

## Introduction

Twitter, a social networking website launched in 2006, has become an effective micro-blogging platform. With more than 336 million monthly active users worldwide, it generates around 500 million tweets per day and 200 billion tweets per year. The content of the tweets covers all aspects of life; social, political and economic. Since Twitter data is easy to acquire and contains large amounts of information regarding different industries, people, and businesses, it provides a perfect platform to access real time data which can be analyzed to provide answers to real life problems, detect patterns and determine people's opinions and feelings regarding certain matters.

Text analysis is an effective tool that reveals how positive or negative the content of a particular tweet is, and can be extremely advantageous, especially in the business environment. For example, it can assist businesses in rethinking their marketing strategy by focusing on promotional strategies that cater to consumers' wants. Moreover, they can also pick up on negative conversations to improve the quality of their customer service and reacting promptly to customer complaints. While text analysis may be particularly useful in a business environment, it can also provide interesting insights regarding controversial topics trending worldwide.

Our analysis will focus mostly on people's opinions regarding political issues and figures within Pakistan, including but not limited to Imran Khan, Manzoor Pashteen etc. We expect our data to reveal interesting results regarding people's feelings and attitudes towards controversial figures and matters, and whether these results are consistent with our preexisting notions and opinions. The topics and hashtags we plan on exploring are as follows:

- #OurPeacefulBalochistan
- #ShameOnFIA
- #WeStandWithFIA
- #Pashteen
- #ImranKhan
- #AntiCorruptionDay

## Methodology

In order to extract data from Twitter to study it to gauge trends and patterns, firstly a Twitter account needs to be created at <https://apps.twitter.com/> to access developer options. A new application is created with a declaration of the purpose for data mining. Once the application credentials are received, a Twitter API (Application Programming Interface) has been set up.

We used the twitterR library in R studio to connect to Twitter Streaming API and download the data. The Twitter API comes with certain constraints. Rate limits prevent the download of more than xxx tweets per window durations of 15 minutes each. For our study, this poses a restriction on the amount of data we can reasonably collect. Furthermore, the data extracted comprises of only 16 columns of basic information regarding the tweets extracted i.e. the

tweet text, the username which posted the tweet, the number of favorites, retweets and replies etc.

Once this framework was created, we extracted tweets using specific keywords in the form of hashtags. Hashtags are words or phrases that uniquely identify trending topics on Twitter. All the hashtags selected for this study were political and current in nature, and were as follows:

#### **#WeStandWithFIA:**

This trend emerged when two members of the Pakistan National Assembly were taken in custody by the Federal Investigation Agency (FIA) as they were about to leave the country for UAE. They were arrested because of their association with the Pashtun Tahaffuz Movement (PTM), a social movement for Pashtun human rights, and their recent “provocative” speeches in its favor. Twitter users expressed their support for the arrest with this hashtag.

#### **#ShameOnFIA:**

Another hashtag emerged in opposition of this action from the politically liberal side of Twitter who expressed support for PTM and the MNAs affected. These Twitter users condemned FIA for the step taken.

#### **#Pashteen:**

Manzoor Pashteen is the chairman of the Pashtun Tahaffuz Movement, therefore directly relevant to the issue being explored.

#### **#OurPeacefulBalochistan:**

This hashtag has emerged to highlight the peaceful and positive aspects of the Pakistani province, Balochistan. This hashtag appears with tweets of two kinds: those actually using the hashtag for its intended purpose, and those that are pointing out the falsehood being depicted by it. This gives us reason to believe that this hashtag may not have emerged organically, and rather was engineered to gear public sentiment in a particular direction.

#### **#ImranKhan:**

This hashtag depicts public sentiment about the current Prime Minister of Pakistan, Imran Khan.

#### **#AntiCorruptionDay:**

International Anti-Corruption Day is celebrated annually on the 9th of December to raise awareness against the threat of corruption.

The following techniques will be employed for conducting our analysis on certain political issues and controversial figures:

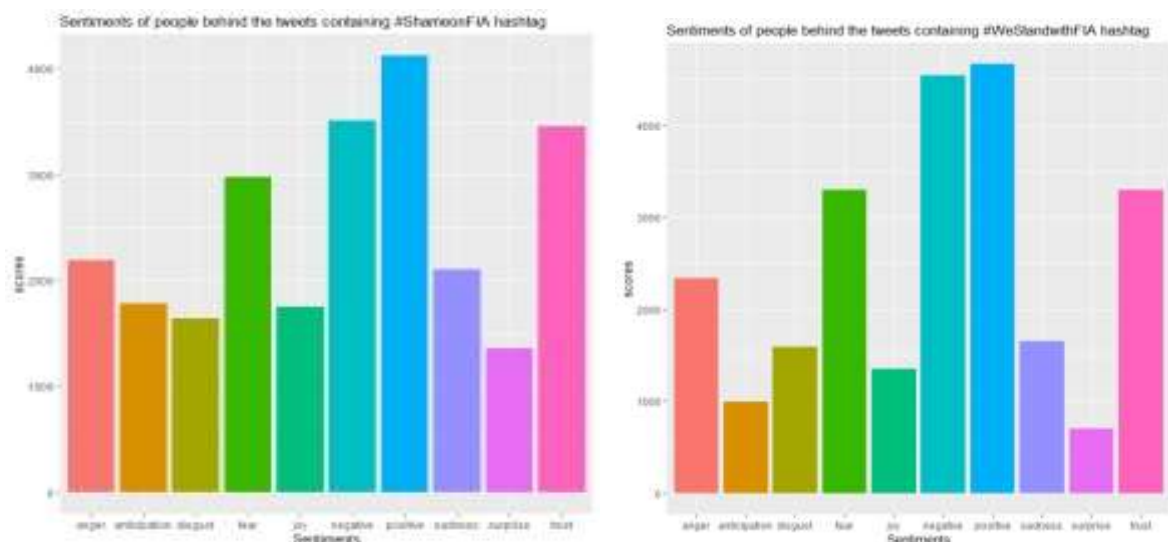
- **Sentiment Analysis:** To determine the general attitude of the tweet regarding something i.e. whether the attitude is positive, negative or neutral
- **Network of words:** To discover words associated with one another; for example association between “love” and “Imran Khan” would indicate generally positive sentiments towards the politician
- **Timestamp Analysis:** Analyzing the time and date of a tweet to determine whether it was actually the writer’s opinion or whether some manipulation was involved

## Sentiment and Network of Words Analysis

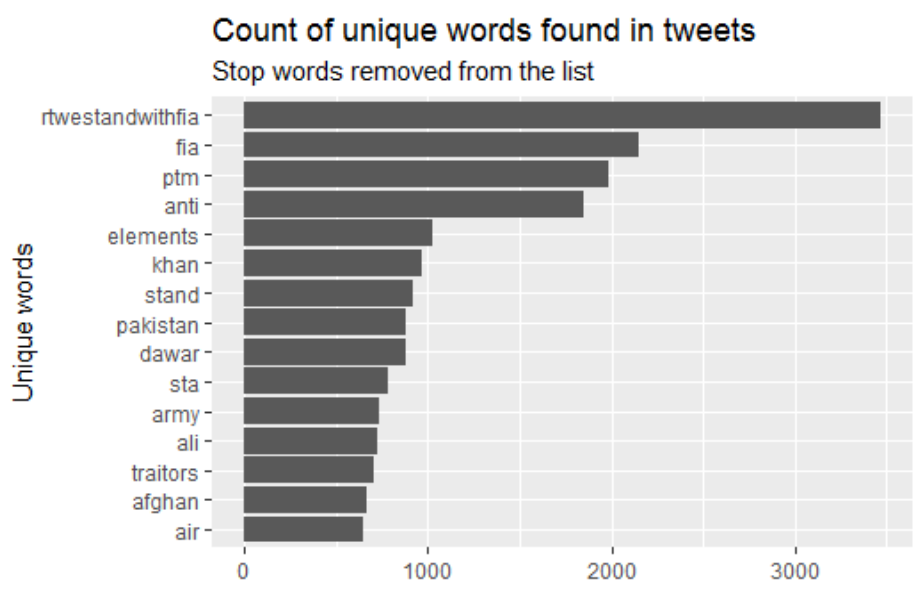
### 1. #ShameonFIA and #WeStandwithFIA:

Two trends have been competing on twitter—#ShameonFIA and #WeStandwithFIA. It all started when Mohsin Dawar and Ali Wazir, MNAs associated with PTM were taken into custody at the Bacha Khan International Airport by FIA. Liberals started speaking out against FIA thus the #ShameonFIA hashtag began trending on twitter. In order to compete with this, #WeStandwithFIA hashtag also emerged.

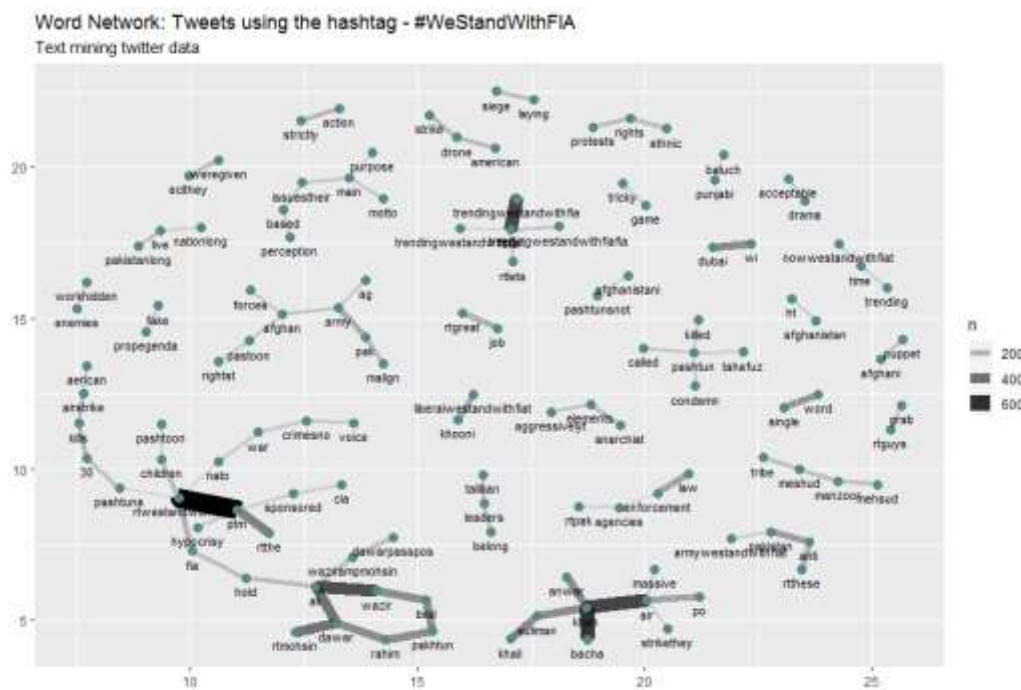
It is likely that the second hashtag was an organized trending campaign to downplay the consequences of the arrest and gear public opinion in favor of the federal body by portraying the action in a positive light and criticizing the efforts of PTM as anti-state, and therefore, did not reflect the actual opinions or viewpoints of the authors of the tweets. We can see that both the sentiment graphs are very similar, and high and low on the same emotions. This is because both hashtags are competing with one another. Where #WeStandWithFIA has expressed favor for FIA and criticism for PTM, #ShameOnFIA has done the opposite.



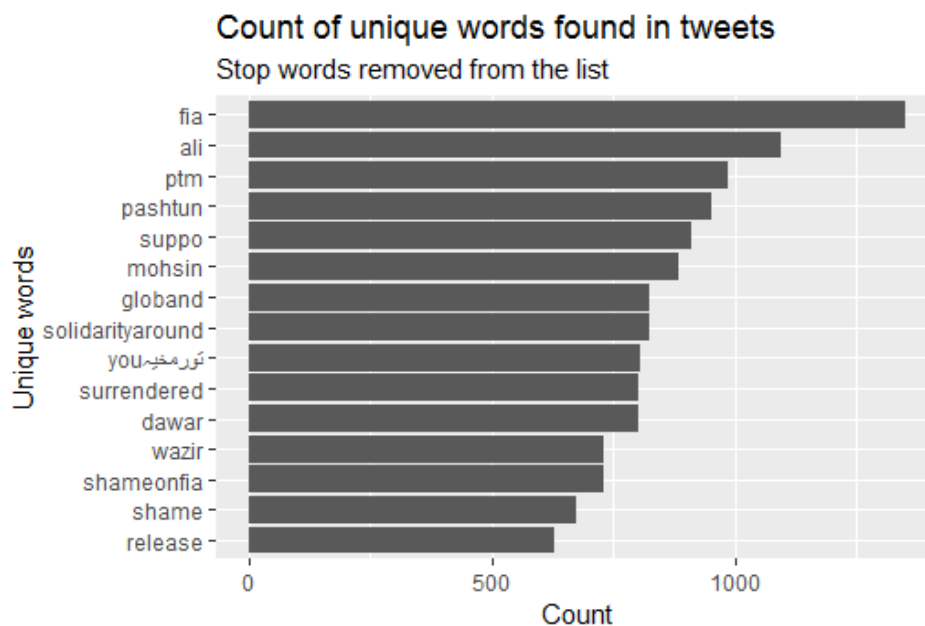
The network of words analysis suggests the same. When looking at the frequency of unique words for #WeStandWithFIA, negative words like “traitor” and what can be assumed to be “anti-state and anti-army elements” appear with Ali Wazir and Mohsin Dawar’s names:



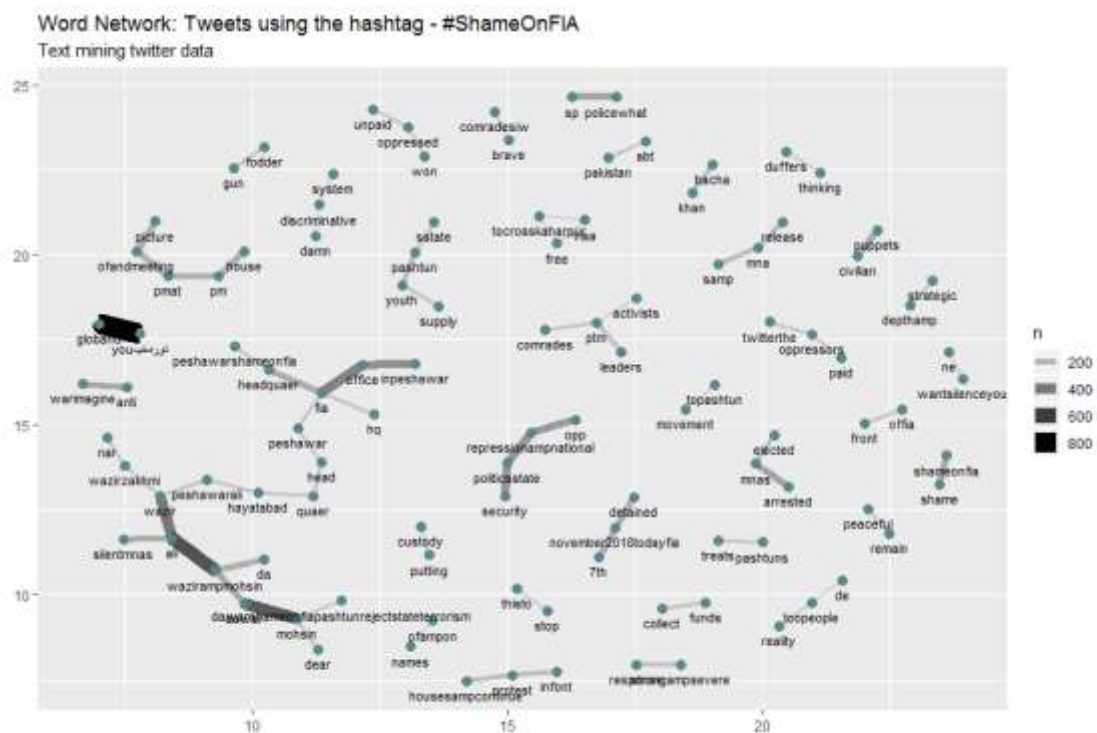
The network of words calls attention to pairings of words like “fake propaganda”, “aggressive anarchist elements”, “CIA sponsored” that represent public outlook towards PTM. The bottom left corner of the network builds a whole narrative. An *American airstrike killed Pashtuns* in Afghanistan but the two PTM members did not speak out against this which highlighted *PTM’s hypocrisy* to the people. Several string associations seen are between keywords regarding the issue (e.g. trendingwestandwithFIA, PTM, Bacha Khan, which is the airport where the arrest took place, etc.)



#ShameOnFIA has the following word frequencies:



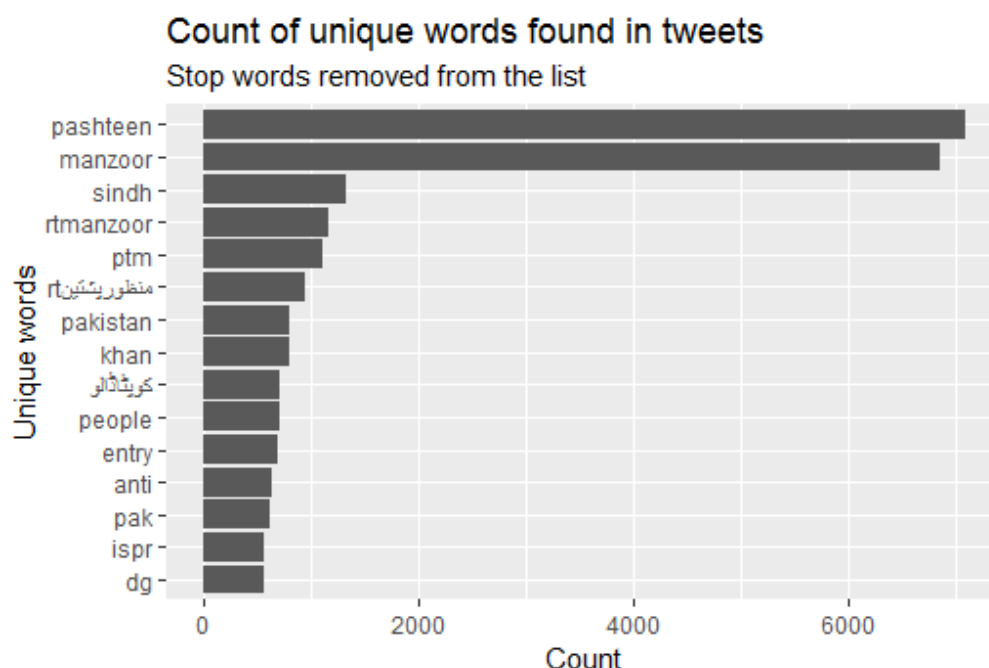
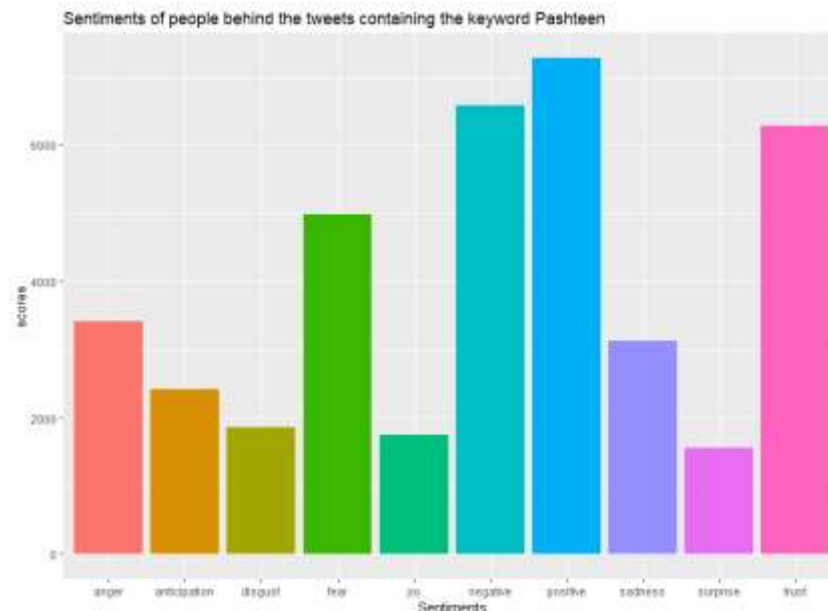
This suggests phrases like “*support for and solidarity around PTM/Pashtuns*”, asking for “*the release of Ali and Mohsin*”, and “*Shame on FIA*”. The network again builds a narrative telling us the *system is discriminative and oppressive*, referring to the *PTM leaders and activists as brave comrades*. The mix of sentiments is evident here.



## 2. Pashteen:

Manzoor Pashteen is a human rights activist and chairman of PTM who was recently denied entry into Quetta and sent back to Karachi, both cities of Pakistan. People were enraged since he was on his way to participate in a protest against “enforced disappearances and extrajudicial killings in Quetta.” This is why the sentiment analysis

shows that the tweets are high on negativity. However, with the recent order that withdrew the ban on Pashteen from Quetta, the graph indicates that trust has been restored. People are also exhibiting positive emotions regarding this decision, commending those involved in lifting the ban.

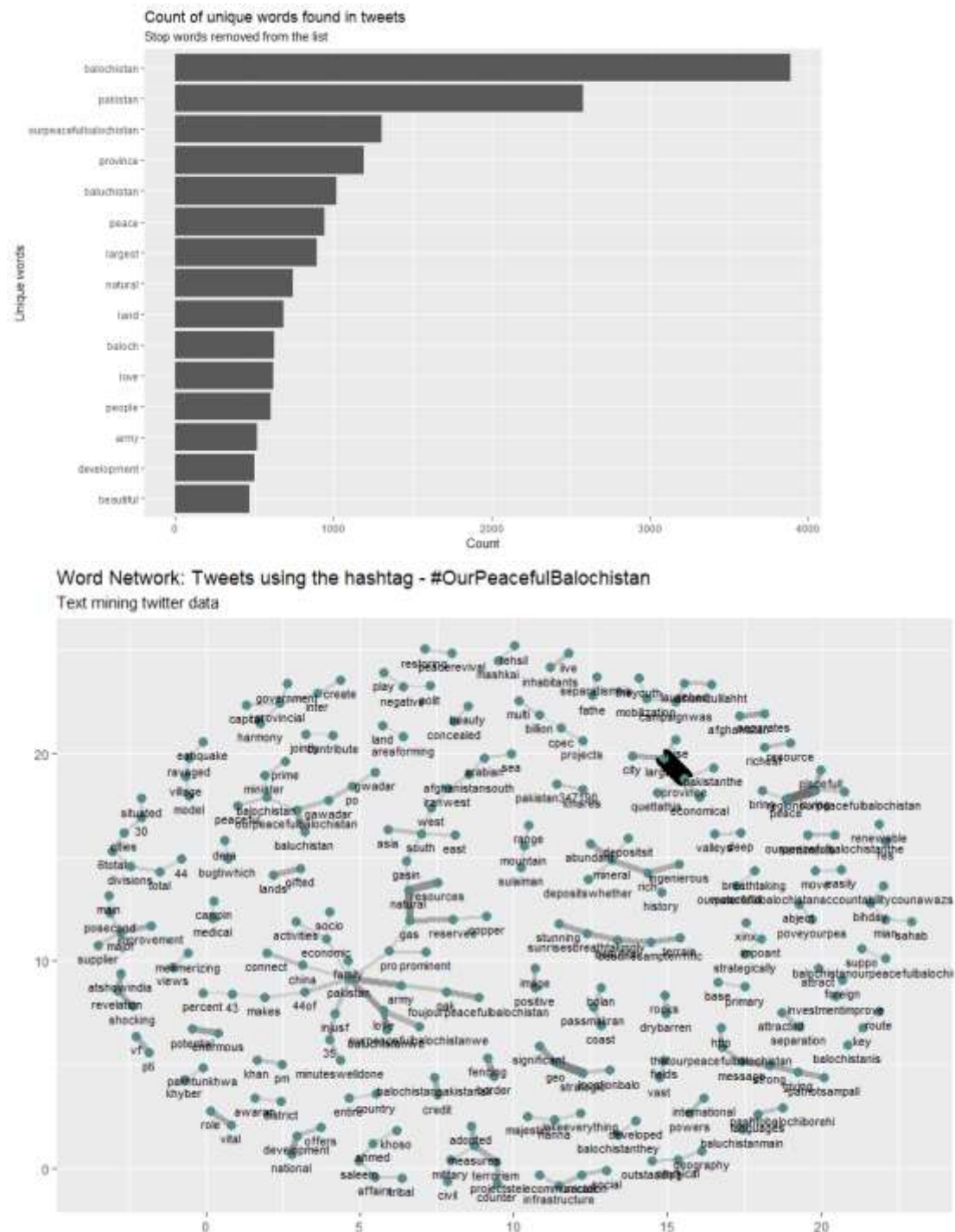


The network points to ideas of “*fundamental human rights*” which Manzoor Pashteen fights for and the army “*banning peaceful political activities*”. There is high mention of Naeqebullah Mehsud (*Naeqeb Mesud*), a 27-year-old man from Waziristan, who was killed in a police encounter for charges of participation in terrorist activity and affiliation with the terrorist group Tehrik-e-Taliban. PTM gained prominence in 2018 when it called for the setup of a judicial inquiry into his killing. There is also a demand to the DG ISPR to *withdraw notification* of Pashteen’s ban from Quetta.





The main topics of interest are also highlighted in the frequency of words and network of words e.g. *natural resources*, *beautiful land*, *development*, *socio economic activities*, *enormous potential* etc.:



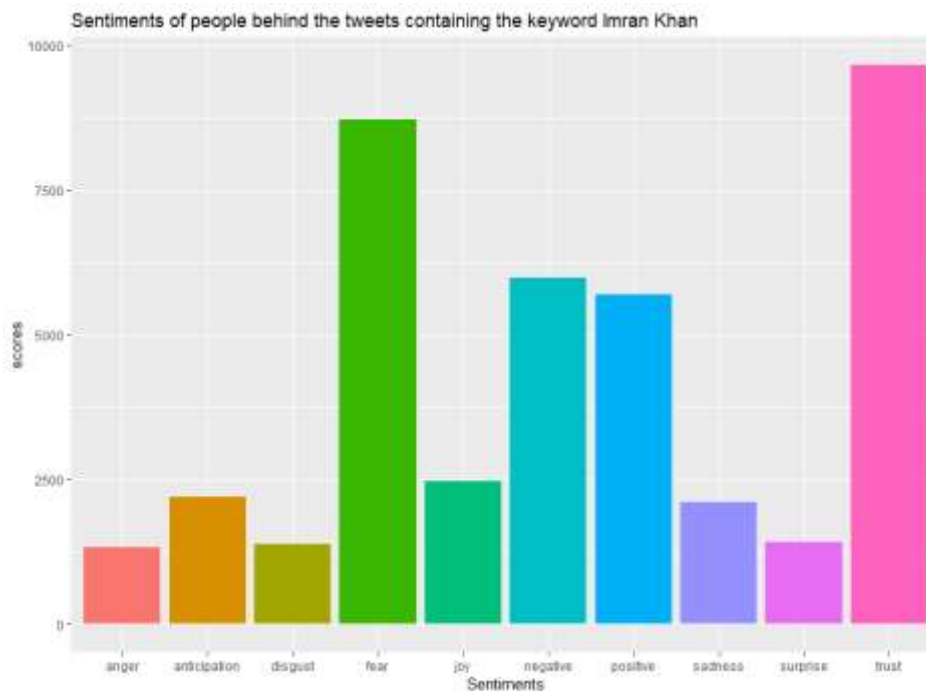
However, the current political context is very different. Balochistan, despite indeed being a beautiful land with an abundance of natural resources, have too often been given the short end of the stick. Other provinces capitalize on its resources while it remains underdeveloped and inaccessible. Moreover, it is riddled in violence and disorder. To see such a positive trend in the current scenario hints at manipulation. There is a possibility that organized tweet campaigns were done (meaning that they do not actually express the



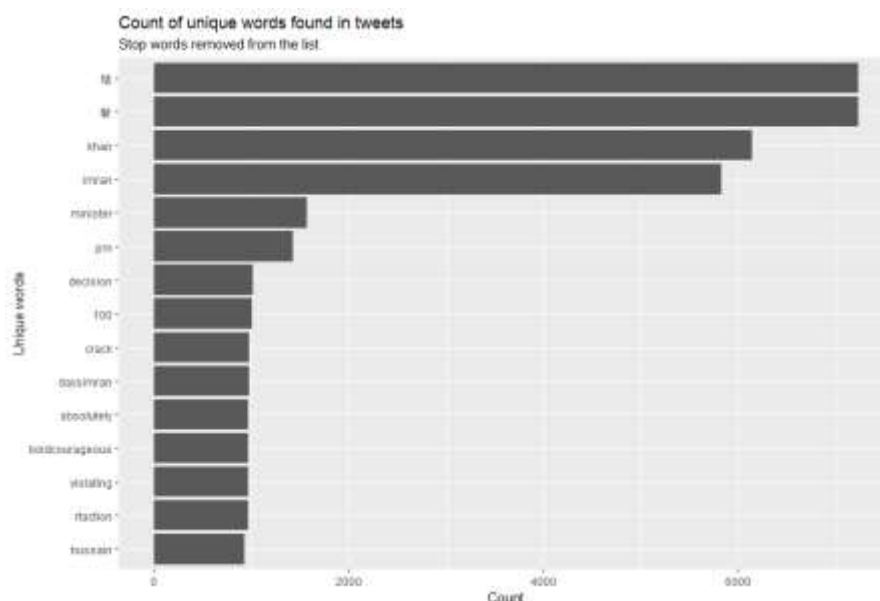
writers' own opinion) since the “positive” emotions seems to be way too high considering the actual sentiments of the Baloch people.

#### 4. Imran Khan:

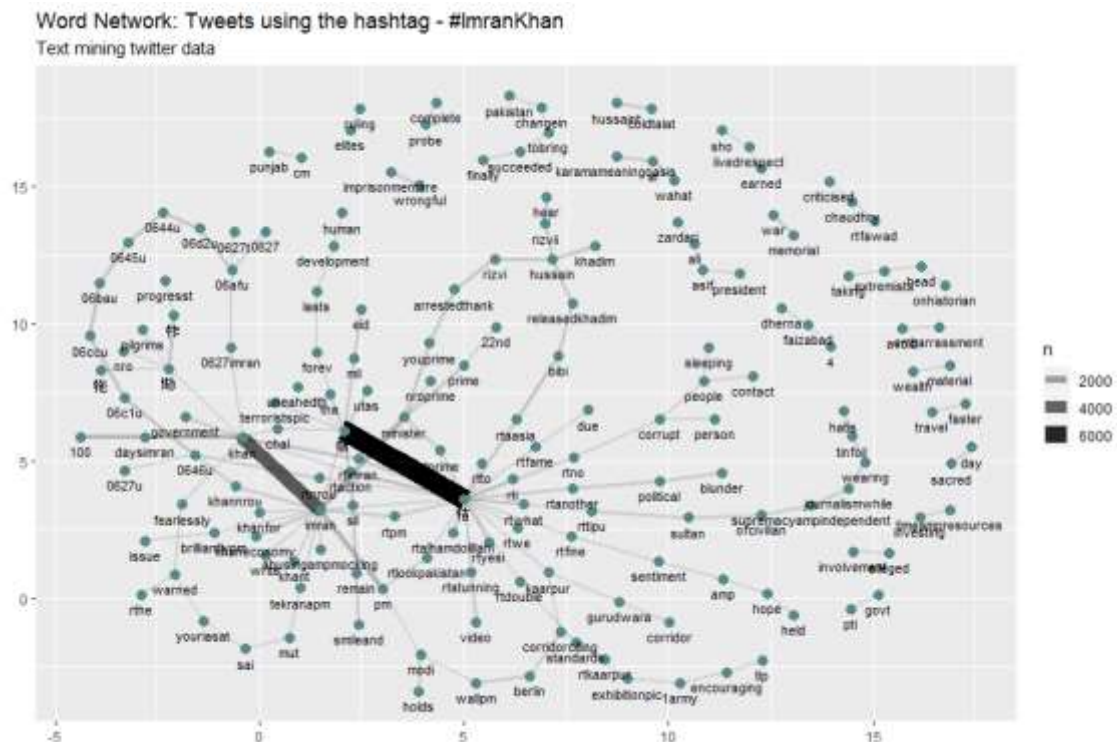
Imran Khan is a fairly controversial politician, now Prime Minister of Pakistan. People hold extreme views regarding his agendas and ability to bring Pakistan on the right track. The sentiment analysis shows high levels of “trust” and “fear” as well as high levels of “positive” and “negative” sentiments. These extremely conflicting emotions (i.e. fear and trust, and positive and negative) just emphasizes the point that people do indeed have very extreme, possible preset views regarding him.



This is also very evident in the frequency of words chart that also depicts positive extremes like *bold* and *courageous* as well as negative extremes like *violation* of his promises.

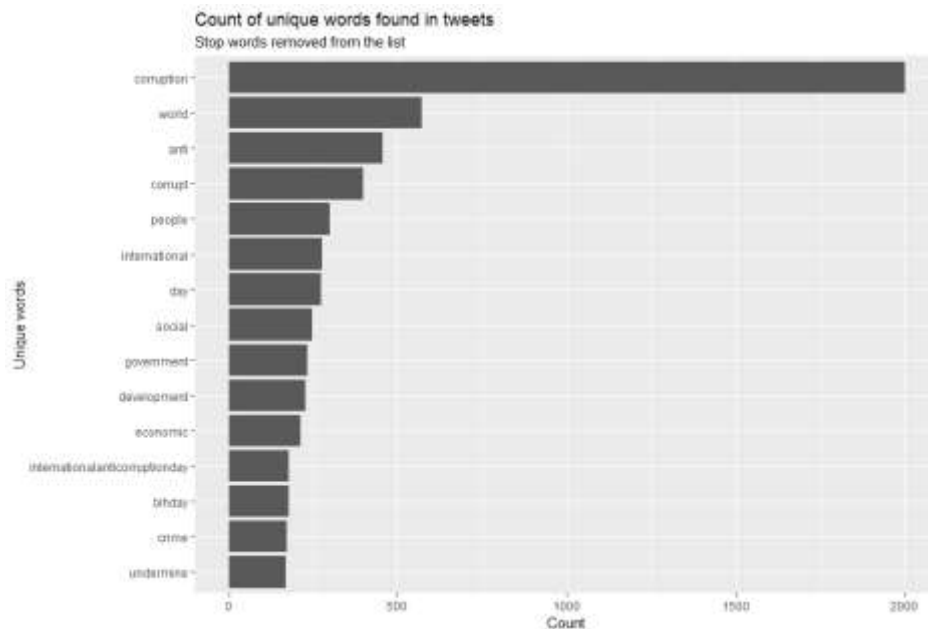
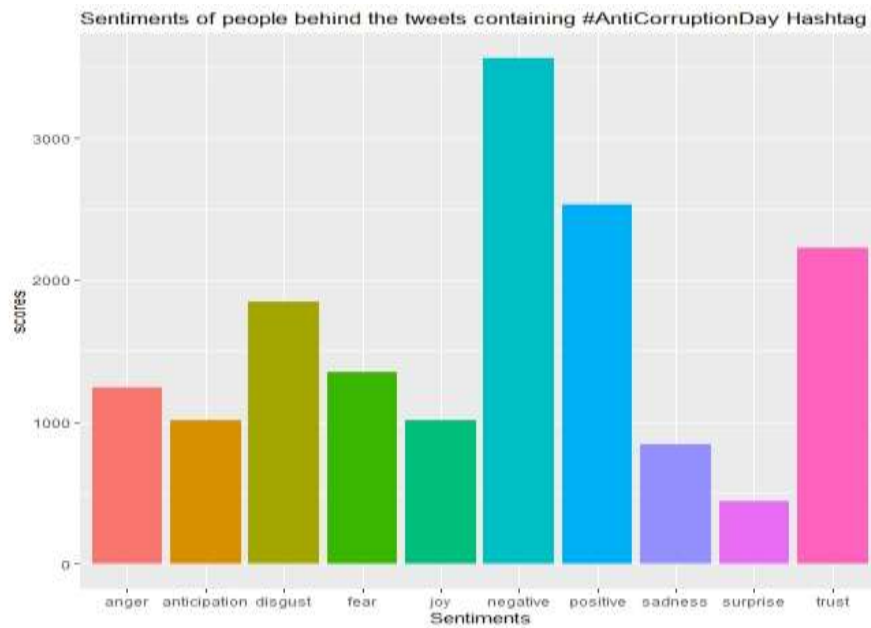


The word network itself speaks of change in Pakistan, lasting human development, appreciation of his action against the provocative political figure, Khadim Rizvi. However, criticism follows with mention of Khan's political blunders, wrongful imprisonment of opponents, the ruling elite and a lack of civilian supremacy.



## 5. #AntiCorruptionDay:

This is the only international topic we have selected for our study, but it is very directly relevant to Pakistan's political scenario where corruption is rampant. We can see here that the analysis generally reveals a negative sentiment since most of the tweets consist of keywords that relate to how corruption adversely impacts Pakistan and its people. Examples include tweets like "The land is bleeding" and "Corruption creates a fertile ground for organized criminal activity" etc. The sentiment analysis is also high on emotions like trust and positivity; this is because the content of the tweets is also about hoping to put an end to corruption and calling upon the people to raise their voice against this evil—for example, "We believe a #corruptionfree world is possible. On this international #AntiCorruptionDay we need YOUR help."



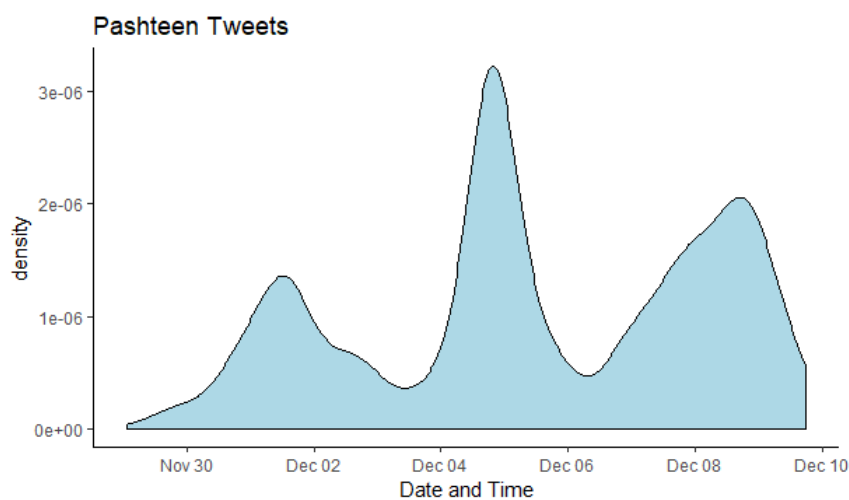
The word network tells us more about the public's sentiments regarding corruption using the phrases *"corruption undermines economic development"*, *"fosters inequality and injustice"*, *"threatens national accountability"* etc.

Since the question of whether or not these trends occur organically or are staged to influence public opinion rose multiple times in our previous analysis, we decided to conduct a Time Stamp Analysis on the times of creation of these tweets. The premise was that when a trend emerges organically, the incline in the trend of occurrence of the keywords is not very stark, even though an initial surge may occur. Moreover, the decline in the trend is generally slow and steady as the Twitter users slowly lose interest in that topic and move on to more current news. Contrarily, for an organized effort to trend certain ideas, the trend graph shows distinctive spikes for when most of the users become active at the same time and start talking about a topic. Similarly, the trend also dies as quickly as it started once the campaign ends. Our findings were as follows:

The trends of these two topics are linked but there is a subtle difference between their trends. The first one started abruptly around 18:00 on 30<sup>th</sup> November while the second one was in response of the first and started around the same time but peaked around 22:00 the same day. The decline of the #WeStandWithFIA one was sudden to a medium level and then another sudden drop to a very low level. The decline of #ShameOnFIA was more gradual and it had a few small bumps after midnight till the afternoon next day.

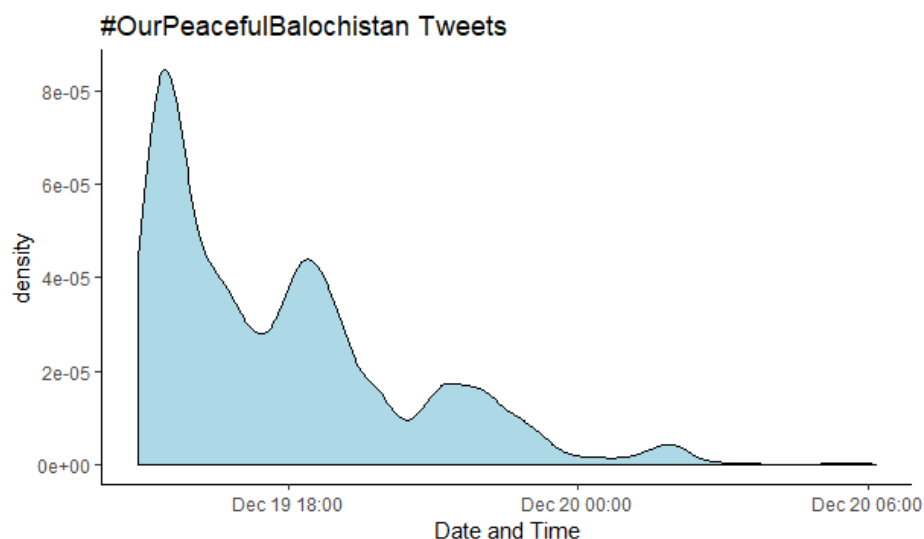
The sharp rise and fall of the first trend show that it may have been due to manipulation while the activity of the other is quite organic. The activities are quite similar for both otherwise which shows their link and the subtle difference in the delay of start of the second trend can also be noted.

### **Pashteen:**



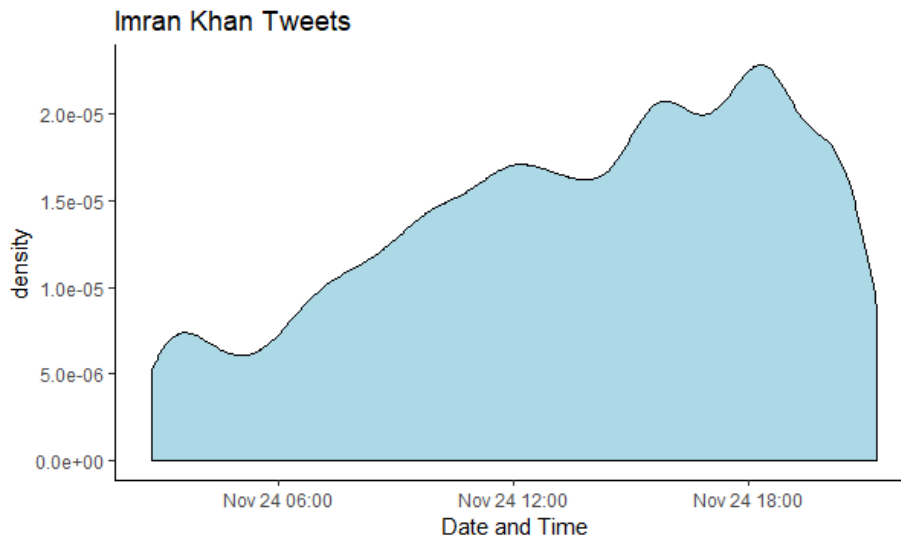
This is clearly an organic trend as there are frequent rises and falls in the activities related to the Pashteen trend. It starts gradually, declines slightly, rises to a peak between December 04 and 05 and then has another high on December 09 before finally declining.

### **#OurPeacefulBalochistan:**



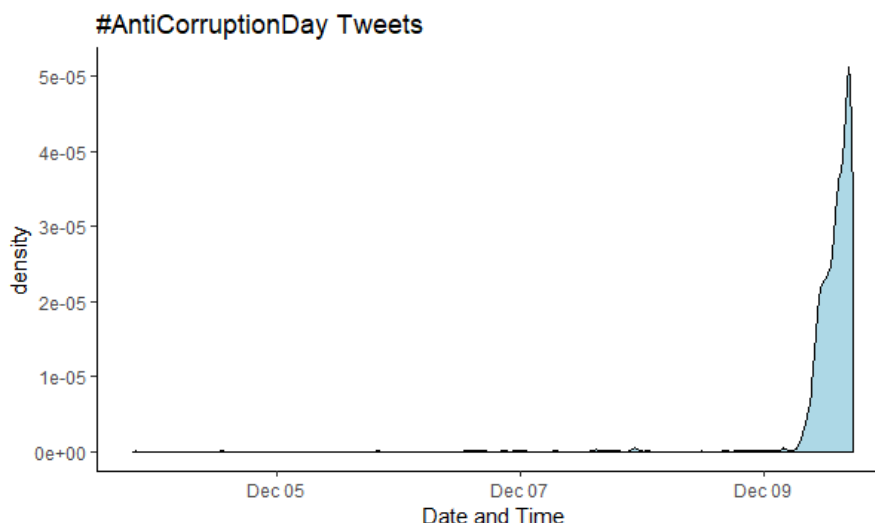
This is again interesting as the sudden rise of the number of tweets for #OurPeacefulBalochistan can be attributed to some manipulation. However, this cannot be asserted as the fall in activity is gradual and there are a few peaks before the trend stops completely.

### Imran Khan:



This topic is different from the others as this is not attributed to a certain trend or hype in activity. It just shows the tweets by the public related to the Prime Minister, Imran Khan. As we have only extracted 10,000 tweets for this topic, the activity covers little less than a day. However, this shows us an interesting trend related to hourly activity. The trend is not active in the early morning but starts rising until it reaches a peak around 18:00 and then falls steadily. As this is a general topic, we can associate this trend to general activity in a certain day.

### #AntiCorruptionDay:



There is a stark difference between the #AntiCorruptionDay tweets' activity and the rest. There is little to no activity until December 09 before it sees a sharp rise and then abruptly ends on the same day. This is not a representative of an organic trend as the chances of that happening are

next to none. The timestamp graph clearly shows that the trend has been manipulated and the rise and fall are difficult to explain otherwise.

### **Discussion:**

Interesting trends have emerged from this study that can allow us to predict public sentiment on political phenomena based on the indicators we have identified as useful. This public sentiment, of course, is limited to that on social media. Pakistan has a total population of approximately 200.8 million. According to a 2013 Open Society Foundations study<sup>i</sup>, Pakistan had a Twitter user base of approximately 3 million. While it is likely to have grown multifold over the coming years, the percentage of the entire population that uses Twitter is still very small. Moreover, the users predominantly belong to an urban, privileged and literate class who might not represent the views of their fellow countrymen, or even the people of their own regions. Hence, any opinions trending on Twitter need further external analysis before assuming them to be the opinion of the majority.

It is also important to take into account how original the tweets are to make any inferences about public opinion. If political trends are artificially engineered by the state, it gives a distasteful statement about state control of narrative as well as manipulation of public opinion through social media instead of allowing true freedom of speech and being open to criticism from its subjects.

While our study has certain reasonable limitations, it can serve as the basis of further analysis. Firstly, we have only studied tweets sent out in English when a lot of Pakistani twitter users also tweet in Urdu (using the Arabic or Roman script). A more comprehensive analysis can be done by including those tweets. Secondly, location data can be mined to study where the tweets are emerging from. It would especially add to our analysis of trend engineering.

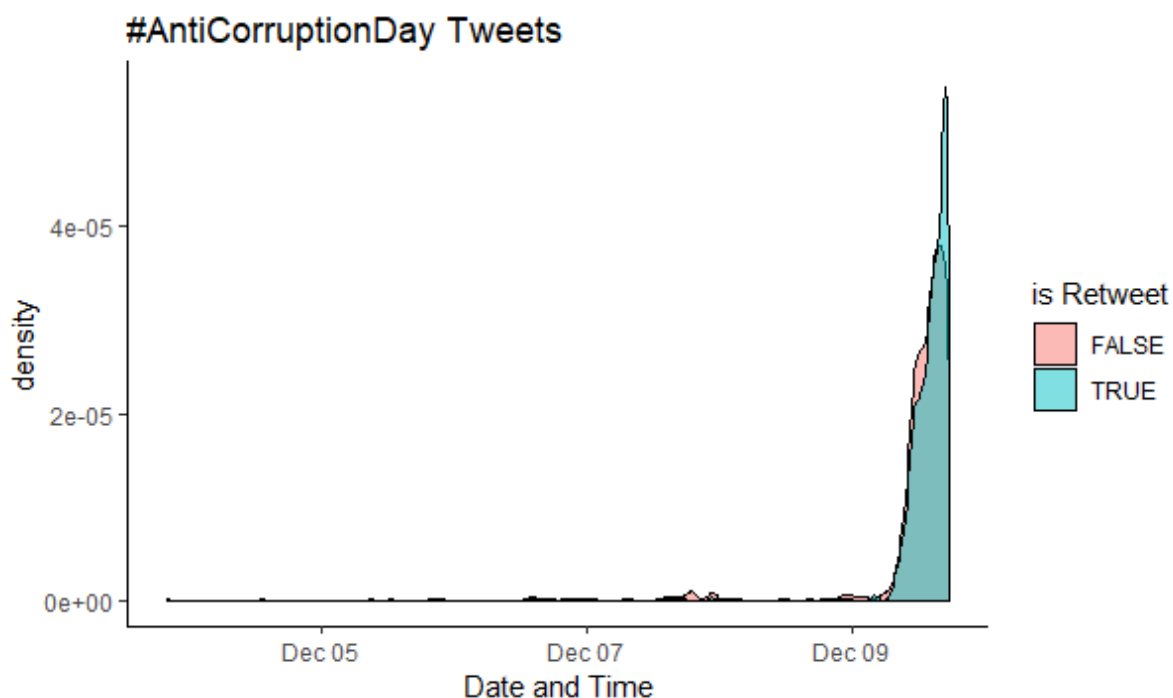


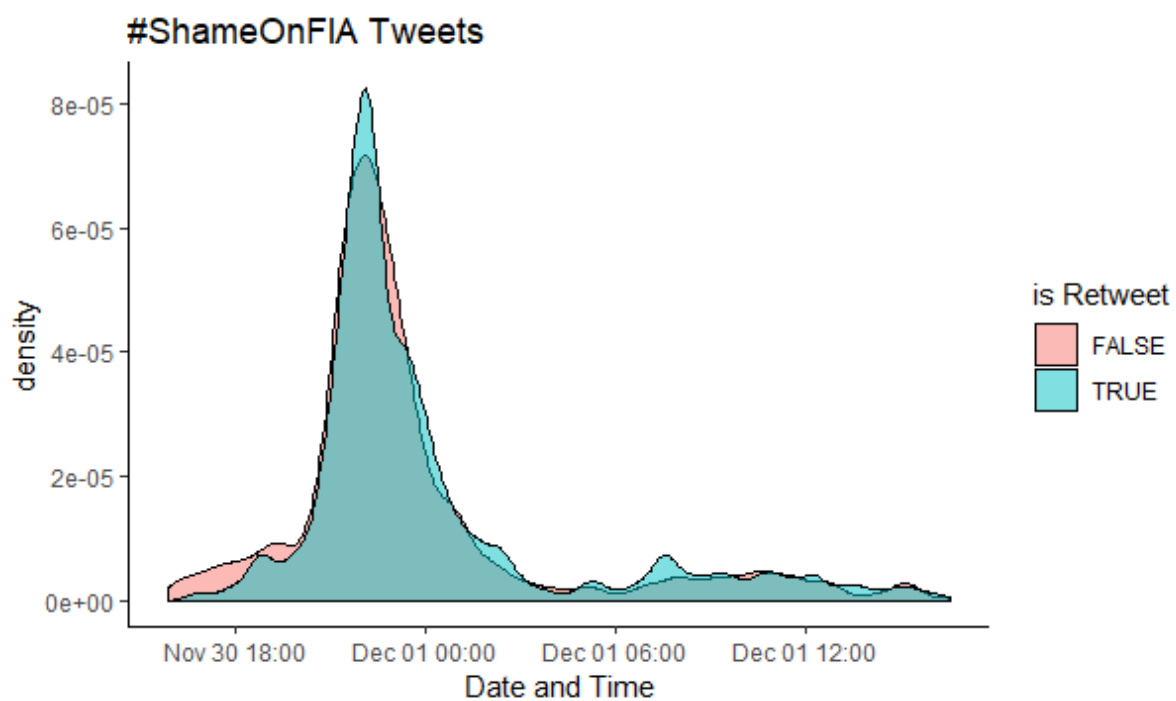
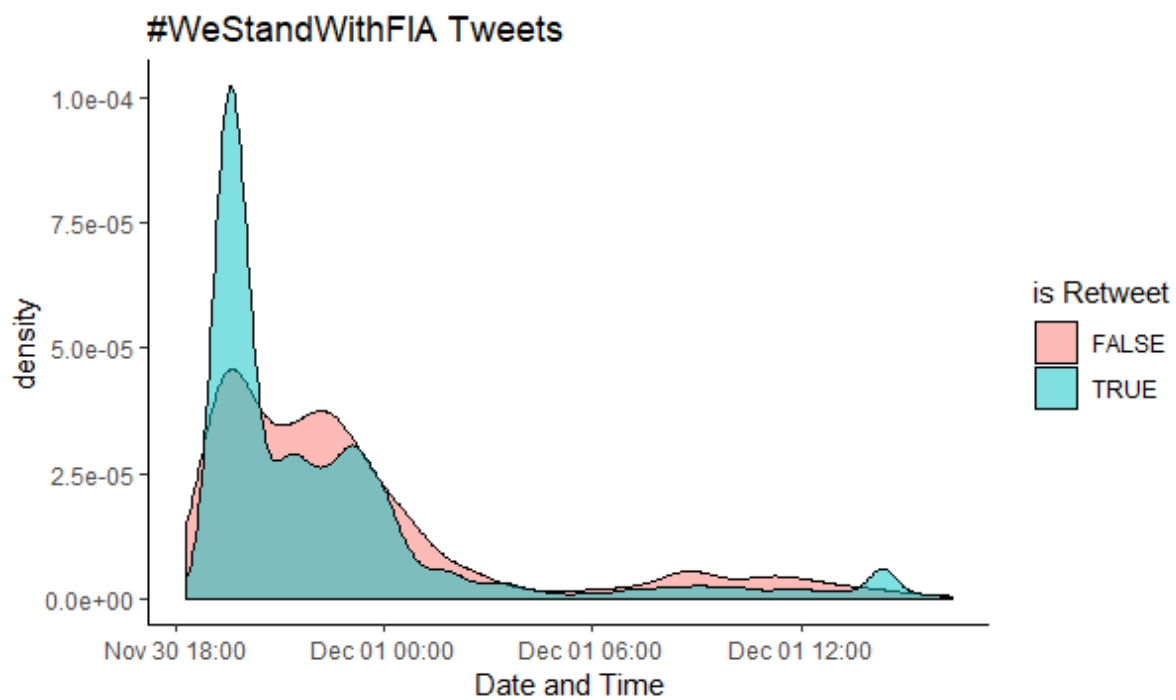
## Appendix

### Sample Table

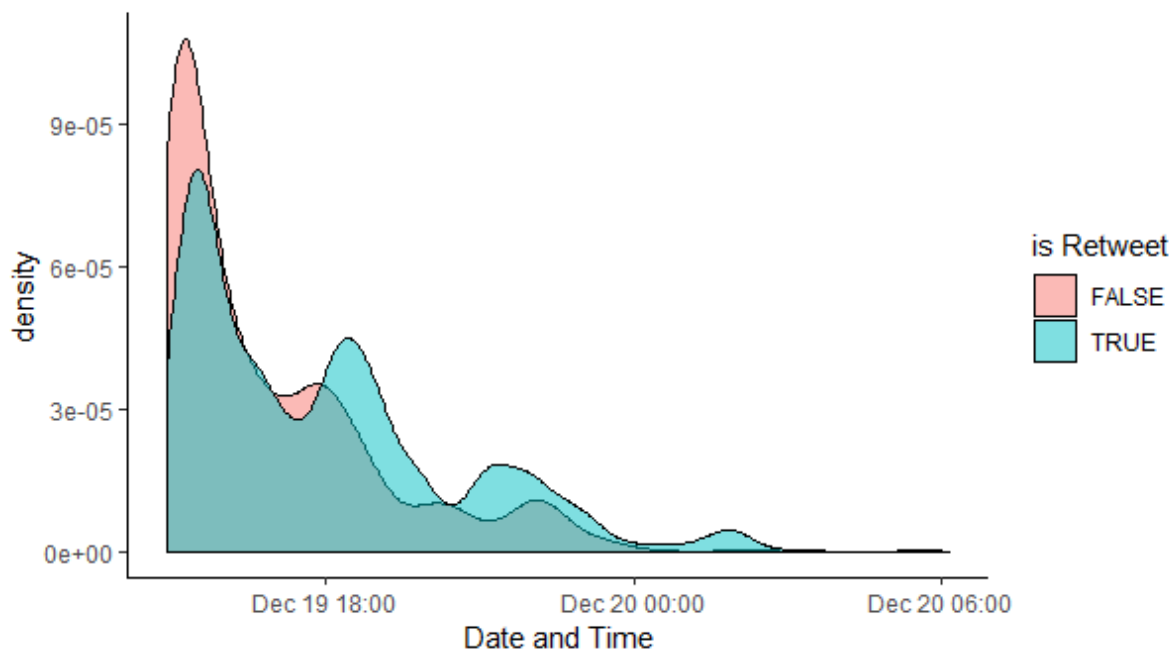
Trend	Stance	Sample Tweet
WeStandWithFIA	Positive	@viitvdotcom: Great job FIA for detaining seditious propagandists Manzoor Mehsud & Ali Wazir
	Negative	@AlamgirJogezai: #WeStandWithFIA by putting the names of democrates,peace lovers and seeking pashtoons and other oppressed nation's rights. Level of hypocrisy #ShameOnFIA
ShameOnFIA	Positive	@mo2005: Under what law democratically elected members arrested, does the mandate of waziristan people count #ShameOnFIA
	Negative	@wajidchattha: I have tried my level best to figure out this combo of so called liberals with so called nationalists. And the only common ground I found is that both are anti army. #ShameOnFIA
Pashteen	Positive	@a_siab: Banning peaceful political activities & expelling Manzoor Ahmad Pashteen from Sindh, in total violation of fundamental rights
	Negative	@DarabDawar @manzoorpashteen @SindhGovt Manzoor Pashteen is a traitor and an Indian agent. Disloyal to Pakistan
OurPeacefulBalochistan	Positive	@BurhanSayss: I My Opinion Balochistan Is The Peaceful And Safest Province Of Pakistan And Is The Best Place To Invest. #OurPeacefulBalochistan
	Negative	@FarhadKarim6: Wish we sensed a peaceful #Balochistan.?
ImranKhan	Positive	@amnajabeen11: We all are thankful to the PM Imran khan. Good strategy so far. #NationThanksTheState
	Negative	@BajiPlease: Another political blunder by Imran Khan. Economy is crippled and he is worried about cold - Talat Hussain
AntiCorruptionDay	Positive	Fighting corruption is not just good governance its self defense its patriotism #AntiCorruptionDay #Pakistan
	Negative	-

### Timestamp Analysis with Retweet

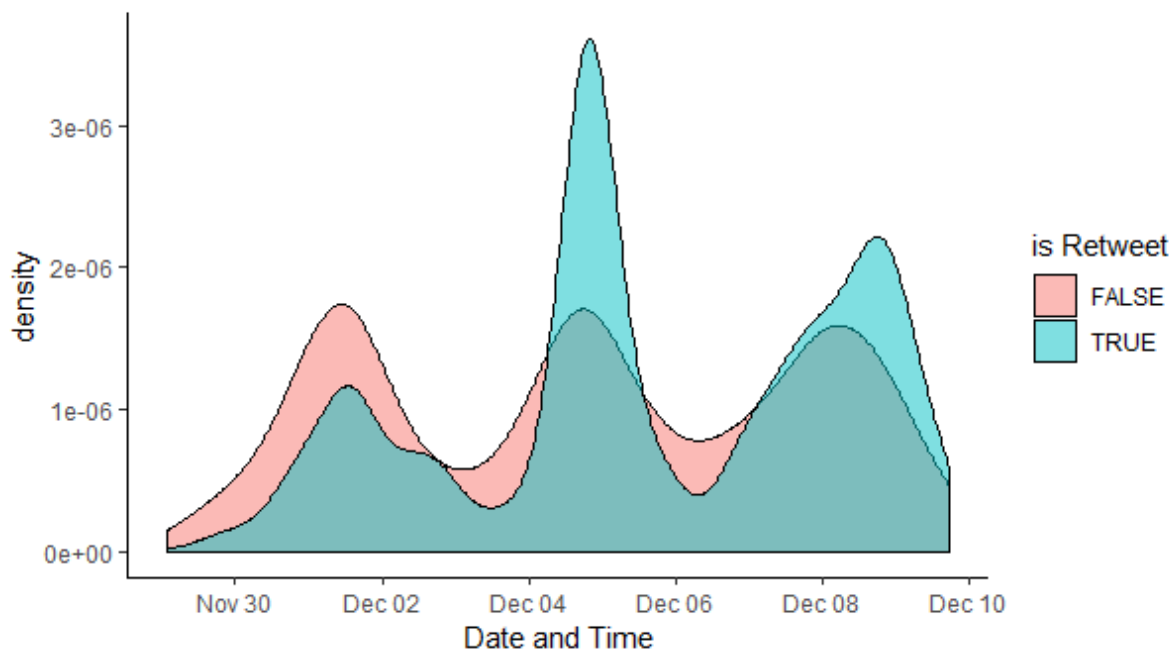


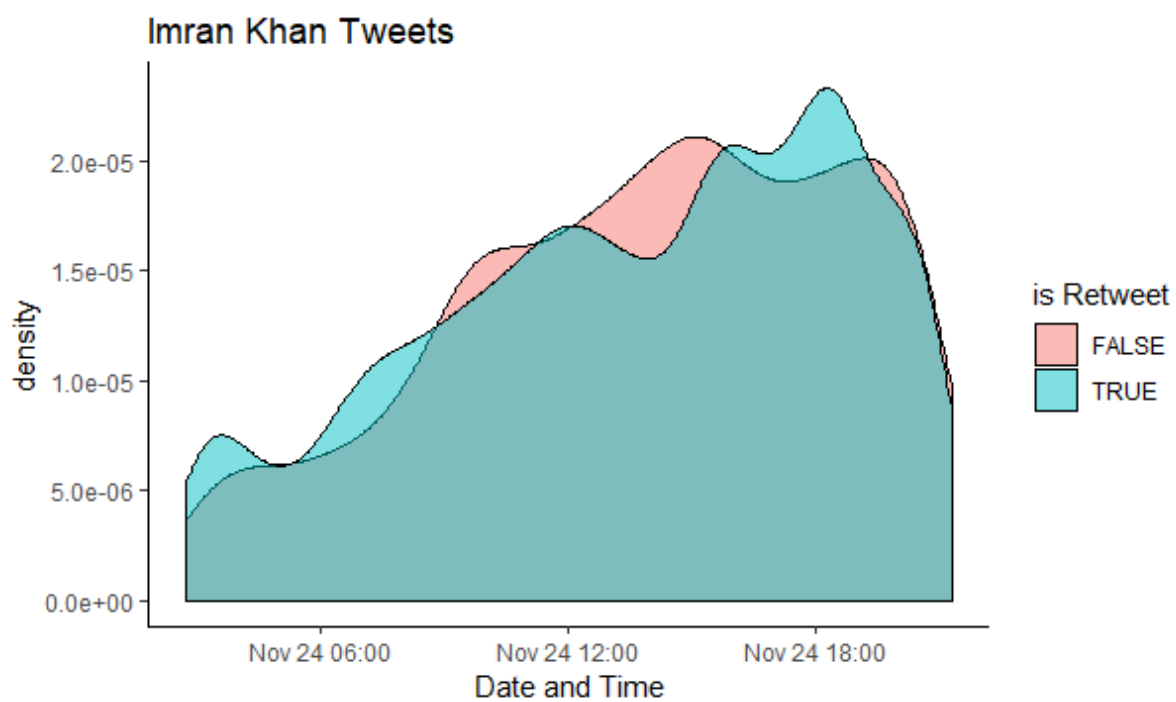


### #OurPeacefulBalochistan Tweets



### Pashteen Tweets





<sup>1</sup> <https://www.opensocietyfoundations.org/sites/default/files/mapping-digital-media-pakistan-20130902.pdf>