Twitter, Trump, and Time

How timing of Trump's tweets affect sentiment of retweets

Mao Li & Curtiss Engstrom

2019-12-12

Contents

Github repository for Data and extended code: https://github.com/cwengstrom/FinalPaper

Introduction

Over the past 10 years, the popularity of Twitter has exploded from 50 million worldwide active users in 2010 to 350 million active users in 2019 (Clement J (2019)). With such a wide user base, it's no surprise that politics entered the Twitter-verse, especially in 2016. Then candidate Donald Trump used Twitter constantly and got his message across to his supporters and to get his message on the news (Newport (2019)). Starting with his presidential run announcement in June 2015, as Trump gained in popularity in politics, as did his Twitter followers (Tsur et al. (2016)). It wasn't until his election as US President, do we see more and more political discourse in Trump's followers, especially from political ideologues, people are strong partisans for one US political party.

Some recent work on this issue by Brian Ott at Texas Tech discusses the inevitable rise of daily political news from Twitter. Twitter itself is impulsive, uncivilly, and simple (60-62, Ott (2017)). This leads to people speaking their minds about politics at any moment, allowing for more discourse on a daily basis. As such, it create echo chambers, where people look for what they want to believe, and argue over things they don't necessarily approve of (64-65). More importantly, Ott discusses how Twitter hit the mainstream media as news, when its opinions and not facts (65-66). With that said, if it hits the news, there will be consequences of it, mostly more divisive tweets from people in an increasingly hostile political discourse.

That said, the language used in discussions becomes more emotional the more words used by people in political discussions with people of the opposite party (Brady et al. (2017)). Does this theory in conversation translate to Twitter, especially in an increasingly politically charged Twitter? Along with that, it is understood that people retweet Trump's tweets at an alarmingly high rate, especially after his election (Carr (2018)). However, no research has been conducted looking at when people retweet political tweets.

Along with political discourse, re-tweeters can also be fact-checkers. A fact-checker is someone who checks the validity of the claim of a tweets and retweets a comments about said validity. There has been some research looking into the spread of fact-checkers on Twitter through true or false rumors and people checking their validity, with mixed results (Vosoughi, Roy, and Aral (2018)) on their effectiveness. There is no telling when these fact checkers retweet someone.

With our motivation stated from previous scholarship and recent news, we have two main hypotheses. Our first hypothesis is that within the first half-hour, more political ideologues will retweet President Trump's tweet. Based on the literature of political discourse in conversation and the prevalence of Trump's tweets in mainstream culture, as well as how popular Twitter is, we believe people with strong opinions towards Trump will be the first to retweet either support or anger towards his tweets. Along those lines, we understand fact-checkers play a role in how prevalent a tweet becomes, based on its validity, therefore, we hypothesize that time will be taken by these fact checkers and will result in more negative retweets towards Trump because, as research suggests, Trump is a liar (Ross and Rivers (2018), A. (2019)) and anything that may correct him, is labeled "fake news" (Ott (2017)).

Data

We used Twitter API as the main data source and rtweet package as a data-gathering tool. Each time we use the search tweet function to request data, we can obtain approximately 18000 records in two hours. Thus, we run the search tweet function at a certain time point (Nov. 25th) to gather retweet data for a whole day. An example of this code is shown below:

```
Trump <- search_tweets('to:@realDonaldTrump',geocode = lookup_coords("usa"),n=3000)
```

After gathering the data, we used row bind to merge each dataset of two hour together. In the end, we got a "Trump" dataset which contains about 125,000 observations. Since the data gathering step can't be replicated, we have attached the data at the end of the paper.

Results

Data exploration

First, we are interested in the overall picture for these retweets about President Trump. To do this, we used the tidytext package to see the Common Sentiment words.

Based on the outcome of the common sentiment words, we can tell there is a polarization trend of attitude about President Trump.

After that, we used the 'reply_to_status_id' variable to identify each retweet response to a tweet posted by Trump or just a general comment about Trump.

We can see most of retweets just a general comment to Trump rather than towards certain tweet. From this, we looked at the three most replied-to retweets from Donald Trump, creating the tweets we will analyze.

```
first_mostly_reply <- filter(Trump,reply_to_status_id=='1198948075928268800')

second_mostly_reply <- filter(Trump,reply_to_status_id=='1198981079723589639')
```

For each tweet's retweet, we used the same method to explore the data. First, using regular expressions to extract the useful information from text variable. Then applying the sentiment analysis package to do the sentiment analysis about the text in the retweets. The outcome data frame has 15 variables and we chose "SentimentGI" as the main variable of analysis since it shows the overall sentiment about each retweet. Then, using ggplot plus the smooth option, we will be able to see the overall trend of sentiments about retweets. After that, we will filter positve sentiments to show the change through time as well as negative sentiments. In the end, we will use the overall data about sentiments to fit the ARIMA model of time analysis to see if there is any robust relationship between time and sentiments for each tweet.

Analysis

```
sentiment_second%>%
  filter()%>%
  group_by(time)%>%
  summarise(n=n())%>%
  ggplot(aes(x=time))+
  geom_histogram(bins=50,color='darkgrey',fill='#ae6bff')+
  labs(title = "Figure 1: Amount of Retweets over Time",
       x='Time',
       y='Number of retweets',
       caption = 'Source: Twitter API')+
```

```
  theme(plot.title=element_text(size=20,
                                color = 'tomato',
                                hjust=0.5,
                                lineheight=1.2),  # title
        plot.subtitle=element_text(size=7,
                                   color = 'tomato',
                                   hjust=0.5))
```

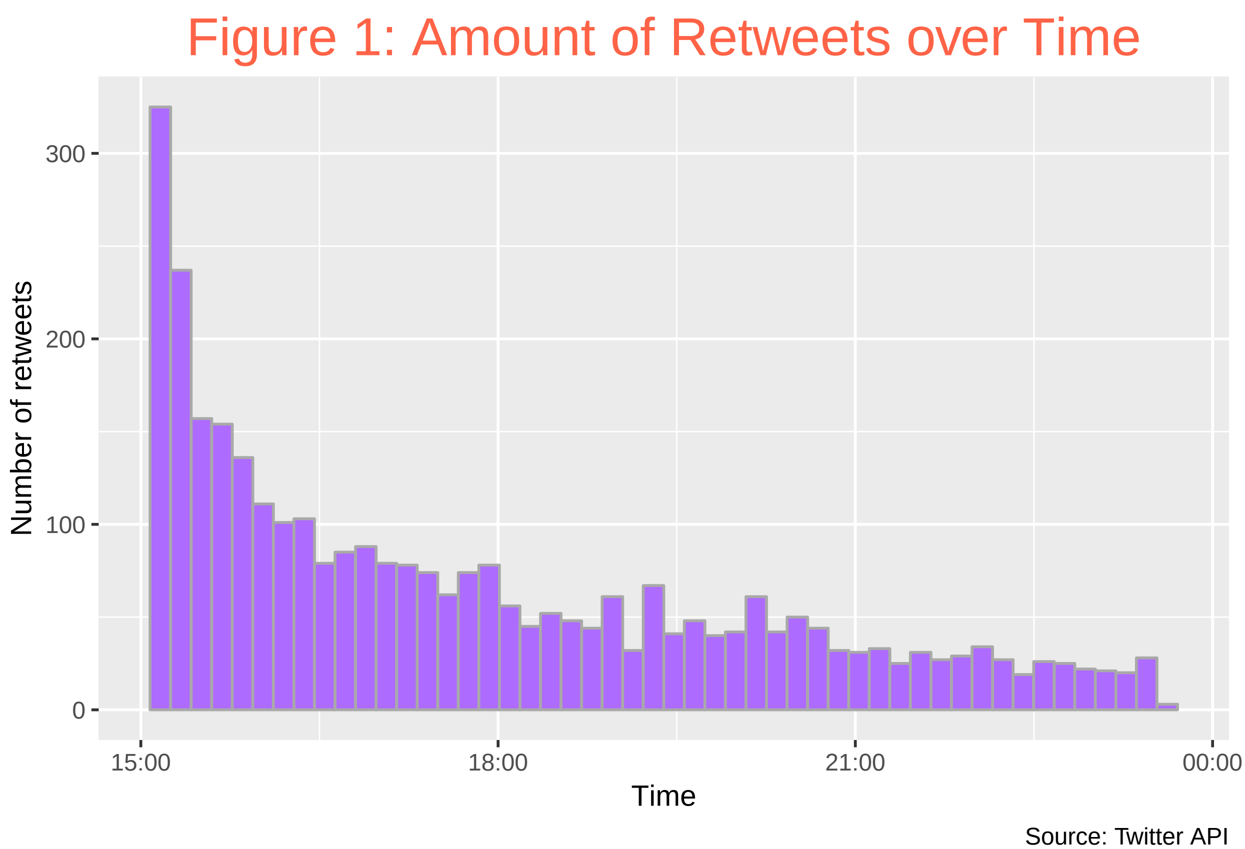


Figure 1 describes the overall distribution of retweets for the first Trump tweet. We see that most of the retweets occur during the first 90 minutes after Trump tweets the original tweet. This gives some evidence to support our first hypothesis that political ideologues retweet right after Trump tweets. However, this does not support our second hypothesis about fact checkers. Though not supportive, until we breakdown the retweets through sentiment analysis, will we get a full picture about who retweets when.

```
sentiment%>%
  ggplot(aes(x=time,y=SentimentGI))+
  geom_smooth(color='#dc4eff',fill='orange')+
  theme_bw()+
  labs(title="Figure 2: Sentiment Trend for Tweet 1",
       subtitle = 'Political Statement | Overall',
       y="Sentiment",
       x="Time",
       caption="Source: Twitter API")+
  theme(plot.title=element_text(size=20,
                                color = 'tomato',
                                hjust=0.5,
```

```
                                   lineheight=1.2),   # title
        plot.subtitle=element_text(size=7,
                                      color = 'tomato',
                                      hjust=0.5))
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

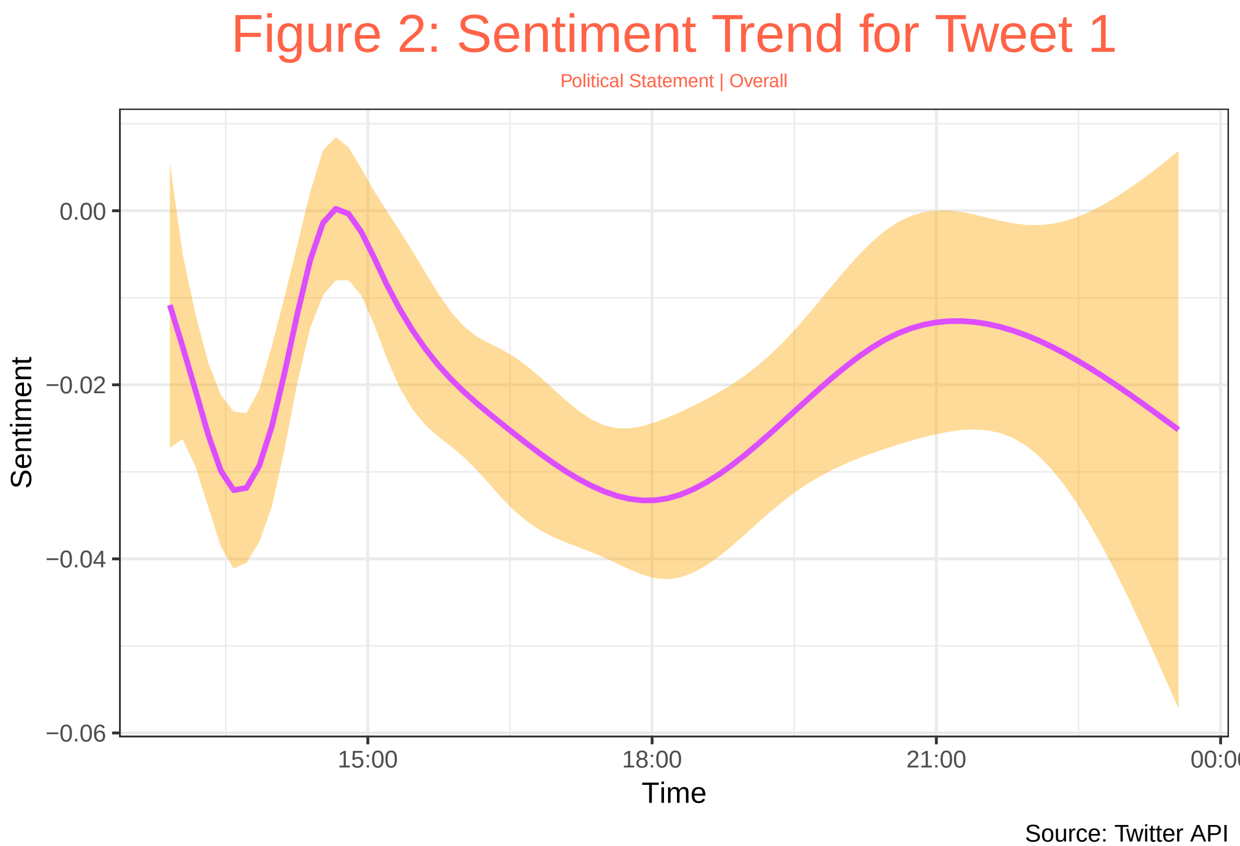


Figure 2 is the overall sentiment trend for the first tweet. At first the trend is fairly neutral leaning negative for the most part. There is some variation through time, but not enough to say the overall retweets are negative.

```
sentiment%>%
  filter(SentimentGI>0)%>%
  ggplot(aes(x=time,y=SentimentGI))+
  geom_smooth(color='#dc4eff',fill='orange')+
  theme_bw()+
  labs(title="Figure 3: Sentiment Trend for Tweet 1",
       subtitle = 'Political Statement | Positive: SentimentGI > 0',
       y="Sentiment",
       x="Time",
       caption="Source: Twitter API")+
  theme(plot.title=element_text(size=20,
                                   hjust=0.5,
                                   lineheight=1.2),   # title
        plot.subtitle=element_text(size=7,
                                      hjust=0.5))
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

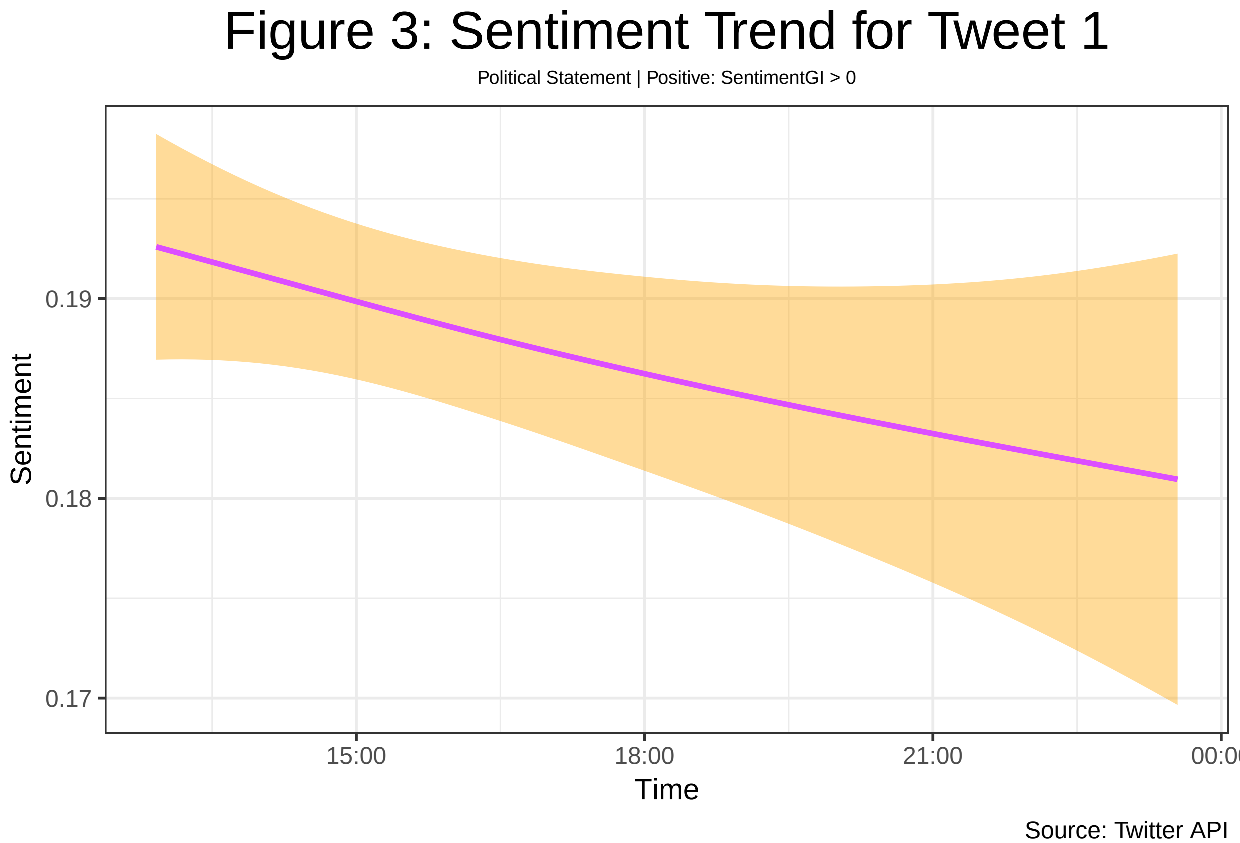


Figure 3 shows retweets that have a positive sentiment for the first tweet. We see there a linear pattern with sentiment leaning neutral over time. However, we see the highest positive sentiment within the first 30-60 minutes of the tweet, supporting the first hypothesis.

```
sentiment%>%
  filter(SentimentGI<0)%>%
  ggplot(aes(x=time,y=SentimentGI))+
  geom_smooth(color='#dc4eff',fill='orange')+
  theme_bw()+
  labs(title="Figure 4 : Sentiment Trend for Tweet1",
       subtitle = 'Political Statement | Negative: SentimentGI<0',
       y="Sentiment",
       x="Time",
       caption="Source: Twitter API")+
  theme(plot.title=element_text(size=20,
                                color = 'tomato',
                                hjust=0.5,
                                lineheight=1.2),  # title
  plot.subtitle=element_text(size=7,
                             color = 'tomato',
                             hjust=0.5))  # subtitle
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

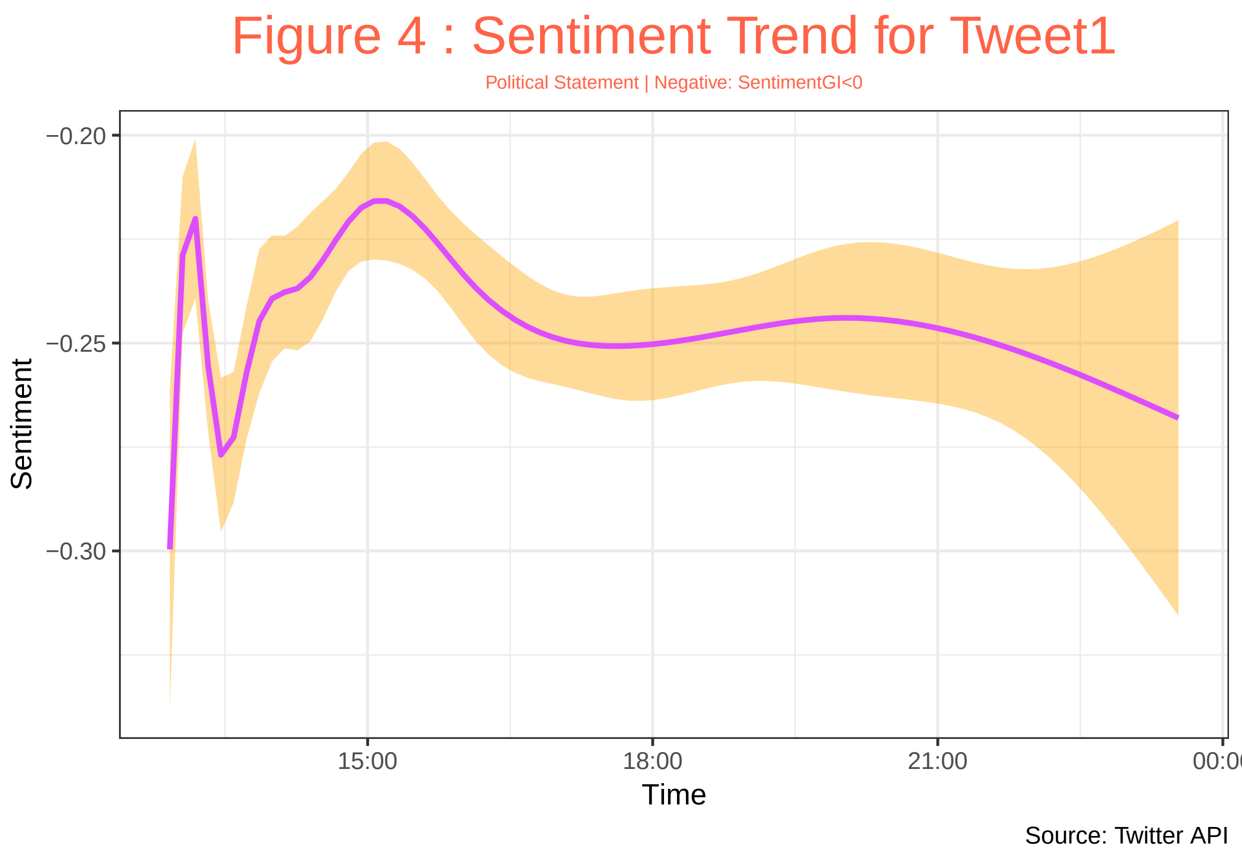


Figure 4 shows retweets that have a negative sentiment for the first tweet. We see an initial high degree of negative sentiment, followed by a shift towards neutral sentiment, then a second downward trend in sentiment. After that, it starts to level off at a middle level of negative sentiment. This trend suggests that people with the strongest negative opinions towards the first tweet by the president tweet right away, followed by a secondary wave of strong negative sentiment, which could be the fact checkers we hypothesize retweeting.

```
sentiment_second%>%
  ggplot(aes(x=time,y=SentimentGI))+
  geom_smooth(color='#dc4eff',fill='orange')+
  theme_bw()+
  labs(title="Figure 5: Sentiment Trend for Tweet 2",
       subtitle = 'Fact Statement | Overall',
       y="Sentiment",
       x="Time",
       caption="Source: Twitter API")+
  theme(plot.title=element_text(size=20,
                                color = 'tomato',
                                hjust=0.5,
                                lineheight=1.2),  # title
        plot.subtitle=element_text(size=7,
                                   color = 'tomato',
                                   hjust=0.5))
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

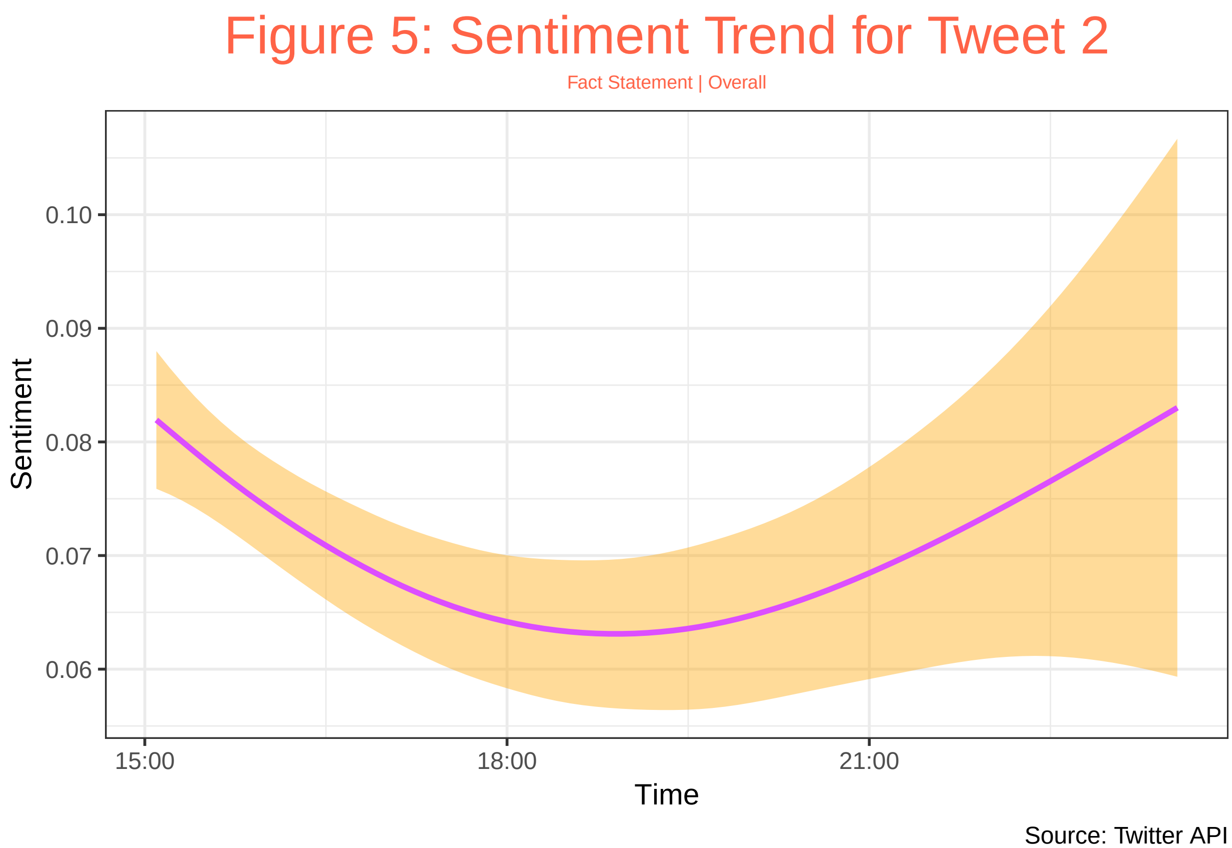
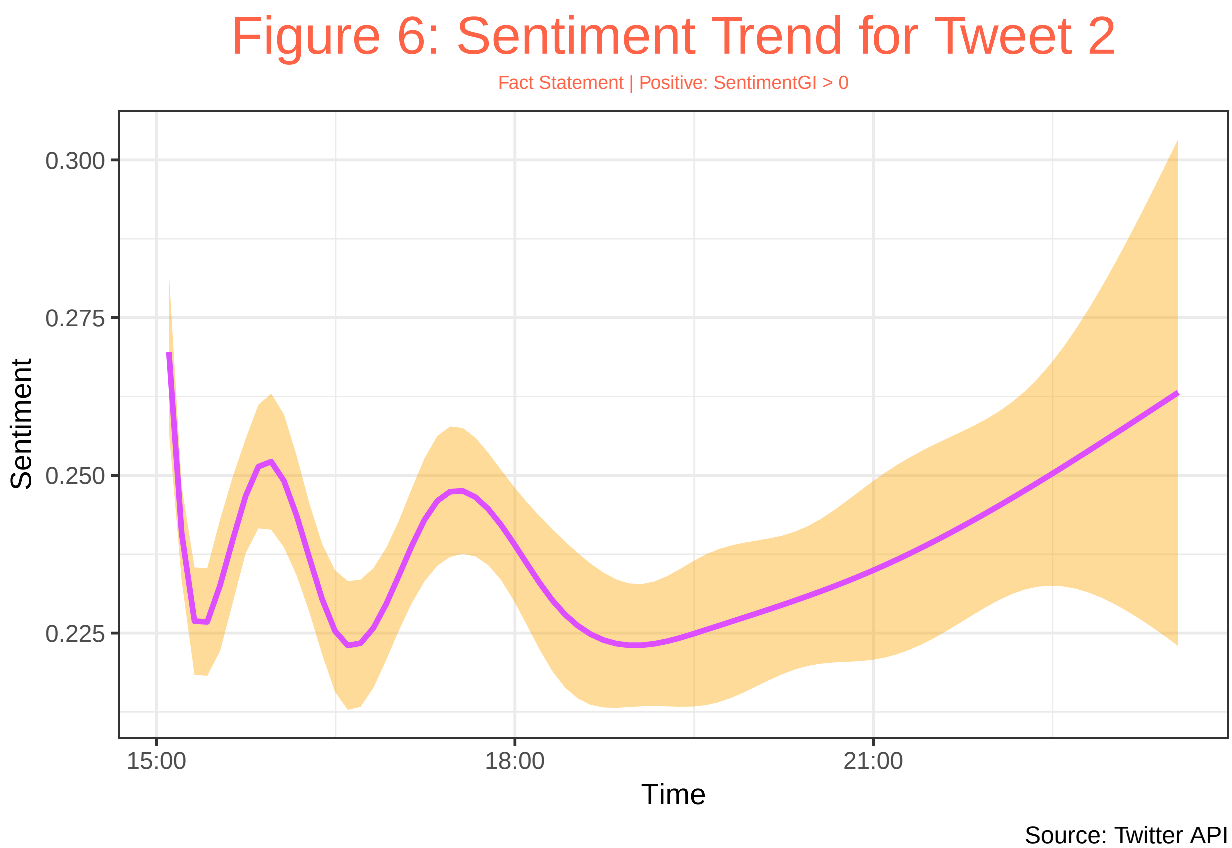


Figure 5 is the overall sentiment trend for retweets of the second tweet. The general trend here is mostly positive sentiment. This suggests some bias, not systemic, but in this tweet for a positive affinity towards it. This will be taken into account when we look through the sentiment broken down by positive or negative sentiment.

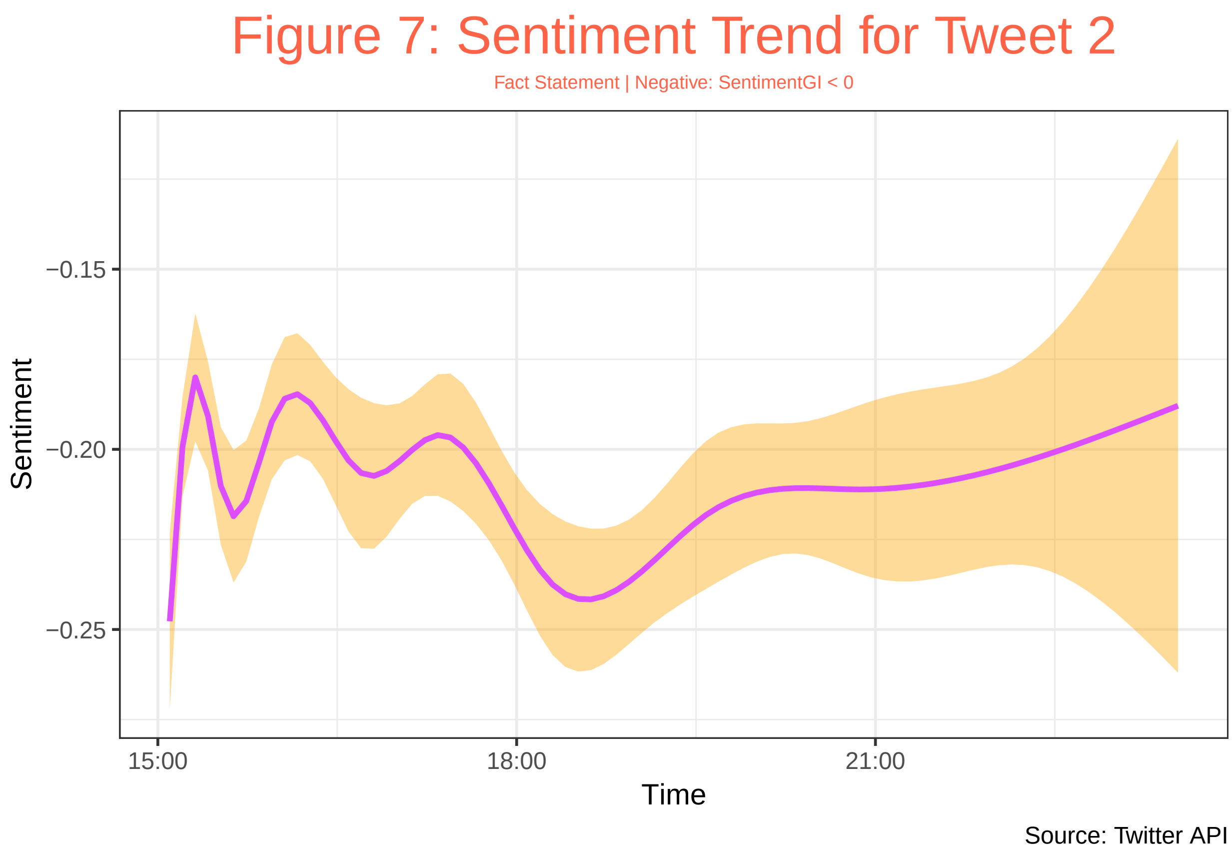
```r
sentiment_second%>%
  filter(SentimentGI>0)%>%
  ggplot(aes(x=time,y=SentimentGI))+
  geom_smooth(color='#dc4eff',fill='orange')+
  theme_bw()+
  labs(title="Figure 6: Sentiment Trend for Tweet 2",
       subtitle = 'Fact Statement | Positive: SentimentGI > 0',
       y="Sentiment",
       x="Time",
       caption="Source: Twitter API")+
  theme(plot.title=element_text(size=20,
                                color = 'tomato',
                                hjust=0.5,
                                lineheight=1.2),  # title
        plot.subtitle=element_text(size=7,
                                   color = 'tomato',
                                   hjust=0.5))
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
sentiment_second%>%
  filter(SentimentGI<0)%>%
  ggplot(aes(x=time,y=SentimentGI))+
  geom_smooth(color='#dc4eff',fill='orange')+
  theme_bw()+
  labs(title="Figure 7: Sentiment Trend for Tweet 2",
       subtitle = 'Fact Statement | Negative: SentimentGI < 0',
       y="Sentiment",
       x="Time",
       caption="Source: Twitter API")+
  theme(plot.title=element_text(size=20,
                                color = 'tomato',
                                hjust=0.5,
                                lineheight=1.2),   # title
        plot.subtitle=element_text(size=7,
                                   color = 'tomato',
                                   hjust=0.5))
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



Figures 6 and 7 show the trend for positive and negative sentiment, respectively. For the initial 30 minutes of retweets, we see a similar pattern found in figures 3 and 4, very strong opinions retweeted right away with a sharp digress afterwards. In figure 7, we see the same secondary bump after about 60-90 minutes as in figure 4. However, that same secondary bump is found in figure 6, though it was not found in figure 3, which may suggest fact-checkers, but it may also suggest type and timing of tweets may play a factor in who retweets.

After running a ARIMA model on our data, we see a overall negative trend for the first tweet for retweet sentiment (-0.0194 (SE= 0.002)). However, in the second tweet, we see an overall positive sentiment trend (0.0726 (SE = 0.002)).

Discussion

Looking at everything, we can draw some conclusions on our hypothesis. Our results support our first hypothesis that political ideologues tend to retweet right away after Donald Trump tweets. With the sentiment analyses for the first two tweets having strong sentiment in either direction for the retweets, people with strong opinions tend to retweet right away. As for our second hypothesis about grouping fact-checkers and ideologues based on timing, we have minimal support for that conclusion. While we do see a secondary bump in negative sentiment retweets for both tweets, we also see a second bump in the positive sentiment of the second tweet, which is not present in the first tweet. Also, we have a lot of "noise" after the first hour and a half for sentiment in either direction, suggesting some else is going on. Along with that, fact-checkers may have their retweets be considered "neutral" and we did not look at neutral sentiment retweets.

With that said, we have some limitations on our exploratory study. First, we only looked at two tweets for sentiment analysis of retweets. If we were to replicate this, we would do it over a higher number of tweets with more diversity of political opinions and type. With that said, we only looked at tweets from Donald Trump. If we want these results to be generalized, we need to look at different political ideologies and differing reputations of politicians. That said, Donald Trump presents a unique case of hyper-twitter usage from an elected official with a strong following and a strong opposition to him. Along those lines, if we do this again, looking at Donald Trump, we need to get a variety of tweets over different times of day. Our tweets mostly come from late morning/early afternoon. If we want to see if time of day affects who retweets, we need to spread the tweets over different days and different times. Also, for variety of tweets, we need a coding mechanism to distinguish between tweet types, such as opinions, factual claims, or quotes. Tweet type may affect who retweets and would be interesting to look at.

Another limitation on our results is that we have no demographic data on who is retweeting. Demographic data would give us an idea on what type of people are retweeting, such as gender of re-tweeters, race, educational background, etc. Also, we need to look out for bots in future analyses. Bots are machines that automatically retweet someone as soon as someone tweets. In future analyses, we need to remove these for a better picture on who is actually tweeting. Lastly, we have a lot of retweets that claim to be "neutral", meaning no positive or negative sentiment within the retweet. This could be problematic as some people retweet a video, GIF, or a picture on someones tweet, and the R package we used does not distinguish that. Thus, in future analyses, we need to look at the neutral tweets and manually place them in either a positive or negative sentiment category.

For future analyses, we will try to incorporate some of the ideas discussed in the limitations noted above, as well as some other ideas. This includes using machine learning to catch sarcasm within retweets to correctly place them in either positive or negative sentiment. Along with that, future analyses would include regression analysis of sentiment on time to understand times effect on sentiment. This would provide for more robust results on times effects. Lastly, for determining fact-checkers versus political ideologues, we should incorporate machine learning to separate tweets based on wording and content into three categories: fact-checkers, political ideologues, and something else. Doing this would create three categories that we could run effect models onto determine each groups effect on time of retweet.

References

A., N. 2019. "Analysis | Tracking All of President Trump's False or Misleading Claims." *The Washington Post*. https://www.washingtonpost.com/graphics/politics/trump-claims-database/; WP Company.

Brady, William J., Julian A. Wills, John T. Jost, Joshua A. Tucker, and Jay J. Van Bavel. 2017. "Emotion Shapes the Diffusion of Moralized Content in Social Networks." *Proceedings of the National Academy of Sciences* 114 (28): 7313–8. https://doi.org/10.1073/pnas.1618923114.

Carr, Nicholas. 2018. "Why Trump Tweets (and Why We Listen)." *POLITICO*. https://www.politico.eu/article/why-president-donald-trump-tweets-and-why-we-listen/. POLITICO.

Clement J. 2019. "Twitter: number of active users 2010-2019." Website. Statista; https://www.statista.com/statistics/282087/n of-monthly-active-twitter-users/.

Newport, Frank. 2019. "Deconstructing Trump's Use of Twitter." *Gallup.com*. https://news.gallup.com/poll/234509/deconstruc trump-twitter.aspx; Gallup.

Ott, Brian L. 2017. "The Age of Twitter: Donald J. Trump and the Politics of Debasement." *Critical Studies in Media Communication* 34 (1): 59–68. https://doi.org/10.1080/15295036.2016.1266686.

Ross, Andrew S., and Damian J. Rivers. 2018. "Discursive Deflection: Accusation of 'Fake News' and the Spread of Mis- and Disinformation in the Tweets of President Trump." *Social Media Society* 4 (2): 1–12. https://doi.org/10.1177/2056305118776010.

Tsur, Oren, Katherine Ognyanova, David Lazer, Juleanna Glover, Jeff Greenfield, and Alexandra Glorioso. 2016. "The Data Behind Trump's Twitter Takeover." *POLITICO Magazine*. https://www.politico.com/magazine/story/2016/04/donald-trump-2016-twitter-takeover-213861.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–51. https://doi.org/10.1126/science.aap9559.