DESAUTELS | McGill

INSY 670-075

Social Media Analytics

Presented to Professor Changseung Yoo

Final Assignment

Prepared by:

Ali Al Samuraee	260739157
Mesaye Bahiru	260634934
Cristina Esposito	260744222
Alya Gabsi	260771940
Gayatri Krishnamoorthy	260975833

Friday, April 15th, 2022

Contents

1. Introduction

Racism can manifest itself in every element of our lives and society, including interpersonal contact, discriminatory policies and practices, prejudiced language, and laws and institutions. More specifically, the United States of America encounters a disproportionate criminal justice system, where from arrest to sentence to death row, Black and Latino males are heavily represented. According to a study conducted by Frank Edwards, of Rutgers University's School of Criminal Justice, African Americans are 2.5 times more likely than whites to be slain by police (Edwards et al., 2019).

Black Lives Matter (BLM) is a movement founded by three Black women in 2013 - Alicia Garza, Patrisse Cullors, and Opal Tometi. The movement began with the hashtag #BlackLivesMatter after George Zimmerman, who was acquitted in 2012 in the shooting death of Trayvon Martin, took center stage on social media. The movement gained national recognition in 2014 after the deaths of Michael Brown in Missouri and Eric Garner in New York. This period also gave rise to the #BlueLivesMatter in response to the BLM movement after the jury decided not to indict NYPD officer Daniel Pantaleo, who was responsible for Eric's death. This countermovement argued that those convicted of killing law enforcement officers should be sentenced under hate crime laws.

Following the death of George Floyd by the Minneapolis police in May 2020, BLM emerged as a worldwide movement. In recent times, BLM has spearheaded demonstrations worldwide in opposition to police brutality and systematic racism that disproportionately affect black people.

It opened the discussion around race to a vast extent, changed the nation's view on race and its underlying implications. The world is now more aware of the challenges faced by them in society and more welcoming to cater to their rights.

The aim of our project is to examine people's stance on these movements, and how it varied across time. Using twitter data and snscrape, we scraped a total of 30,000 tweets for #BlackLivesMatter and #BlueLivesMatter for 3 key time periods based on Google trends: pre-event (April 24th, 2020 – May 24th, 2020), during event (May 25th, 2020 - June 8th, 2020), and post-event (June 9th, 2020 - June 22nd, 2020). Please refer to "scraper.ipynb".

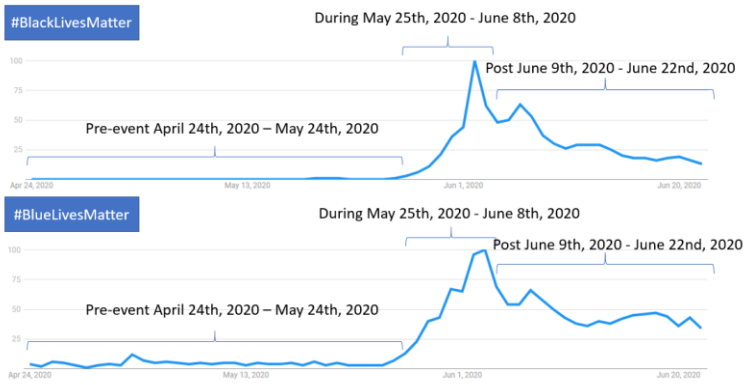

Figure 1 Google trend for #BlackLivesMatter and #BlueLivesMatter

For our analyses, we have planned the following:

- Determine the most influential people behind these two movements and their popularity amongst the public with network analysis.
- Study the overall sentiment for each of the hashtags and discover words that are frequently associated with them.
- Examine their opinion, the nature of speech with toxicity analysis to get a better understanding about how strongly they feel towards these causes.
- Compare the impacts of both movements and uncover important insights based on our key findings.

2. Network Analysis

In order to identify who the biggest influencers were between #BlackLivesMatter and #BlueLivesMatter, we carried out a network analysis on for each time period for each hashtag and used three metrics: betweenness, degree, and pagerank. Degree looks at which people are the most connected within the network. Betweenness focuses on the number of paths a person may be on. Lastly, pagerank is a measure of how important a person may be. To visualize the networks, we used NodeXL. Additional records were scraped for users that appeared within the original scraped datasets in order to ensure that all information was captured for the network analysis (please refer to "Network data.ipynb". When visualizing the networks, we only included non-tweets to help us focus on the connections that occurred throughout each time period.

2.1 Black Lives Matter

Before the George Floyd event, #BlackLivesMatter was still a prominent hashtag on Twitter. Many of the discussions revolved around Breonna Taylor, who was shot and killed by police in March 2020 in her own home during a drug investigation operation. Some notable figures who appeared within the network are celebrities who were for the movement like Beyonce, Diddy, and Drake, various American political figures like Joe Biden (pro BLM), Donald Trump (pro law enforcement), AOC (pro BLM), and Bernie Sanders (pro BLM), as well as journalist organizations like NBC, Fox (pro law enforcement), and ABC. There were various groups/cliques that were apparent, more prominently accounts that focus on the BLM movement, such as the official page for the movement, Blklivesmatter. Interestingly, Fedex appeared to be prominent within the network, mostly because news sources at this time reported that Fedex had fired two African American employees for filming a racist customer verbally abusing them, further sparking outrage online (Price, 2020). Please refer to the nodexl file "Blacklivesmatter_pre_nodexl.xlsx".

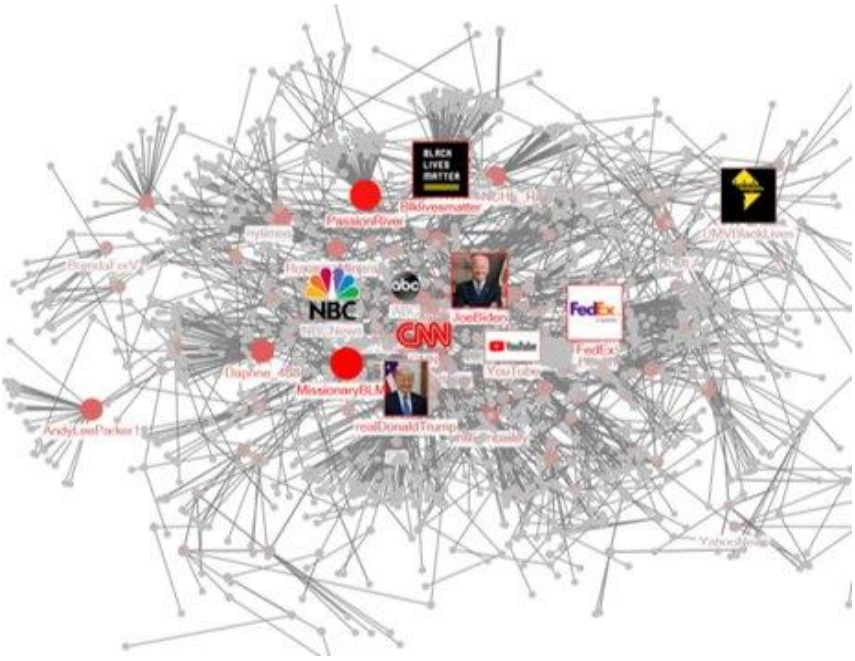

Figure 2 Network for #BlackLivesMatter Prior to George Floyd's Death

Shifting over to the network during the few days after George Floyd's death, the hashtag inevitably exploded on a global scale. Notable politicians like Donald Trump (pro law enforcement), Mitt Romney (pro BLM), and Barack Obama (pro BLM) appeared in the network, along with many celebrities who were for the movement like Oprah, Tyler Perry, and Taylor Swift. Many different

cliques and branches of discussions had formed, most interestingly that of Korean pop group BTS (BTS_twt), whose own fan accounts continued to drive the conversation with their own audiences (onlyBTS1030, BTS_ARMY_NL_). This is mostly due to the band having expressed their support for the movement and donating $1M to the Black Lives Matter efforts (BBC News, 2020). Please refer to the nodexl file “Blacklivesmatter_during_nodexl.xlsx”.

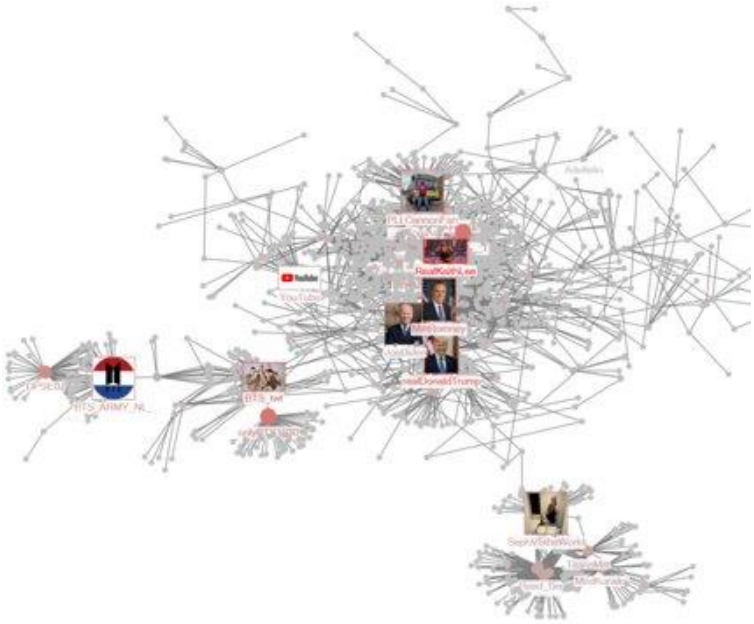

Figure 3 Network for #BlackLivesMatter During the Days After George Floyd's Death

The conversations continued during the month after George Floyd’s death, more specifically surrounding discussions on prominent athletes like Lewis Hamilton and Romelu Lukaku. For Lewis, tweets discussed how he was the only F1 driver to speak out publicly about the event, further sparking discussions about the F1 industry and their stance on the issue. For Romelu, he had taken a knee and a fist to the air after scoring a goal in a soccer match in support of the Black Lives Matter movement, which sparked conversations (XtraTime, 2020). Please refer to the nodexl file “Blacklivesmatter_post_nodexl.xlsx”.

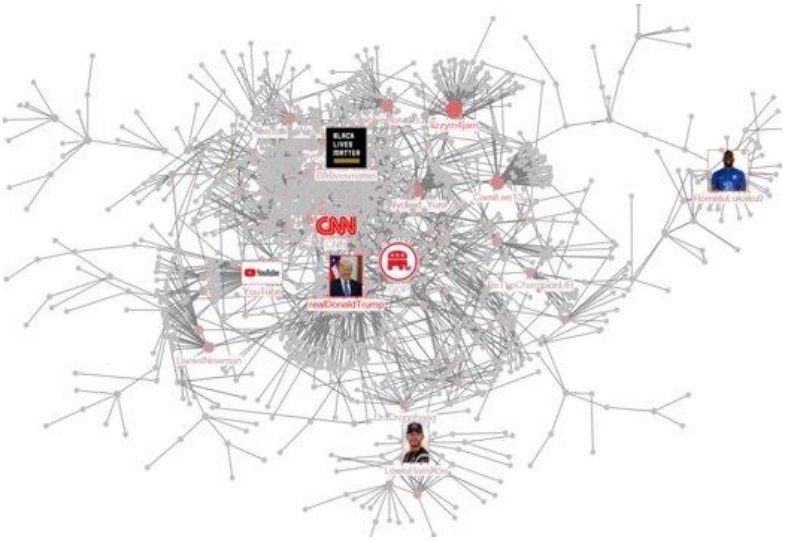

Figure 4 Network for #BlackLivesMatter During the Month After George Floyd's Death

When looking at the network metrics for the BLM movement, we can see a shift over time of who was most influential. Prior to the event, MissionaryBLMM had the most connections and highest importance. They were then replaced by RealKeithLee, who is an African American WWE fighter. This was because he had sparked conversation online after showing his support for the movement by wearing BLM pants to a fight. Unsurprisingly, Donald Trump became the biggest influencer post George Floyd's death, considering that he was the president at the time. What sparked conversations online were his own controversial tweets and retweets, inaction towards reforming current law enforcement policies, and even threatening to bring in the military to stop the protests.

	Pre	During	Post
Highest Degree	**MissionaryBLMM**	**RealKeithLee**	**realDonaldTrump**
Highest Betweenness	**JoeBiden**	**realDonaldTrump**	**realDonaldTrump**
Highest Pagerank	**MissionaryBLMM**	**RealKeithLee**	**realDonaldTrump**

Table 1 Network Metrics to Identify Top Influencers for #BlackLivesMatter

2.2 Blue Lives Matter

Blue lives matter is a countermovement to Black Lives Matter. The movement gained support in 2014 after the death of NYPD on-duty officers Rafael Ramos and Wenjian Liu. The Blue lives matter movement advocates for people attacking law enforcement to be charged with hate crimes and supports families of law enforcement officers in need. A network analysis was performed to understand who the main influencers were and how the network evolved with time starting with pre-blue lives matter, during blue lives matter, and post blue lives matter.

In pre-blue lives matter, the network analysis showed the main influencers were politicians President Donald Trump, Democrat Strategist Adam Parkhomenko, a non-profit organization "bluelivesmatternyc", far-right syndicated news website "Breitbart", and far-right militia group Oath Keepers and its founder Stewart Rhodes. In the early stage of the pre-blue lives matter, the analysis showed far-right advocates were the main players behind this movement, and these players were more likely using this movement to push their political agenda disguised as pro- blue lives matter. Moreover, influencers like "bluelivesmatternyc" with legitimate cause to the movement had a lesser role in the network. Please refer to the nodexl file "Bluelivesmatter_pre_nodexl.xlsx".

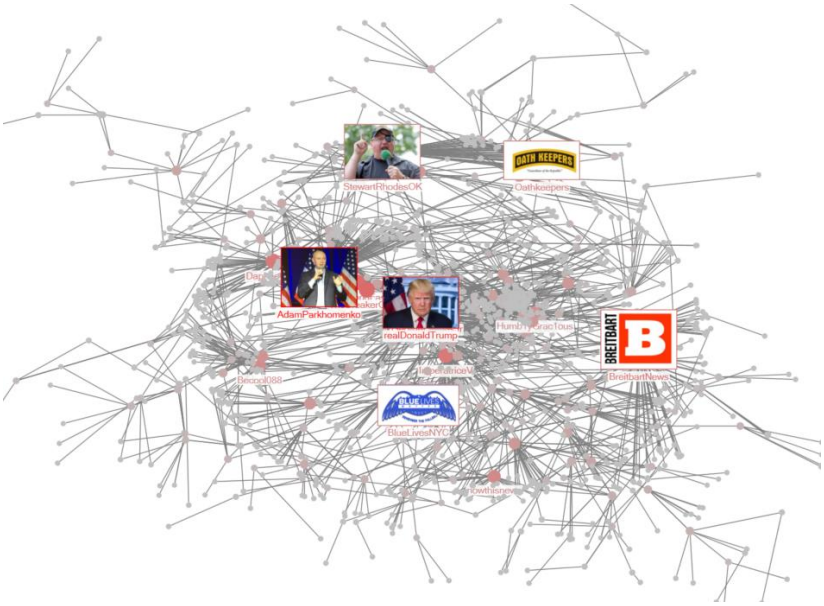

Figure 4: Network Prior to #BlueLivesMatter

During blue lives matter, the network showed Donald Trump was the main focal point in the network and he was surrounded by lesser-known influencers. These influencers were highly

connected and displayed a similar pattern in the network, which prompted further investigation to determine if they were fake or bot accounts. To better understand these accounts, a community analysis was performed at a later stage. Please refer to the nodexl file "Bluelivesmatter_during_nodexl.xlsx".

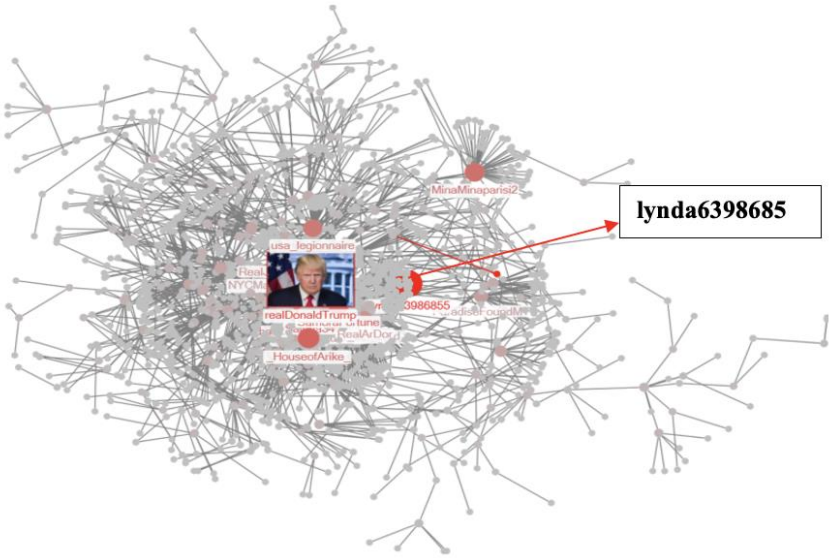

Figure 5: Network During #BlueLivesMatter

In post blue lives matter, the network analysis displayed Atlanta and Tulsa Police departments as top influencers, which revealed broader adoption of the movement across different police departments. Additionally, far-right activists/groups like Joey Poynter, and bugoffdear were also picked up by the network. The movement provided a platform for far-right movement to push their agenda, and these groups had diluted the real cause of the movements, which was to show solidarity to law enforcement and support families of officers' needs. Instead, the far-right influencers in the network from before post-movement were creating a source of division, and distraction from the real cause of the movement. Please refer to the nodexl file "Bluelivesmatter_post_nodexl.xlsx".

Figure 6: Network Post #BlueLivesMatter

	Pre	During	Post
Highest Degree	**AdamParkhomenko**	**Lynda63986855**	**joey_poynter**
Highest Betweenness	**realDonaldTrump**	**realDonaldTrump**	**realDonaldTrump**
HighestPagerank	**AdamParkhomenko**	**Lynda63986855**	**joey_poynter**

Table 2 Network Metrics to Identify Top Influencers for #BlueLivesMatter

Community detection analysis was performed during blue lives matter movement to better understand the communities located within the network using the Girvan Newman clustering algorithm. Aside from Donald Trump, the clustering output showed 3 different communities consisting of liberal voices like Former NYC Major Bill DeBlasio, conservative Fox news anchors (Sean Hannity, Jesse Watters, Laura Ingraham), and communities consisting of questionable accounts (lynda6398685, houseofArike, minaminaparisi2, usa_legionnaire, simorafortune). When examining these questionable accounts, they exhibited similar stances and behavior in the network, and almost all information was being relayed either to Trump or Fox News communities. There was consorted effort to transfer information into these communities. This is indicative of fake activity, and the network did provide us with insight into who might be potential bot accounts.

To confirm these findings, an open-source platform for bot detection was utilized (www.botsentinel.com) and the result revealed the questionable accounts were indeed labeled as being a bot with problematic/ disruptive activities on twitter.

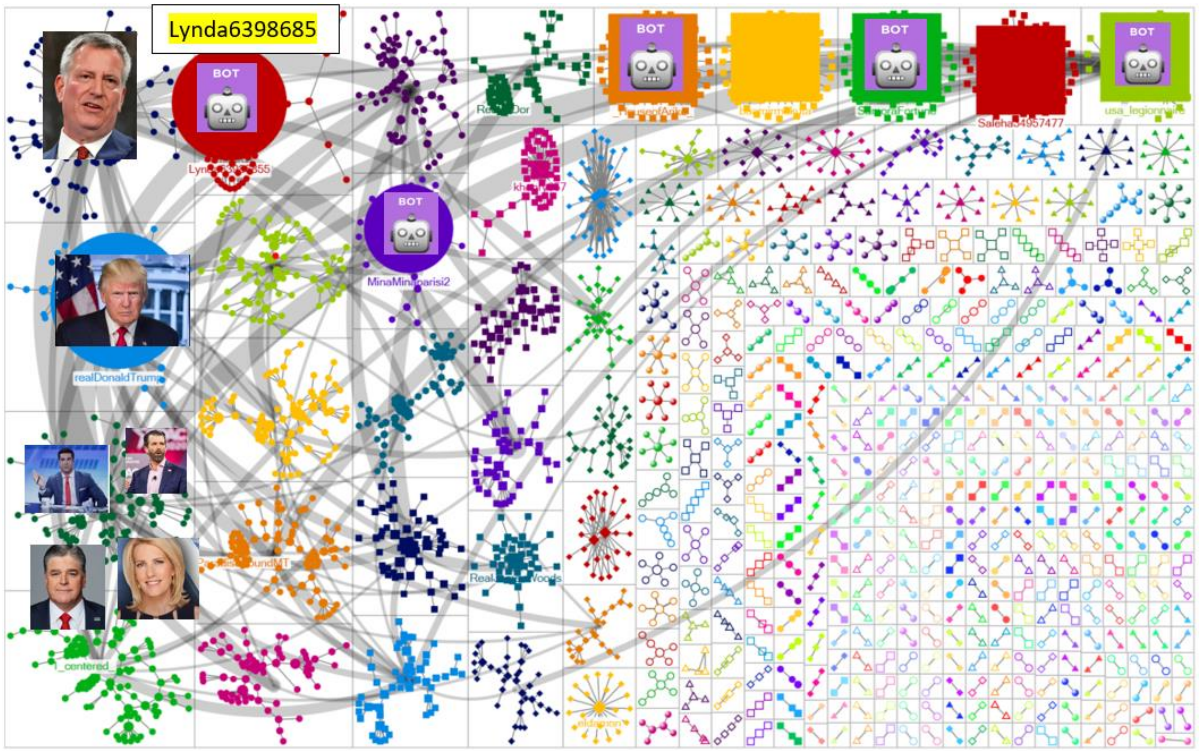

Figure 7: Network Post #BlueLivesMatter

3. Toxicity Analysis

After performing the network analyses, we sought to better understand the dynamic between both movements, as these are opposite movements. More specifically, we wanted to determine which movement had the most hate speech and evaluate both network's toxicity overtime (pre-event, during, post-event). We decided to perform a toxicity analysis of the networks. This was done first through calculating an overall toxicity score using Google's Perspective API. Then, to refine our analysis, we dove deep into understanding which movement used the most swear words in their speech using statistical inference to validate our results.

3.1 Data Pre-processing

To increase the reliability of our analyses, we had to perform a lot of data pre-processing on both sets of tweets (Black Lives Matter and Blue Lives Matter). We had to remove emojis –using

python's emoji library–, URLs, hashtags, numbers, punctuations, leading and trailing whitespaces and substitute multiple whitespaces by a single whitespace. Furthermore, there were a few tweets that were not in English, these were identified using python's langdetect library and removed from the dataset. After selecting English tweets, english stopwords were removed. Please refer to the jupyter notebook files "#blacklivesmatter.ipynb" and "#bluelivesmatter.ipynb".

3.2 Toxicity Analysis - Perceptive API

After pre-processing the data, we ran a for loop to get a toxicity score for each tweet in our dataset using the Perspective API. In fact, the perspective API predicts the perceived impact that a comment may have on a conversation by evaluating it across a range of emotional concepts. We decided to use the "TOXICITY" concept which provides a measure of the probability that a comment is rude, disrespectful, unreasonable, or more likely to make people leave a discussion. This means that the toxicity score does not only look at curse words but also other words that are inherently violent such as "kill" or "murder" for example.

The toxicity scores were then averaged over the three periods for both movements to get an overall toxicity score per period per movement. The following bar chart summarizes our findings.

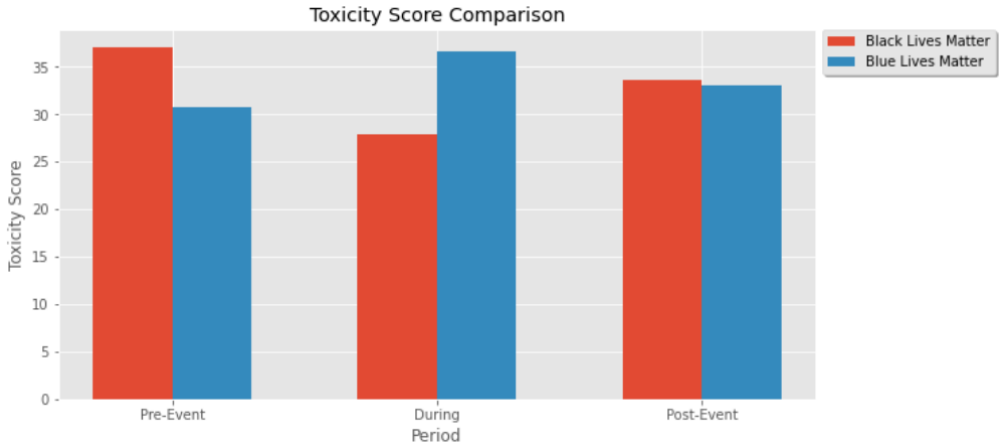

Figure 8: Toxicity score per period per movement

	#BlackLivesMatter			#BlueLivesMatter		
Period	Pre-Event	During	Post-Event	Pre-Event	During	Post-Event
Score	37.07	27.96	33.63	30.82	36.63	33.12
Growth		-25%	20%		19%	-10%

Figure 9: Evolution of toxicity score per movement throughout time

The results of this analysis were remarkably interesting. In fact, we can interpret our results as such:

1. **Pre-event period:**

   The Black Lives Matter movement had on average a higher toxicity score than the Blue Lives Matter Movement in the pre-event period. This might be due to the fact that Black Lives Matter supporters were already condemning crimes against black folks before the murder of George Floyd. In particular, there was a lot of conversation around the murder of Ahmaud Arbery, Breonna Taylor and Stephon Clark that might explain why the toxicity score of the Black Lives Matter movement was higher than the Blue Live Matter movement's. Furthermore, the Blue Lives Matter movement was less active than the Black Lives Matter movement before the protests that followed George Floyd's murder which might also explain why their score was lower.

2. **During period:**

   The Black Lives Matter movement's toxicity score decreased by 25% during the George Floyd protests. In fact, during that time, Black Lives Matter supporters were seeking solidarity and raising awareness to their cause more than they wanted to confront the Police. In fact, there was a lot of conversation about police brutality victims through the #SayTheirNames hashtag for example, but there was a lot of conversation about organizing protests, supporting protestors who faced police brutality and black-owned businesses as well. This attitude of Black Lives Matter protestors also shone at the George Floyd's protests as they were peaceful in nature. On the other hand, the Blue Lives Matter's toxicity score increased by 20%. The Blue Lives Matter supporters were most probably threatened by the condemnations of the Black Lives Matter protestors as they were protesting the current U.S. police system.

3. **Post-event period:**

   In response to the toxicity of the Blue Lives Matter movement, the Black Lives Matter movement reciprocated. In fact, the toxicity score of the latter movement increased by 19% until matching that of the Blue Live Matter movement's.

Through this analysis, the intricate relationship between both movements is clearly apparent and we can distinctly understand how both movements influenced each other.

3.3 Swear Word Analysis - Hypothesis testing

To deepen our analysis, we wanted to know exactly which movement used the most curse words. This was done using python's ProfanityFilter library and two-sample statistical inference. Please see jupyter notebook file called "Comparison Analysis.ipynb".

In fact, we calculated the proportion of tweets that contained at least one swear word for both movements. We then performed a one-sided statistical inference test for two proportions using python's statsmodels package.

In particular, we performed the following test:

$$H_o: \ p_{blueLM} - p_{blackLM} \leq 0$$

$$H_a: \ p_{blueLM} - p_{blackLM} > 0$$

The proportions were calculated to be:

$$p_{blueLM} = 6.48\%$$

$$p_{blackLM} = 5.43\%$$

The test came out statistically significant at the 5% confidence level with a p-value of 0.00168%.

To understand whether both movements curse differently we have also created a frequency graph of the top 10 swear words for both movements.

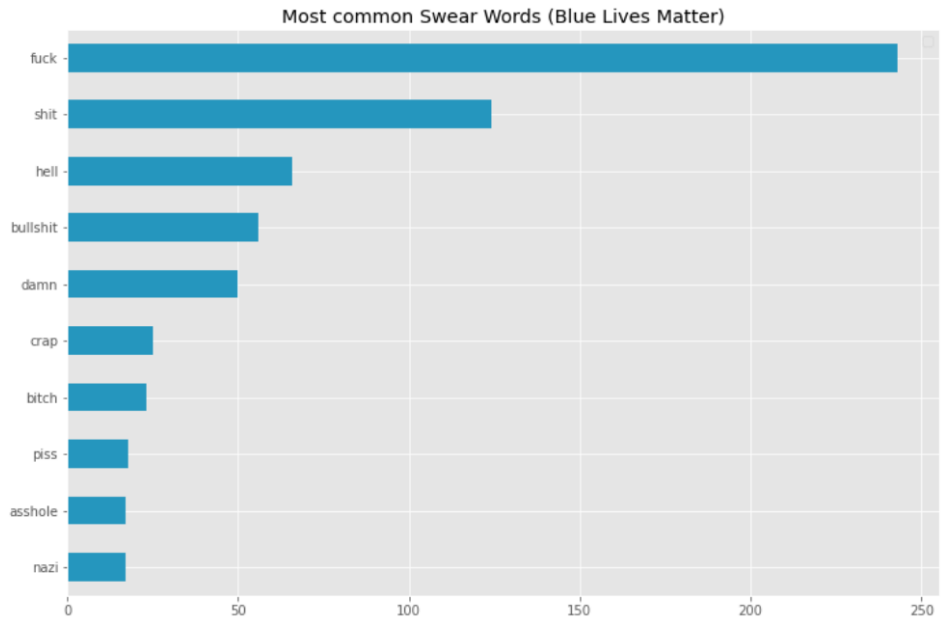

Figure 10: Top 10 swear words for the Blue Lives Matter movement

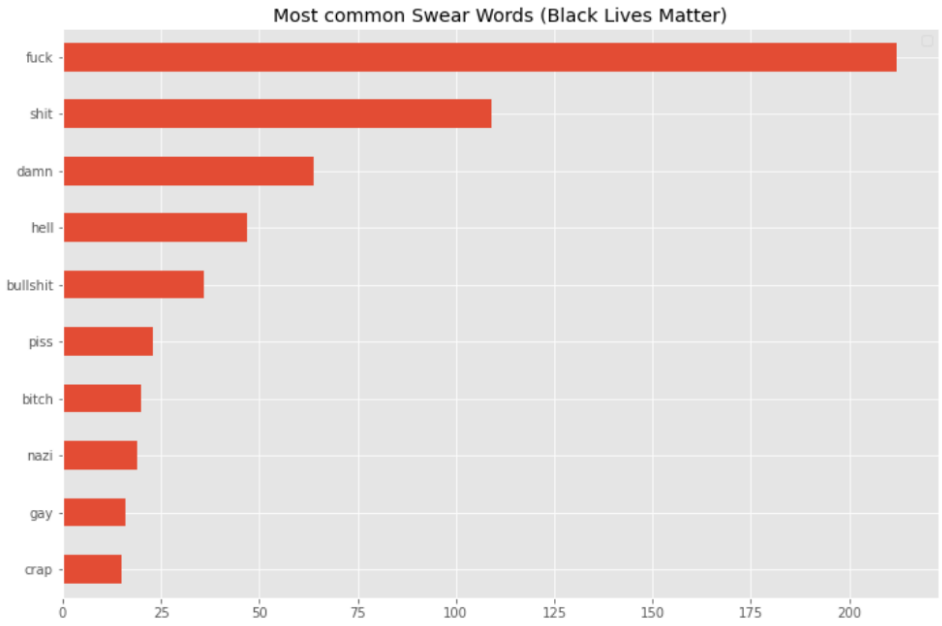

Figure 11: Top 10 swear words for the Black Lives Matter movement

As we can see both movements use the same curse words. It is also very interesting to note that the n-word does not appear at the top of the list in both cases.

To conclude this analysis, both movements had times when they used more toxic language than the other depending on the context and time period at hand. However, overall, we have proven with a confidence level of 95%, that the blue lives matter movement uses more cuss words and is thus overall, more toxic.

4. Conclusion

By using Twitter data with snscrape, we were able to find interesting insights with 30,000 tweets and identify the most influential people through network analysis on both Black Lives Matter and Blue Lives Matter movements. Having noticed irregularities on the later, we analyzed Blue Lives Matter tweets with community detection and discovered several bots that were confirmed using BotSentinel. Moreover, we examined the nature of speech with a toxicity analysis of each movement. The objective was to get a better understanding about how aggressive the speech is towards these causes. We noticed a trend between both movements, in which the effect of the death of George Floyd attenuated the Black Lives Matter speech and Blue Lives Matter felt threatened by using more hate speech. We also performed statistical inference to conclude that Blue Lives Matter speech swear more than Black Lives Matter at a 5% level. We then proceeded, to showcase slur words that are frequently associated with each movement and compare results.

To conclude, we believe that this extensive analysis could be further improved by increasing the scraping threshold to be above 5,000 tweets per period and hashtag, as 30,000 tweets is not large enough for such a topic. In addition, another layer of segmentation could be implemented to separate pro-tweets and against tweets. By adding this layer, we could add more interpretability to our analysis. Finally, we believe it could be interesting to compare these network analyses with other movements such as Capitol Invasion.

Works Cited

BBC News. “BTS Black Lives Matter: Fans Match Band’s $1m Donation.” BBC News, 8 June 2020, www.bbc.com/news/world-asia-52960617.

Edwards, Frank, et al. “Risk of Being Killed by Police Use of Force in the United States by Age, Race–Ethnicity, and Sex.” Proceedings of the National Academy of Sciences, vol. 116, no. 34, 2019, pp. 16793–98. Crossref, https://doi.org/10.1073/pnas.1821204116.

Price, Joe. “FedEx Shares Statement After Workers Claim They Were Fired for Filming Racist Customer.” Complex, 14 Oct. 2020, www.complex.com/life/2020/05/fedex-releases-statement-firing-workers-viral-video-of-racist-customer.

XtraTime. “Serie A: Romelu Lukaku Takes Knee to Show Support towards Black Lives Matter Movement | XtraTime | To Get the Best and Exclusive Sporting News, Keep Watching XtraTime.” Xtratime, 22 June 2020, xtratime.in/serie-e-romelu-lukaku-takes-knee-to-show-support-towards-black-lives-matter-movement.