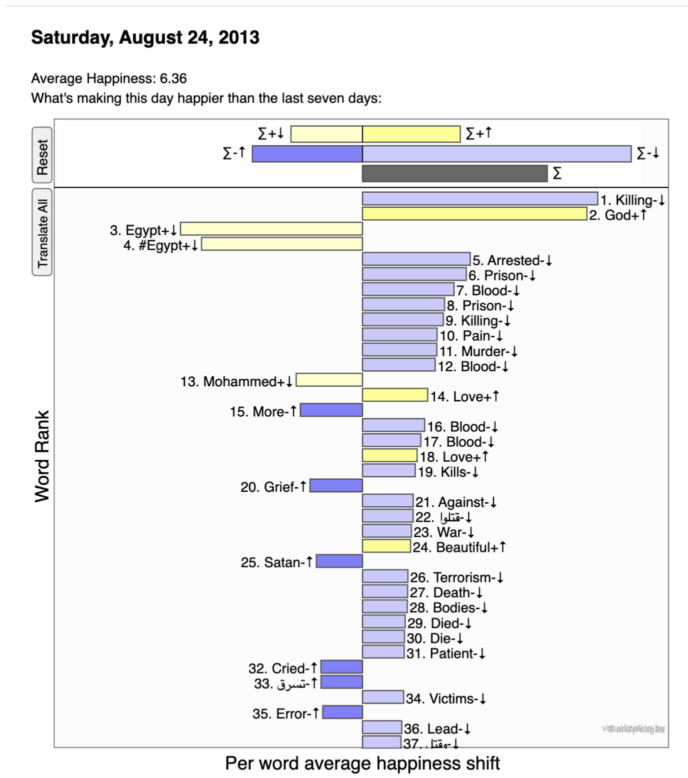


Figure 3.10: Hedonometer: Random day after Rabaa Massacre

passing other significant spikes throughout the years. However, the hedonometer also highlights several events related to shootings and bombings at mosques and churches. A notable decline in happiness score occurred on January 1st, 2011, in the city of Alexandria, Egypt. This decline was attributed to the presence of negative words such as "bombing," "accident," "victims," and "explosions," which reflected in a less happy day.

To gain deeper insights, a Twitter search was conducted during that specific time frame using these keywords. The search revealed discussions about the 2011 Alexandria bombing on church, including aspects such as the actions taken by then President Hosni Mubarak, the individuals involved in planning the

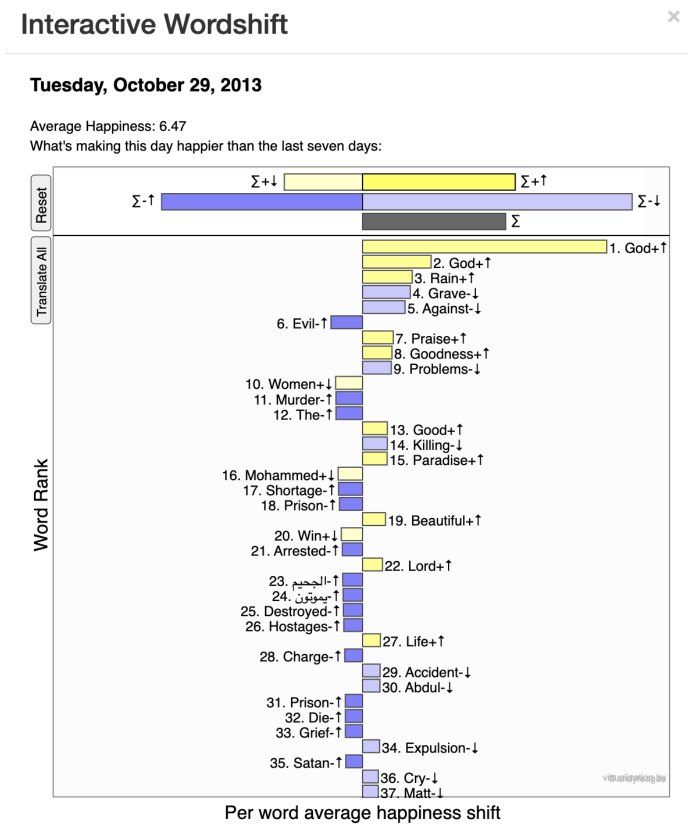


Figure 3.11: Hedonometer: Random day after Eid-al-Adha

attack, details about the funeral, and concerns raised about the safety of the Christian community due to the bombing taking place at a church. This finding suggests that the occurrence of this bombing had a negative impact on the overall happiness levels during the Arab Spring.

Another incident recorded on March 15, 2019, was labeled as the "Mass shooting at Mosques in Christchurch, New Zealand" on hedonometer.org, as depicted in Figure 2.4. This event was accompanied by an increase in the usage of words like "Terrorism," "Bombing," and "Killing," leading to a decrease in the happiness score. These incidents highlight that acts of violence associated with religion

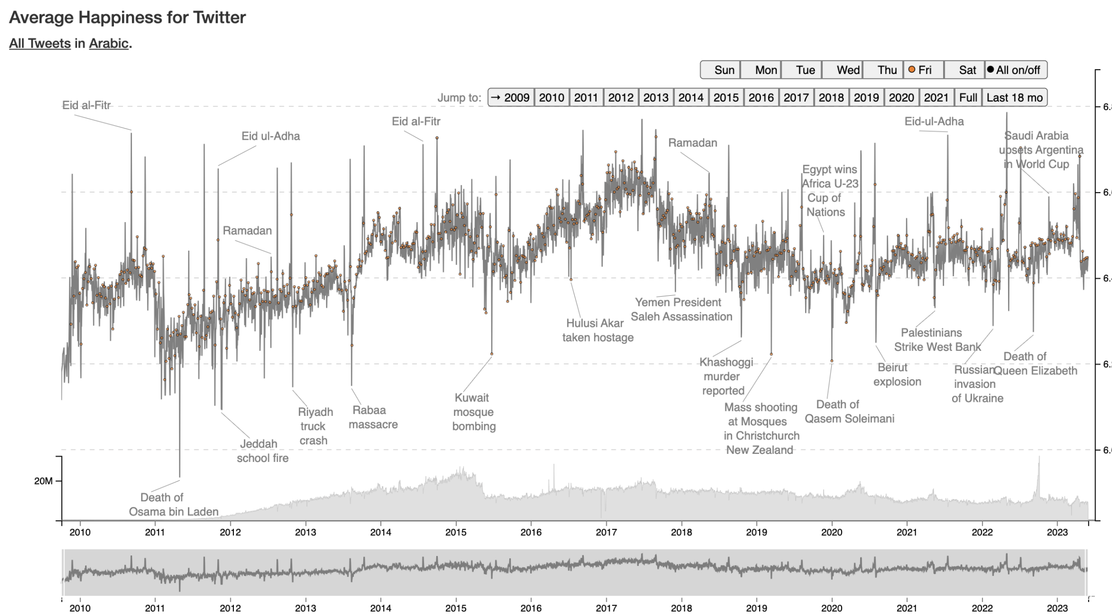


Figure 3.12: Hedonometer: Friday highlighted

predominantly contribute to lower levels of happiness, whereas religious festivals significantly boost happiness levels.

Upon investigating for "Mass shooting at Mosques in Christchurch, New Zealand" by following the similar approach, we found that it was a horrific act of terrorism carried out by a lone gunman targeting Muslim worshippers during Friday prayers. The incident took a further distressing turn when the attacker chose to livestream the shooting on social media platforms. As a result, the entire incident reflected a significant downturn in sentiment. The concerns raised in relation to this event were particularly alarming for the Muslim community, as it hinted at potential bias or prejudice based on religious affiliation. In response, individuals expressed their sentiments through tweets and messages, emphasizing that terrorism should not be associated with any specific religion.

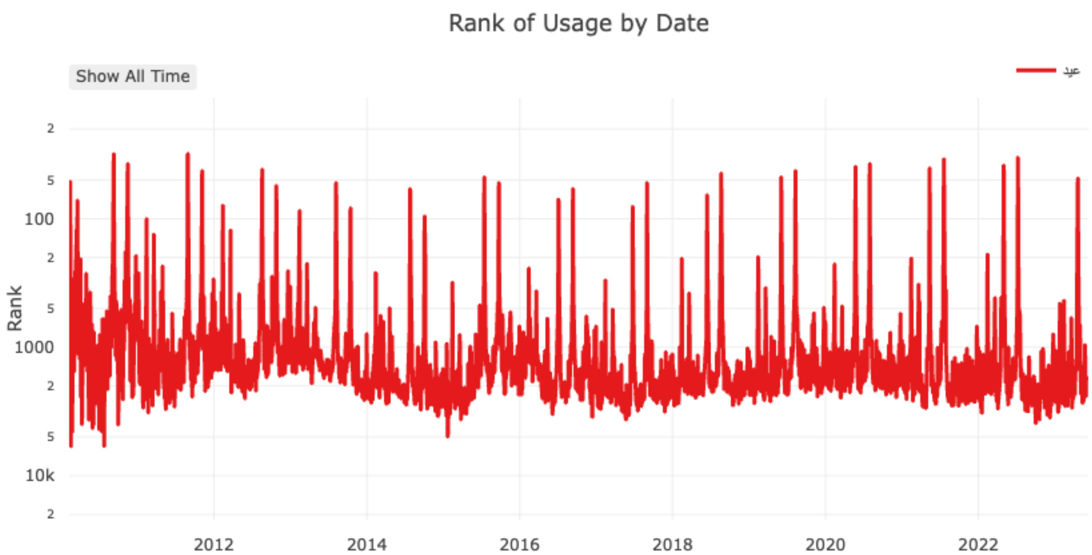
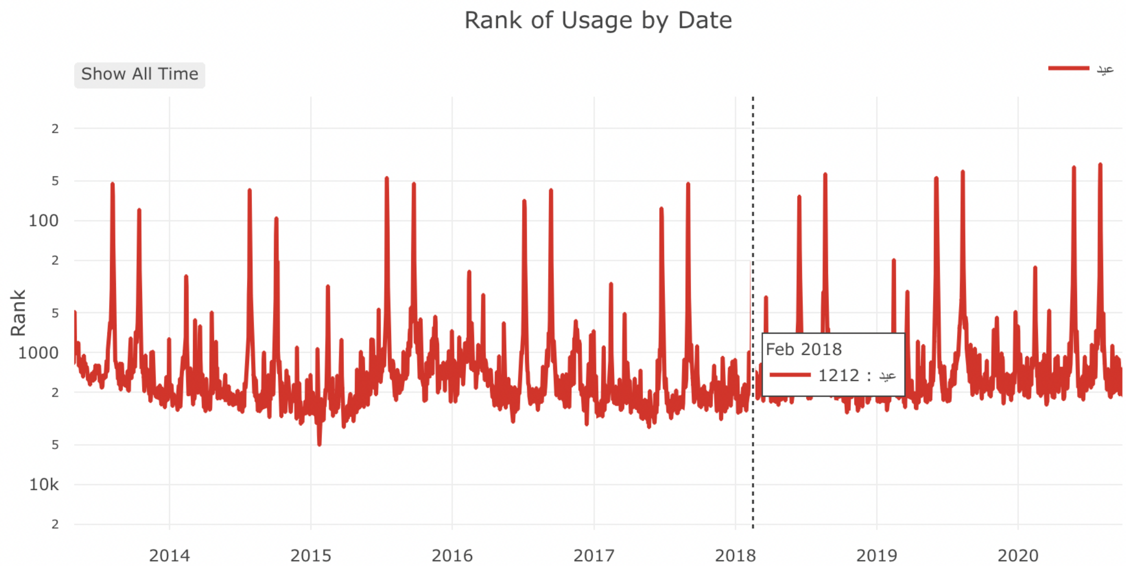


Figure 3.13: StoryWrangler: Keyword Eid highlighted

The phrase "terrorism has no religion" emerged as a common sentiment, aiming to counteract any misguided generalizations or stereotypes.



"عيد"

Figure 3.14: StoryWrangler: Arab Valentines Day highlighted

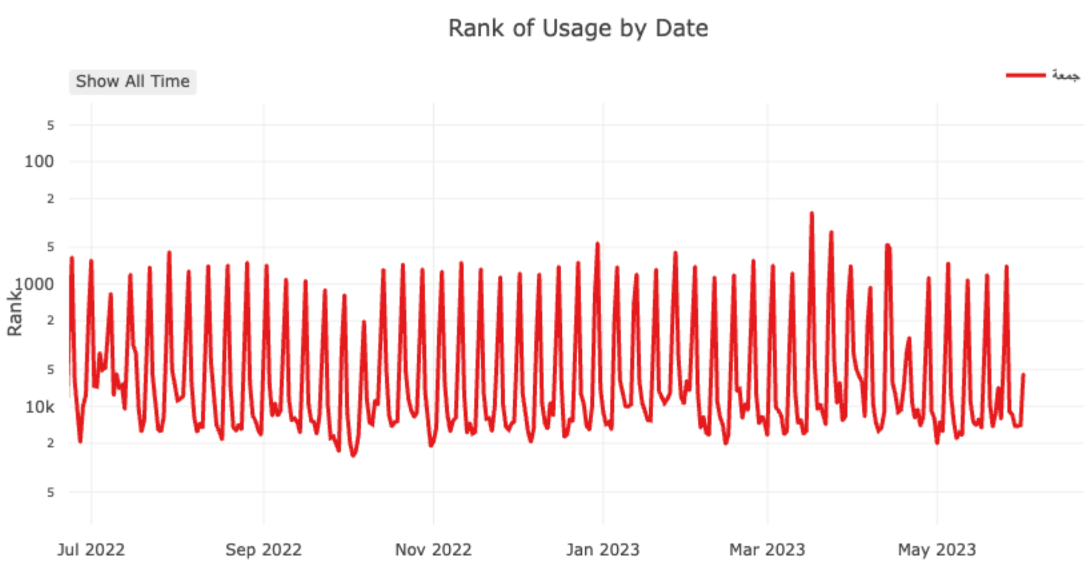


Figure 3.15: StoryWrangler: Keyword Friday highlighted

Chapter 4

Conclusion

To summarize, our study sheds light on the impact of religious and socio-political factors on happiness levels within Arabic speaking countries. We found that the onset of the Arab Spring in 2010 reflected in a noticeable decline in Twitter happiness, which persisted until around 2013, despite the unofficial conclusion of the Arab Spring in 2012.

The wave of uprisings and protests known as the "Arab Spring," which shook the Middle East in the early 2010s, can be understood as a complex web of feelings and motivations. The fact that Mohamed Bouazizi's self-immolation in Tunisia is sometimes seen as the movement's turning point, it indicates that its roots are much deeper and may go back to years of political economic injustice, corruption, and social unrest. Viewing the Arab Spring through the lens of happiness and sadness is a simplification of the myriad emotions and motivations that were at play during that period. While there were certainly elements of anger, frustration, and unhappiness, there were also feelings of hope, empowerment, and solidarity among the people seeking change. Among all those emotions, it certainly was a sad event causing the deaths of people and events for violence for several years across different countries.

While introducing a open-ended hypothesis in the conclusion, it is interesting to think about the possible effects of a translation feature on the development of the movement in the context of examining the function of social media and language during the Arab Spring. A hypothetical situation arises when we consider the impact of a translation tool, even though the addition of Arabic language support on platforms like Facebook and Twitter undoubtedly played a crucial part in facilitating communication and mobilization within Arabic speaking communities. One might hypothesize that having access to a feature like this, which enables messages to get across language barriers and find non-Arabic-speaking audiences, may have increased the movement's influence. This hypothesis raises intriguing questions about the potential spread of the Arab Spring beyond the immediate Arab region, promoting intercultural awareness and possibly gaining more support from around the world and spread of the Spring for injustice. By offering this hypothetical situation as food for thought, we can emphasize the complex interaction between language, technology, and the spread of ideas in influencing historical events and making culture, while also acknowledging the need for data to support such hypotheses.

We also noticed that Arabic culture is heavily influenced by religion, with religion often being the driving force behind happy events, while acts of violence and conflict are associated with sad events.

4.1 Limitations and Future Work

While our findings provide valuable insights, it is important to acknowledge certain limitations. The findings from hedonometer.org, which analyze Arabic tweets, cannot be generalized to all Arab countries. The happiness scores are calculated based on ratings of Arabic words, which were provided by speakers of Arabic language living in Egypt in 2012. This means that the tool may be biased towards Egyptian perspectives. To mitigate this limitation, it would be beneficial to collect ratings from individuals across different Arab countries to ensure a more diverse representation.

Once we have the ratings from different countries, we will continue this research to do comparative analysis of Arab countries to study more about the culture. To narrow the scope of this study, we would only focus on top 4 countries where the most tweets are coming in as shown in Figure 2.3. It's is challenging to determine the country of tweets as only less than 1% of tweets are geolocated. We propose a approach to get the location of arabic tweets.

Extracting countries from non-geolocated tweets: Twitter users have the freedom to write anything in the location field. Some individuals provide a location, while others do not. To extract the location information from non-geolocated tweets where location is given, we will use regular expressions to match the text in the location field with the names of the mentioned countries. An expression that defines a search pattern is known as a regular expression. It is used to look for patterns in text, such as specific words or character sequences [33]. For example, to search for Kuwait in the location field, we used a regular expression pattern that searches for specific words and phrases related to Kuwait, such as "Kuwait", "KWT", "الكويت",

"Jahra", "الجهراء", "Qortuba", "KW", and "K U W A I T".

This method used to extract the location of non-geolocated tweets may not be foolproof as it relies on the location information provided by the user in their profile, which may not necessarily reflect their actual location. However, given that the tweets analyzed are in Arabic, there is a high likelihood that the user posting the tweet resides in one of these Arab countries. Therefore, while the location information obtained through this method may not be completely accurate, it can still provide valuable insights into the geographical distribution of the tweets and help facilitate comparative analysis between different Arab countries.

To further enhance our understanding of Arabic sentiment, it is recommended to incorporate alternative approaches, such as utilizing transformer models, to capture the intricate nuances of language [34]. These advanced techniques can provide deeper insights into the sentiment expressed in Arabic tweets, surpassing the limitations of relying solely on word ratings. For example, in this research, we observed that the word ‘mubarak’ consistently received a high score, primarily due to its positive connotations in Arabic, such as ‘congratulations’ or ‘blessed.’ However, its association with the former president, Hosni ‘Mubarak’, influenced the overall happiness score. Although the measures were taken to neutralize this impact and prevent it from skewing the results on Hedonometer, but this measure would also impact the tweets where ‘mubarak’ actually plays means ‘blessing’ and plays an important role. In this case, capturing the semantic meaning of the words and their relationships with other words in the tweet becomes crucial that could be achieved by word embedding in transformers.

By addressing these limitations and utilizing more sophisticated methodologies, future studies can offer a more comprehensive understanding of the factors influencing happiness levels in Arabic-speaking countries. This knowledge is crucial for promoting well-being and improving the lives of individuals within these vibrant communities.

Bibliography

[1] Daniel L. Everett. *Language: The Cultural Tool.* Vintage, 2012.

[2] Thayer Alshaabi, David R Dewhurst, Joshua R Minot, Michael V Arnold, James L Adams, Christopher M Danforth, and Peter Sheridan Dodds. The growing amplification of social media: Measuring temporal and social contagion dynamics for over 150 languages on twitter for 2009–2020. *EPJ Data Science*, 10(1):15, 2021.

[3] Ali Al Shehhi, Julie Thomas, Roy Welsch, Ian Grey, and Zeyar Aung. Arabia felix 2.0: a cross-linguistic twitter analysis of happiness patterns in the united arab emirates. *Journal of Big Data*, 6(1):1–20, 2019.

[4] M. H. Bakalla. *Arabic culture: Through its language and literature.* Routledge, 1984.

[5] A. Farghaly and K. Shaalan. Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 8(4):1–22, 2009.

[6] M. A. Attia. *Handling Arabic morphological and syntactic ambiguity within the LFG framework with a view to machine translation.* PhD thesis, The University of Manchester (United Kingdom), 2008.

[7] Alan Ritter, Sam Clark, Mausam, and Oren Etzioni. Named entity recognition in tweets: An experimental study. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1524–1534, 2011.

[8] Alan Ritter, Mausam, Oren Etzioni, and Sam Clark. Open domain event extraction from twitter. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*, pages 1104–1112. ACM, 2012.

[9] Julia Hirschberg and Christopher D. Manning. Advances in natural language processing. *Science*, 349(6245):261–266, 2015.

[10] Peter Sheridan Dodds, Kameron Decker Harris, Isabel M Kloumann, Catherine A Bliss, and Christopher M Danforth. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PLoS ONE*, 6(12):e26752, 2011.

[11] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing*, 9(6):811–824, 2012.

[12] V. Kharde and S. Sonawane. Sentiment analysis of twitter data: A survey of techniques. *International Journal of Computer Applications*, 139(11):5–15, 2016.

[13] Y. Kryvasheyeu, H. Chen, N. Obradovich, E. Moro, P. Van Hentenryck, J. Fowler, and M. Cebrian. Rapid assessment of disaster damage using social media activity. *Science Advances*, 2(3):e1500779, 2016.

[14] J. A. Russell and A. Mehrabian. Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294, 1977.

[15] W. T. Jones. *The Classical Mind*. Harcourt, Brace, Jovanovich, New York, 1970.

[16] P. S. Dodds and C. M. Danforth. Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of happiness studies*, 11:441–456, 2010.

[17] B. Russell. *A History of Western Philosophy*. Allen & Unwin, London, 1961.

[18] M. Argyle. *The Psychology of Happiness*. Routledge, New York, 2 edition, 2001.

[19] R. Layard. *Happiness*. The Penguin Press, London, 2005.

[20] D. Gilbert. *Stumbling on Happiness*. Knopf, New York, 2006.

[21] C. R. Snyder and S. J. Lopez. *Positive Psychology*. Oxford University Press, New York, NY, 2 edition, 2009.

[22] C. E. Osgood, G. J. Suci, and P. H. Tannenbaum. *The measurement of meaning*. University of Illinois Press, 1957.

[23] Peter Sheridan Dodds, Eric M Clark, Suma Desu, Morgan R Frank, Andrew J Reagan, Jake R Williams, and Christopher M Danforth. Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, 112(8):2389–2394, 2015.

[24] Eric M. Clark, Ted James, Chris A. Jones, Amulya Alapati, Promise Ukandu, Christopher M. Danforth, and Peter Sheridan Dodds. A sentiment analysis of breast cancer treatment experiences and healthcare perceptions across twitter, 2018.

[25] Sharon E. Alajajian, Jake Ryland Williams, Andrew J. Reagan, Stephen C. Alajajian, Morgan R. Frank, Lewis Mitchell, Jacob Lahne, Christopher M. Danforth, and Peter Sheridan Dodds. The lexicocalorimeter: Gauging public health through caloric input and output on social media. *PLOS ONE*, 12(2):1–25, 02 2017.

[26] Henry H. Wu, Ryan J. Gallagher, Thayer Alshaabi, Jane L. Adams, Joshua R. Minot, Michael V. Arnold, Brooke Foucault Welles, Randall Harp, Peter Sheridan Dodds, and Christopher M. Danforth. Say their names: Resurgence in the collective attention toward black victims of fatal police violence following the death of george floyd. *PLOS ONE*, 18(1):1–26, 01 2023.

[27] Zeynep Tufekci. *Twitter and Tear Gas: The Power and Fragility of Networked Protest*. Yale University Press, New Haven, 2017.

[28] Sune Haugbolle. Local solidarities: How the arab spring protests started. *Sociology Compass*, 10(2):197–207, 2016.

[29] Joseph Pugliese. Permanent revolution: Mohamed bouazizi's incendiary ethics of revolt. *Law, Culture and the Humanities*, 10(3):408–420, 2014.

[30] Jason Goodwin. Why we were surprised (again) by the arab spring. *Swiss Political Science Review*, 17(4):452–456, 2011.

[31] Aymon Kreil and Rihab Ben Othmen. Valentines day: An array of egyptian and tunisian lifeworlds in 2016. *Journal of Arabic and Islamic Studies*, 21(3), 2021.

[32] Barbara Michalak-Pikulska. Girls from riyadh by rajaa alsanea as a literary statement of changes within contemporary saudi society. *Rocznik Orientalistyczny/Yearbook of Oriental Studies*, 2013.

[33] Jing Liu, Yufan Wu, Hu Qin, and Zhixiong Li. Recurrent neural network based maritime anomaly detection. In *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, pages 219–228. Springer, 2019.

[34] Usman Naseem, Imran Razzak, Katarzyna Musial, and Muhammad Imran. Transformer based deep intelligent contextual embedding for twitter sentiment analysis. *Future Generation Computer Systems*, 113:58–69, 2020.