**Why Twitter ?**

What’s so interesting about Twitter ?

People use it more than we probably realize, and are affected by it more than they realize.

Since it’s creation in 2006, Twitter has become the largest communication tool in the world. Worldwide, monthly active users are in the 330 million range, with 79% of all users outside the United States. In the US, there are 67 million users. Currently, Twitter can handle 18 quintillion users, which would account for future growth. That number looks like this: 18,000,000,000,000,000,000!

With such a worldly impact from Tweets, the communication tool from Twitter, it makes sense to understand the origination of a Tweet, it’s sentiment and the source and with possible effect on readers.

That is the basis of our look into Twitter. What is real or fake?

A few questions to start our inquiry:

**WHO ARE THE TOP 10 TRENDING TWITTER USERS?**

**WHAT ARE PEOPLE SAYING AND HOW DO WE INTERPRET WHAT IS SAID?**

**IS THERE A CORRELATION TO THE USE OF TYPES OF DEVICES AND TWEETS?**

**HOW DO WE IDENTIFY AUTOMATED TWITTER ACTIVITY AND ITS SENTIMENTAL EFFECT?**

Our inquiry starts with the identification of the Top 10 Trending Users as found on the website, Tweeplers. URL: <http://tweeplers.com/?cc=WORLD>   
Taking a snapshot in time, a data set was created for the most recent Tweets directed towards those Top 10 Trending Users of 500 Tweets each.

[Zachary]

The compound sentiment values from the Vader Sentiment tool can be misleading when looking at the average of the compound values. When combining the average compound score with a scatter plot view of the sentiment, it is much easier to deduce the overall distribution of tweet sentiment.

Though many fluctuations occur, the top ten trending accounts overall attract a positive sentiment at any given time, readily outpacing negative sentiment. This suggests that twitter users generally are more likely to retweet a status with a positive sentiment as compared to negative sentiment.

[Mandy]

**Wordcloud**

Word cloud is a graphical representation of frequently used words within text. The height (size) of each word in this picture is an indication of frequency of occurrence of the word in the entire text.

Word cloud can be a used as tool to help analyze unstructured data. To be able to count the frequency of data STOPWORDS is required to eliminate commonly occurring words.

Data cleanup (removing characters, HTTPS, RT’s) is required to see word clouds that are of value. Word cloud analysis to some extent is less precise for text with 240 characters or less.

**Language**

There are some common challenges of sentiment analysis such as emoji analysis, word order, spelling and certain words that may have a different or opposite meaning in certain situations. After observation is was clear that the sentiment analyzer is not accurate for multi-language text. No values recorded for Russian as an example. English text (without multi language) appeared to be more accurate.

[James]

Analyzing the source of each Tweet, we can see that most Tweeters, globally, are using Android Phones,

and users are mostly using cell phones to tweet. An interesting thing we discovered, many people are using a program called, "IFTTT", which stands for “If this then that.” It's an app that automates different services from app to app and might be a good place for further exploration in finding BOTS or people that have automatic responses.

In one bar chart, I users in bins according to Tweets per day. The first (0-100) and last, (900-1000) were removed because the data was not desired to explore. Most users over, (4,000+) are only Tweeting 0-100 times per day, so the data skewed the visualization too much, so it was removed. The first element is (100-200) tweets per day, and you can see a little over 200 people are Tweeting over 100 times a day! You can also gather that there is an odd quantity of users Tweeting 800-900 times a day!

This would be a great place to further explore BOT presence.

[Verna]

This segment of the analysis looks into the first of the 10 Most Trending Users captured (realDonaldTrump), and the Tweets directed towards them from other Twitter users.

The first graph importance of this inquiry is to see how the sentiment of the dataset of 500 Most Recent Tweets shows in a sorted line chart. It is clear by the length of the segments of the chart, that the sentiments of Tweets are fairly even between positive, neutral, and negative sentiments. But this chart does little for understanding the authorship of Tweets and the source of the Tweet. A further look into the Tweet user is needed.

Unnderstandimg the validity of Tweets and their content, necessitated the use of the integration of the Botometer ratings into the information matrix. An actual photo ‘url’ for each Tweet was used as a graph marker and then exploded to help understand Tweet activity and identify possible BOT suspects. With the interactive programming, higher BOT rated Tweets can filtered for, along with selectable sets of Tweets to analyze.

It was found that suspected BOT activity existed for Donald Trump in all 3 categories of negative, neutral and positive sentiments, however trending towards negative, the higher the BOT value threshold. Understanding that the dataset given was a mere 500 tweets, that was obtained from a time period of less than a minute, a larger dataset would give a more reliable sentiment.

With this method of filtering out or including BOTS in a dataset, and then looking at the results, a more genuine sentiment could be seen for any user collected Tweet data on a given user. Hopefully this would reflect a truer overall communication.

This coding, with a few enhancements could become a useful tool for analyzing any Twitter user’s account who desires to see how the world is reacting to their activity- a tool that could help identify possible mischievous actors for security or other purposes. A useful application of this analysis would be to select the data set of Tweets that occurred after a particular User Tweet to analyze its affect on the public audience by analyzing their Tweets that came after, such as for political or security reasons as examples.

Summary

To summarize this Twitter analysis, there are methods to look at sentiment of Tweet text that can give different results. However, they are both useful In an estimated way of interpreting Twitter for its users. Understanding trends such as devices, are also clues into how Twitter is being used and is worthwhile to investigate further. With data changing every fraction of a second, it is also understood that the results of any analysis will change as the topics of interest change. However, those moments in time are capturable and analyzable, it is clear that Twitter has actors whose Tweets are not a standard human entered responses, but automated with purposeful intent that could have any sentiment. Filtering out these actors is needed for a better understanding of

“the Real DonaldTrump” or any user!