An Exploration of Methods Using Social Media to Examine Local Attitudes Towards Mask-Wearing During a Pandemic

Nathan J. Elrod
University of Cincinnati
elrodnj@ucmail.uc.edu

Pranav Mahajan
University of Cincinnati
mahajapp@mail.uc.edu

Monica Katragadda
University of Cincinnati
katragma@mail.uc.edu

Shane Halse
University of Cincinnati
halsese@ucmail.uc.edu

Jess Kropczynski
University of Cincinnati
kropczjn@ucmail.uc.edu

ABSTRACT

During the COVID-19 health crisis, local public officials continue to expend considerable energy encouraging citizens to comply with prevention measures in order to reduce the spread of infection. During the pandemic, mask-wearing has been accepted among health officials as a simple preventative measure; however, some local areas have been more likely to comply than others. This paper explores methods to better understand local attitudes towards mask-wearing as a tool for public health officials' situational awareness when preparing public messaging campaigns. This exploration compares three methods to explore local attitudes: sentiment analysis, n-grams, and hashtags. We also explore hashtag co-occurrence networks as a possible starting point to begin the filtering process. The results show that while sentiment analysis is quick and easy to employ, the results offer little insight into specific local attitudes towards mask-wearing, while examining hashtags and hashtag co-occurrence networks may be used a tool for a more robust understanding of local areas when attempting to gain situational awareness.

Keywords

Social media analytics, situational awareness, sentiment analysis, N-grams, social network analysis.

INTRODUCTION

In 2020, the coronavirus disease 2019 (COVID-19) was the cause of a global public health crisis. Even with evidence that wearing a mask can reduce the spread of the virus (Prather et al., 2020), doing so became a controversial and politicized action in the United States. As a result, the American Medical Association (AMA), American Hospital Administration (AHA), and American Nurses Association (ANA) teamed up to encourage nationwide mask-wearing, and also to dispute misinformation that discouraged their use (A.M. Association, 2020). Public messaging campaigns were bolstered at all levels by government agencies and public health departments, while additional marketing efforts were used to distribute messaging in areas that were hardest hit by the virus (I.D.S. of America, 2020). Public messaging campaigns were often tailored to local areas' needs, but it is not always obvious what the predominant attitude of particular community might be. This paper explores the merits of several commonly used social media analysis techniques and evaluates their merits with regard to helping local officials understand local attitudes towards mask-wearing while planning targeted messaging campaigns.

The ISCRAM research community has a history of using social media analytic tools to meet the needs of practitioners responding to crises (Imran et al., 2013; Hiltz and Plotnik, 2013). These tools aim to extract relevant information from social media posts to inform practitioner decision-making; research teams have devised various strategies that help with different types of situational information needs. This study explores the use of commonly employed social media labeling methods including sentiment analysis, n-grams, and hashtags as they apply to understanding citizens local attitudes towards mask-wearing. We pose the following overarching research questions to guide this work:

RQ1: *What are the descriptive characteristics of local data analyzed using three labeling approaches (sentiment analysis, n-gram analysis, and hashtag analysis)?*

RQ2: *How do these three labeling approaches compare to manual coding when used as methods to understand local attitude towards mask-wearing?*

The first of these research questions is aimed at understanding the analytic value of each approach, while the second aims to assess the utility of each approach in accurately identifying the attitude towards mask-wearing expressed in a post. To compare these methods, Twitter data were collected from a geographic area corresponding to a local US county between the dates of 7/16/2020 to 8/22/2020. These were first analysed through a popular Python sentiment analysis package called VADER. Next, 1-, 2-, and 3-word token combinations (also known as *n-grams*) were extracted from a mask-related subset using another popular Python package called *NLTK* and evaluated for apparent mask-wearing attitude. Finally, hashtags relating to pro-mask and anti-mask attitudes were identified as a method to label Tweets. All methods were compared to mask-wearing attitude in Tweets as assigned by human coders.

We found that both sentiment analysis and n-gram filtering yielded low (<50%) consistency in identifying Tweets that represent either pro-mask or anti-mask attitudes; however, hashtags more consistently identified (>90%) local attitudes towards masks. We further utilized hashtag co-occurrence networks as a method to understand the landscape of nuanced local attitudes that exist within a region to provide initial hashtag starting points to explore attitudes on social media.

This research contributes to ongoing literature that uses social media aggregation tools for local situational awareness as it applies to the context of the current global pandemic. The process of analyzing these data in this context contributes to the development of a new analytic tool that is specific to the diffusion of public attitude through Twitter posts. We conclude by discussing how these methods can be used in the design of new social media analysis tools used to understand local attitudes towards masks, other COVID-19 preventative measures, and additional use cases.

BACKGROUND

Situational Awareness in Times of Crisis

Crisis informatics examines the social, information, and technical aspects of a crisis. A genre of technologies called "decision support systems" have been successfully used to help in this regard when trying to predict the scale of an event (Chatfield and Brajawidagda, 2012), assist in knowledge distribution (Van de Walle and Turoff, 2008), detect and assess damage (Avvenuti et al., 2014), and facilitate courses of action (Mendonca et al., 2001). The objective of many crisis informatics tools is to provide situational awareness for crisis responders. *Situational awareness* is a concept that was first introduced by Endsley (1995) after observing pilots orienting to complex situations in short periods of time; simply defined, it is a comprehensive understanding of an ongoing situation that can be based on spatial, geographic, and temporal data. Hiltz et al. (2013) extended Endsley's situational awareness concept to the field of crisis informatics, emphasizing the importance of having a dynamic model when trying to gain understanding of unpredictable phenomena during a crisis. As a source of near real-time updates of information (Imran et al., 2013; Hiltz and Plotnik, 2013), social media data aggregation, labeling, and analysis have become commonly explored methods to support situational awareness by the ISCRAM community.

Social media is a data rich, but chaotic information environment, therefore crisis informatics researchers wishing to employ this data have developed a variety of techniques to filter for information most relevant to decision-making. We review some common methods in the following sections.

Sentiment Analysis

Sentiment analysis has been one of the most active research areas in natural language processing in the last 20 years (Liu, 2012). The aim of sentiment analysis is to use automatic tools to extract this subjective information from natural language text, using this information to create structured and actionable knowledge to be used by either a decision support system or a decision-maker.

Sentiment analysis has been used successfully in crisis informatics when examining how the polarity of a message can contribute to the situational awareness of the crisis. For example, Nepalli et al. (2017) explored how polarity of social media messages was correlated with location and distance from a crisis event, and Halse et al. (2018) found that both the sentiment and the emotion of a message affected its perceived trust and actionability. Further works have examined how sentiment analysis can be used to measure the public's response to an organization's message (Singh et al., 2018), enabling information to be leveraged to adjust and customize messaging hyperlocally.

Although many researchers have found relative success using these tools, prior studies have illustrated accuracy issues with generalized sentiment analysis algorithms and social media. For example, Hassan et al. (2013) found poor performance when evaluating seven publicly available sentiment analysis tools against three large twitter datasets, citing an imbalance in neutral Tweet sentiment and "feature sparsity" of Tweets due to character limits and limited sentiment cues. Mishra et al. extend the definition of sentiment in the context of social media to the full typology of emotion, including general mood, attitude, or overarching personality traits as well (Mishra et al., 2017). This added complexity in social media expression can create a problem for automated sentiment analysis techniques—as the authors note, while many online platforms allow for the express indication of sentiment in a post (e.g., a product review), researchers evaluating social media are required to mine the textual content of a Tweet for signals indicating sentiment. Given varied results with this labeling tool, our analysis examines the relative merits of other Tweet labeling techniques in addition to sentiment analysis.

N-grams

N-grams, sometimes referred to as a "shingles" or "tokens", are truncated word combinations that are *n*-length sequence of words. For example, in the following sentence: *"Please wear your mask."* we can extract the following two-word n-grams (also known as *bigrams*): ("please","wear"), ("wear","your"), ("your","mask").

N-grams have been used for web-based scientific research for nearly a quarter century (Broder et al. 1997) and have been a standard in language analysis for over 50 years (Damerau, 1966). Collections of commonly occurring n-grams can be derived from a particular body of text to form a corpus for various methods of statistical analysis (Rabiner and Juang, 1993; Brown et al., 2001). By focusing on key recurring word groups, which are presumably more likely to contain greater context than single words alone, training data can be made for various statistical, machine learning and AI techniques. N-gram analysis has assisted crisis researchers in ways as varied as event detection (Toosinezhad et al., 2019), classification (Nguyen et al., 2017), enhancing geolocation capability (Middleton et al.), and—very recently—in gauging where public attention is being directed in the midst of an ongoing global crisis (Alshaabi et al., 2021).

Given the established use of n-grams in both general language analysis and the domain of crisis response, we extend its use to explore if Twitter n-gram sequences might possess clear meaning with regard to mask-wearing attitude. If it is possible to link attitude with n-gram combinations, this will pave the way for context-independent and potentially less computationally costly methods for identifying mask-wearing sentiment across the broader corpus of Tweets.

Hashtags

Hashtags are an omnipresent feature of modern social media communication, used extensively on Twitter and across all other major social media platforms, and indeed have proliferated through other forms of web related and non-web related media as well. First introduced to the Twitter platform in 2007 (Structured journalism, 2014), it was not long before researchers noticed their communicative value and began systematically analyzing hashtag metrics (Bruns et al., 2012). During this time hashtags have evolved their own sense of conversational logic, and can be employed by users in a number of ways. They can, for instance, be used to identify the "semantic domain" of a post, link it to a collective ideal, or be used to reinforce the emotional content of the message (Zappavigna, 2015).

The language structure of Twitter posts tends to be very disorganized; of the structure that does exits, some is explicit, as in the form of structured Twitter JSON data, or the presence of geolocation coordinates. Other structure is less obvious, such as the inferred geolocation of a Tweet, or semantic cues as to a Tweet's relevance towards a particular topic. Zappavigna (2015) observes that in addition to whatever formal structure or linguistic/semantic value they may possess, hashtags may also serve as a method of markup or what she terms "social metadata". Indeed, as Zappavigna points out, the presence of the hashtag is a form of markup structurally similar to those found in other markup languages such as XML and HTML; meaning users are quite literally organizing their own content.

In 2009, hashtags received the now indispensable feature of being hyperlinked on Twitter (Rao, 2009). While metadata has existed in information management for quite some time (Zappavigna, 2015), the hyperlinking of the hashtag in particular allowed for information organization across the spectrum of online social discourse, not only internal to a particular social media platform and eventually between platforms, but into the wider culture beyond, creating a multidimensional scaffolding for an otherwise extremely chaotic mode of personal communication (Van den Berg, 2014) The tendency for Hashtags to self-organize by co-occurrence has been exploited by a number of researchers. Wilson (2018) lays out a simple strategy for using hashtag co-occurrence graphs to help evaluate semantic networks, and Lee et al. (2018) developed a method to use hashtag co-occurrence strength ties in determining relevance to a particular root topic or line of reasoning. Hachaj and Ogiela (2017) found that in a business context, utilizing hashtags as opposed to natural-language word morphology was able to reveal far more structure among association for trending topics.

In crisis informatics research, the tendency for social self-organization and the network analysis of hashtag co-occurrences has been no less important. Wukich and Steinberg look at levels of community self-organization during crises involving social media (Wukich and Steinberg, 2013) of which hashtags play a crucial role in message proliferation, both locally and to wider jurisdictions. In 2012 the Australian Research Council released a report following the use of popular hashtags during the 2011 Queensland floods (Bruns and Stieglitz, 2013), and in 2019 Son et al. (2019) looked at how hashtags and other Tweet variables facilitate communication during the 2013 Colorado floods, in terms of Retweeting time. Taking all of this into account, in addition to looking at the relevance of particular keyword combinations in locating attitude towards mask-wearing, we explore the context-rich content of hashtags in our analysis. Our work will extend this previous work to briefly explore co-occurrence of anti-mask and pro-mask hashtags to better understand the diffusion of attitudes within these networks. The following methods section will describe how we have approached the use of the above social media analysis tools in our investigation.

METHODS

Data Collection Methods

Data collection for this study occurred between the dates of 7/16/2020 and 8/22/2020, the summer following the onset of the coronavirus pandemic in the United States. Using the free Twitter Developer API, the Python programming language and the *python-twitter* package, Tweets were collected from a particular region by means of a geographic query using zip code approximation (Elrod et al., 2021), the region corresponding to a single mid-western U.S. county that includes a moderately sized city. Data was aggregated by geography rather than keyword because not relying on Tweet selection by *a priori* keyword assumption yielded what we ascertain to be a full set of geographic Tweets from the time period, and allowed the freedom to apply post-collection text filtering as desired.

For sentiment and n-gram analysis, a subset of Tweets from our geo-located data was generated by filtering the dataset for content containing the words ‘mask’ and ‘face cover’.

Overall Study Design

Data were analyzed in the context of a concatenated study. In order to accurately assess Twitter users’ attitudes towards mask-wearing, a number of labeling techniques were explored: conventional sentiment analysis, n-gram evaluation, hashtag evaluation, and a brief qualitative analysis of a hashtag co-occurrence network, each compared against the analysis of mask-wearing attitude as arrived upon by manual coders. We describe each of these steps of the study design in further detail in the following subsections.

Sentiment Analysis

Valence Aware Dictionary for Sentiment Reasoning (VADER) is a text sentiment analysis tool available on the Python programming platform (Github). VADER purports to detect polarity of opinion expressed in a Tweet as well as the intensity of the emotion expressed, and provides output as a series of vector scores (Gilbert and Hutto, 2014). These scores show individual polarity values (positive, negative, and neutral) for each piece of text, ranging from 0 to 1, and also a composite score ranging -1 to 1 representing the overall sentiment of the text. Values below -0.05 are interpreted as having a negative sentiment and values above 0.05 are scored as positive; scores between -0.05 and 0.05 are considered neutral.

N-grams

To obtain n-grams from the dataset, Python and the *NLTK* natural language processing package were used. In particular we utilized the class *TweetTokenizer*, a word token extraction utility optimized specifically for Twitter content (Source Code, 2020). All Tweets from the previously described mask subset were compiled into a single block of line- separated text, filtered to remove punctuation, and converted to lower-case. *TweetTokenizer* was then used to extract individual word tokens from the text body, and the nltk *BigramCollocationFinder* and *TrigramCollocationFinder* packages used to assemble value counts of n-grams of these respective lengths.

The top 200 most frequently occurring n-grams from each category were selected to be manually scored as positive (+1), negative (-1), or neutral/not relevant (0) as indicating attitude towards mask-wearing. The mask Tweet subset was then looped back through programmatically using Python and the total n-gram mask-wearing attitude summed per each Tweet. If the net score was greater than zero, the Tweet was assessed as being pro-mask if and if below zero it was assessed as anti-mask; a score of zero was interpreted as neutral or not-related (Algorithm 1).

```
Data: mask_tweets, ngram_mask_attitudes
Result: tweet_attitude
for tweet ∈ mask_tweets do
    ngram_score = 0;
    for ngram ∈ tweet do
        if ngram ∈ ngram_mask_attitudes then
            get ngram_attitude;
            ngram_score + = ngram_attitude;
        end
    end
    if ngram_score ≥ 1 then
        tweet_attitude = positive;
    else if ngram_score≤  1 then
        tweet_attitude = negative;
    else if ngram_score == 0 then
        tweet_attitude = neutral;
end
```

Algorithm 1: N-gram evaluation algorithm pseudocode

Hashtags

Using Python and the *Pandas* package the full Tweet set was cycled through, and all unique hashtags containing the substring "mask" were extracted from the *entities_hashtags* tag of the JSON data. All of these hashtags were assigned a value of +1, -1 or 0 (pro-mask, anti-mask, or neutral/not-related) to represent their respective attitude towards mask-wearing. Using the same algorithm as the n-gram analysis, the sum hashtag attitude values were calculated per Tweet and interpreted as an overall attitude.

Manual coding

To determine actual Tweet sentiment and attitude towards mask-wearing among Tweets, an inter-coder reliability check was performed. Inter-coder reliability refers to the extent to which two or more independent manual coders agree on the coding of the content of interest with an application of the same coding scheme (Lavrakas, 2008). Coders analyzed each Tweet individually to determine the sentiment associated within the Tweet, classified as pro-mask, anti-mask, or neutral/unrelated. The coders then established consensus for each Tweet where there was disagreement. This labeled data set was used as the point of comparison to the results of the various aforementioned techniques in order to establish labeling accuracy of automated methods.

Hashtag Co-occurrence Network

As a hashtag can serve not only as a vehicle for sentiment or meaning but to link a Tweet to a broader community of thought, we explored whether hashtag network forces might also assist in the assessment of mask-wearing attitude. In particular, we look at hashtag co-occurrence in terms of semantic networks. We did so by generating a bipartite graph of Tweet-to-hashtag relations - a graph possessing two distinct classes (or modes) of vertices, one representing

Tweets and the other hashtags, the edges indicating the presence of a hashtag in a certain Tweet. This graph was also collapsed into a one-mode hashtag co-occurrence matrix (i.e. hashtag-by-hashtag, as opposed to hashtag-by-Tweet) in which vertices represent hashtags and the edges between two nodes one or more co-occurrences of that hashtag pair in a Tweet.

We explore these networks for additional heuristic cues as to a hashtag's actual mask-wearing attitude, and affordances to first responders and crisis researchers greater ability to fill in the blanks, in the absence of any further syntactic content.

RESULTS

We begin this section by reviewing the overall characteristics of our dataset and the presence of conversations focusing on masks. Then we review the type of descriptive data available by implementing each of the three labeling methods (RQ1) and also compare how each compares to manual approaches to understanding local attitudes towards masks (RQ2).

Overall Characteristics of Social Media Data Collected

The Tweet JSON files were collected and parsed in Python, with nested JSON data flattened into a one-dimensional dataframe using the Pandas package. 259,456 Tweets were retrieved in total, associated with 51 zip code geographies. Because a number of these geographies were overlapping, duplicate Tweets were present in the set; using the *ID* property unique Tweets were filtered out, resulting in 129,877 Tweets posted by 17,898 individual users. 17,149 (13.2%) of these Tweets were Retweets (see Table 1); Retweets were not removed, however, the rationale being that a Retweet can still be an expression of attitude towards mask-wearing, even if the that attitude is not original in content. The average user posted 7.26 times during this period, though this varied considerably (*SD* = 36.68), while the average 'Rewteeter' Retweeted 1.58 times with notably less variation (*SD* = 2.47). 18,328 Tweets (14.11%) contained hashtags. The mask filtered Tweet set resulted in a dataset of 1,331 Tweets by 786 users—1.02% of the total number Tweets, Tweeted by 4.39% of the users. 271 (20.36%) were Retweets. 194 (14.58%) of these Tweets contained hashtags.

Table 1. Unduplicated Tweets Descriptive Statistics for Sentiment Analysis and Manual Coding

	Unfiltered	Filtered for mask-wearing
Total Tweets	129,877	1,331
Unique Users	17,898	786
Mean Tweets/User	7.26 (*SD* = 36.68)	1.69 (*SD* = 2.4)
Total Retweets	17,149	271
Unique Users Retweeting	10,842	248
Mean Retweets/Retweeting user	1.58 (*SD* = 2.47)	1.09 (*SD* = .49)
Total Tweets Containing Hashtags	18,328	194

Sentiment Analysis

Descriptively, of the Tweets that were labeled by VADER in the mask subset, 36% were identified as positive, 27.5% were identified as negative, and 36.5% were neutral. Some examples:

Positive:

"Saw some of our students at curbside pick up today. It was great to see you and we are excited to get you back to school virtually and in person. Excellence in learning, leadership, innovation, and service! We are IHMS! #IHPromise #bravemasks"

Negative:

"@sassyfrassy25 Probably the same reason people don't wear masks...rules don't apply to them"

Neutral:

"E513_315E SenTedCruz senjudiciary UN Ambassador Power unmasked hundreds of people herself."

Consistent with Hassan et al. [18] the VADER sentiment analysis package evaluated only 112 of the 200 (56%) in agreement with the sentiment analysis conducted by manual coders. There was even less parity between the VADER general sentiment analysis score and attitude towards mask-wearing as determined by manual coders (i.e., negative sentiment corresponding to anti-mask attitudes, positive sentiment to pro-mask), with 41.5% correspondence between the two. Some examples of mask-wearing sentiment:

Pro-mask:

"School bullying needs to come back for the kids who refuse to wear masks"

Anti-mask:

"Mask of the day! If the pandemic was real this is more along the lines of what would be needed to stop the spread! Cloth and paper masks do Zero Ziltch Nada. #MasksDontWork #Scamdemic #perfectcrisis #GreatAwakening"

Neutral/unknown:

"Are they all wearing masks though?;"

Comparisons of the various automated analyses to methods of manual coding results are shown in Figure 1; in comparison to trigrams and hashtags, sentiment analysis was the least consistent with manual coding when it came to identifying pro-mask and anti-mask attitudes in a local area. We will further describe the n-gram and hashtag descriptives and comparative situational awareness affordances in the following subsections.

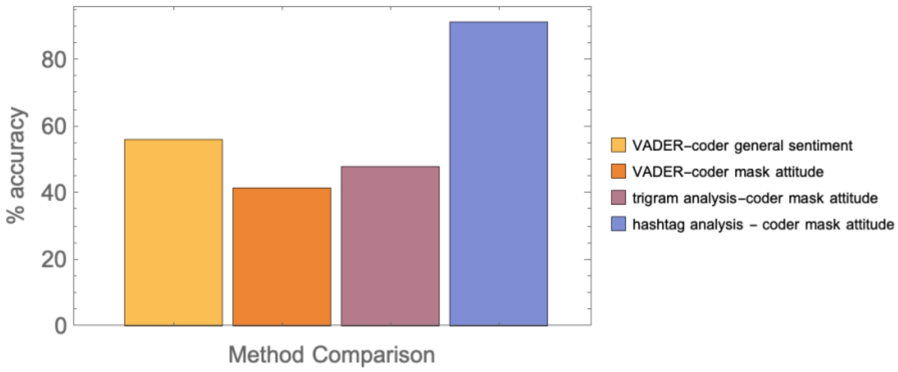

Figure 1. Comparisons of sentiment/mask-attitude accuracy to manual coding results

N-grams

After stripped for punctuation, the filtered dataset yielded 6,181 unique unigrams, 17,948 unique bigrams, 20,868 unique trigrams (Figure 2). Of these, each were filtered for the top 200 occurrences. In attempting to match uni- and bi-grams to local attitudes towards masks, most simply did not contain enough contextual information to draw a conclusion among coders as to their semantic meaning. Even trigrams - which we anticipated would more closely resemble common language structure - turned out to be largely equivocal in their interpretation, and though after a low initial agreement, a 100% inter-coder consensus was finally reached.

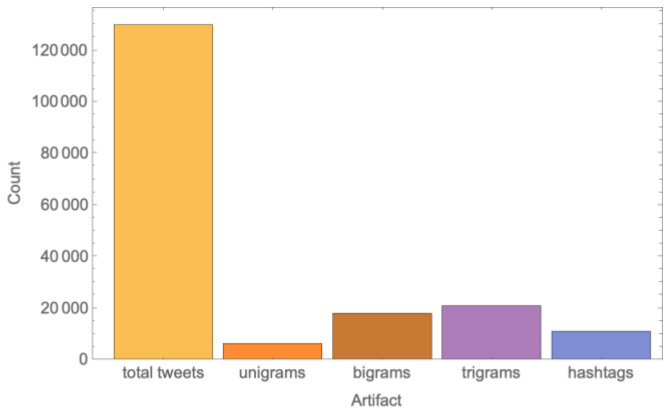

Figure 2. Comparison of Tweet Artifacts

For n-gram analysis comparsion to manual coding, 200 Tweets were randomly selected from the body of mask Tweets; by manual coder evaluation of trigrams we found with respect to attitude towards mask wearing, these Tweets were 59% pro-mask, 6% anti-mask, and 35% neutral or not-related. Trigram analysis fared only slightly better than VADER sentiment analysis, with 48% of trigram attitudes corresponding to the manually coded mask-wearing Tweets (see Figure 1; hashtags will be further explored in the next section). A comparison of labeling by pro-mask and anti-mask is shown in Table 2. While the three-word combinations initially appeared to possess a great deal of meaning, it appears that this may largely have been the coders imparting their own preconceptions on to the trigram—in isolation, for instance, the phrase "wear a mask" appears to be a call to action, we know nothing of the remaining content. The following example, for instance, is clearly a pro-mask Tweet:

"@WyomingSports: Student athletes- how many games we get to play depends a lot on the way you control what you can control. Wear a mask"

However, the same phrase can be found in the following Tweet that is not expressing a pro-mask attitude:

"@Local12 Just one step closer to locking people up in concentration camps ridiculous. They let people out of jails because of the virus why don't they make them wear masks if they are that effective instead of letting criminals out? Wear a mask or don't should be a choice"

Given these discrepancies, we further explore hashtag analysis and explore merits of each further in the discussion section.

Table 2. Manually Coded Attitudes Towards Masks

	N-gram Sample Tweets	Hashtag Sample Tweets
Pro-mask	118	123
Anti-mask	60	7
Neutral/Not Related	22	32

Hashtags

Overall descriptives of hashtags are shown in Table 3. The full Tweet set contained 34,225 hashtag occurrences in 17,804 Tweets, 10,930 hashtags which were unique. The mean number of hashtags per Tweet was .26 (*SD* = .95); of only hashtag-containing Tweets, this value increased to 1.92 hashtags per Tweet (*SD* = 1.85). The maximum number of hashtags in a single Tweet was 24. Individual hashtags occurred an average of 3.13 times (*SD* = 15.74), the maximum number of occurrences being 1,246.

Table 3. Overall Descriptives of Hashtags in Data Set

	All Hashtags	Mask Hashtags
Unique Hashtag Count	10,930	48
Total Hashtag Occurrences	34,225	184
Mean Hashtags per Tweet	.26 (*SD* = .95)	1.14 (*SD* = .45)
Mean Hashtags per Tweet (Hashtag-Containing Tweets Only)	1.92 (*SD* = 1.85)	1.14 (*SD* = .45)
Max Hashtags per Tweet	24	4
Mean Hashtag Occurrence	3.13 (*SD* = 15.74)	3.83 (*SD* = 8)
Max Hashtag Occurrence	1246	44

When filtered for hashtags containing the word 'mask', the Tweet set returned 48 unique hashtags (.44% of all unique hashtags), occurring 184 times across 162 Tweets. While the dataset yielded far fewer hashtags than n-grams, we found that the hashtags were far more likely to contain contextual information, with 26 (54.17%) containing what appeared to be clearly positive attitude towards masks, 6 (12.5%) containing what appeared to be clearly negative attitudes towards masks, and 16 (33.33%) being neutral or unknown.

Compared to the relatively low parity offered by the other methods, 149 of the 162 of the Tweets containing mask hashtags (91.98%) were evaluated correctly for mask-wearing attitude when compared to inter-coder consensus. Mask-wearing attitude in hashtags appears to be quite unequivocal; in fact in many cases, the hashtag is the sole bearer of information relating to attitude towards mask-wearing. Take the following Tweet:

"Had a blast at the @cincinnatizoo today! #maskup #staysafe #cincinnati #fatherhood #parenthood #dadsdontbabysit @Cincinnati Zoo & Botanical Garden https://t.co/frXXvlIs3V"

In this instance, if we were only look at the textual component we see no opinion regarding attitude towards mask-wearing; in fact, but for the presence of the hashtag "#maskup", this Tweet would have no relevance to coronavirus or mask-wearing whatsoever. The inclusion of a simple term imbues the Tweet with pro-mask attitude. We see that when present, hashtags have the potential to carry a great deal of contextual meaning in a Tweet, which is evident when comparing hashtag keyword clouds of the full Tweet set compared to Tweets related to mask-wearing.

Hashtag Co-Occurrence Network

We conclude our exploration of this data with a brief qualitative examination of hashtag co-occurrence in a network of posts, in order to further our descriptive analysis of this data. The relationship of Tweets to mask-wearing hashtags can be seen in Figure 3, depicted as a two-mode/bipartite hashtag-by-Tweet network graph. The small blue vertices represent tweets and the larger multicolored vertices are representative of the mask hashtags contained within those tweets as indicated by edge relationships. Hashtag attitudes toward mask-wearing are indicated by the vertex colors green (pro-mask), red (anti-mask), and yellow (neutral or unknown). This graph illustrates that Tweets containing a particular mask hashtag tend to contain very few other mask hashtags on average, with a mean of 1.14 mask hashtags per tweet, (SD=0.45), the maximum being 4. Moreover, wee see that while the average mask hashtag occurs in 3.83 tweets (SD=8), three hashtags - "#wearadamnmask", "#maskup", and "#wearamask" - are distinct outliers, showing up in 16, 34 and 44 tweets respectively

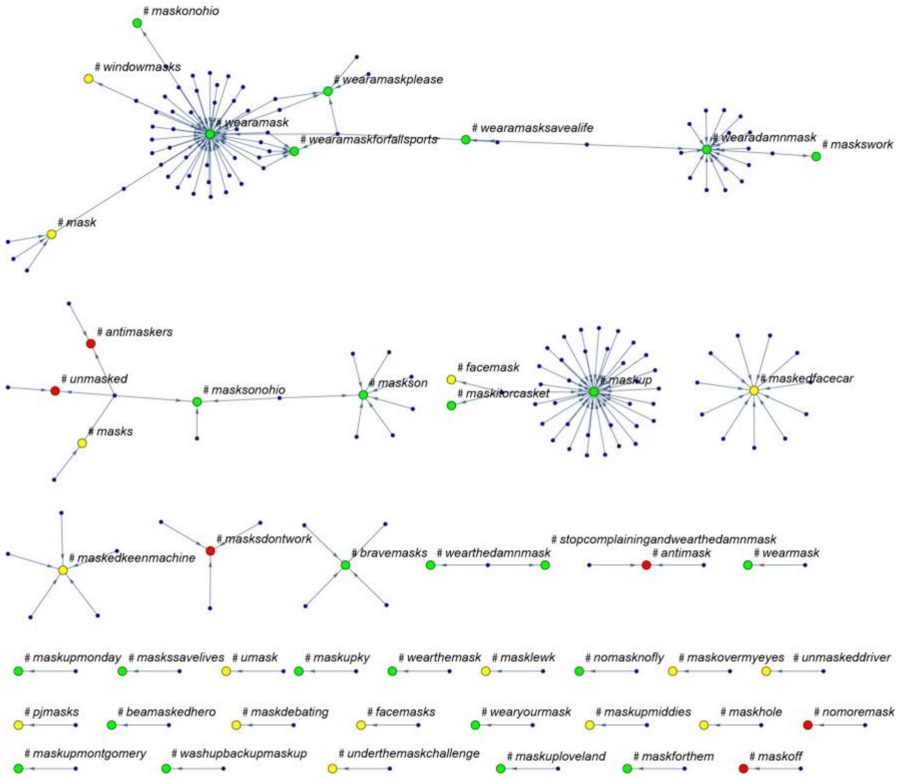

Figure 3. Mask Hashtag-by-Tweet (2-mode) Co-Occurrence network

In Figure 4 the two-mode network is reduced into a single-mode, hashtag-by-hashtag graph, with edges representing hashtag co-occurrences within Tweets. This is a strategy frequently employed in social media research, focusing on dyadic hashtag relationships within tweets. We find that the 48 mask hashtags organize into 4 subgraphs, each appearing to represent a broader pro- or anti-mask attitude. In this graph there are also 29 isolated mask-wearing hashtags that do not co-occur with any other mask-wearing hashtags. When including these isolated hashtags, the average hashtag co-occurs with another hashtag .92 times (*SD* = 1.42), and the mean of these co-occurrences is 1.32 (*SD* = 1.13).

There is little co-occurrence in these clusters of pro-mask and anti-mask keywords, such that each isolated subgraph appears (with some exception) to generally represent a different attitude towards masks. Only in one cluster do we observe what might be considered an ambivalence in mask-wearing attitude, with apparently pro- and anti-mask hashtags co-occurring inside the same tweets. This may be interpreted in a number of ways; again, this is a matter of lack of context. In sections below we contemplate ways in which these graphs may be augmented in future work. Also in the following section we will further discuss the relative contribution that each of the four social media analysis methods offer when contributing to situational awareness about local attitudes.

Figure 4. Mask Hashtag-by-Hashtag (1-mode) Co-Occurrence Network

DISCUSSION

In response to our first overarching research question, which focused on the descriptive characteristics of local data analysis, we examined our data overall and also the descriptive qualities of labeled data. Our exploration of this local data set for the given time period showed that approximately 1.02% of all Tweets gathered were related to mask-wearing (see Table 1). In the deluge of semantic data present in our Twitter dataset, we found that by examining common vehicles of automated and semi-automated natural language analysis provided by sentiment analysis and n-gram evaluation, the information was far too noisy to provide clear paths towards identifying mask-wearing attitude. This seems to be in part due to general issues in language sentiment detection, but also technical and use-driven challenges specific to the Twitter mode of communication and the way in which mask-wearing is publicly discussed. Hashtags are the clear flag-bearer of unequivocal meaning in our dataset. As asserted by the prior literature, this seems likely due to the self-organizing nature of hashtag use, including the tendency of users to self-annotate their posts for the purpose of increased clarity or emphasis. As Tweets containing hashtags comprised only 12% of our mask-wearing dataset, however, the key seems to be building a bridge between attitude among hashtaggers and general mask-wearing attitude, as found inside the textual content of Tweets.

In response to our second overarching research question, which focused on a comparison of labeling approaches to manual coding, we found that labels that were derived from hashtags were most consistent with manual coding (see Figure 1). For this reason, we believe that hashtags are most useful to establishing a baseline attitude towards masks for a given region. We further explored hashtag co-occurrence networks which revealed that pro-mask and anti-mask hashtags tend to be clustered together in subgraphs, which offers opportunities for better categorizing the attitude that may commonly be associated with a particular hashtag. In the following section, we further discuss the relative limitations of each social media labeling method for gaining local situational awareness.

Comparing Limitations of Common Social Media Labeling Methods to Understand Attitude

A limitation present in ascertaining both general sentiment and mask-attitude appears to be a more general limitation inherent in all natural language processing—a lack of situational context. We saw this not only in our proposed VADER and n-gram methodologies, but even in simply arriving at human intercoder consensus: very often, the attitude represented in a Tweet is simply not clear.

Natural human (face-to-face) communication is replete with contextual cues and redundancies, both verbal and non-verbal, that can reinforce the message of the communicator and reduce the chance of miscommunication; as Rabiner & Juang (1993) point out, verbal elements are interpreted very differently between auditory/speech recognition and the evaluation of textual content, not to mention the role of volume, tempo, inflection, and the presence of parallel visual cues. It is extremely easy for subtleties in communication to be missed via forms of digital communication, and thus the impetus for more formalized modes of digital communication such as "structured journalism" (Structured journalism, 2014; Graves and Anderson, 2020). However, owing to technical limitations of the platforms and the overwhelmingly pedestrian/conversational tone of the medium, this measured approach to social media communication is the exception rather than the norm.

One clear limitation in dealing with hashtags is that relevant topical hashtags were few in number; indeed, hashtag-containing Tweets in general were comparatively lacking compared to the full number of Tweets in our dataset. This, among other things, makes it difficult to say definitively if the attitude of hashtagging users is reflective of Twitter users in general, or if they are representative of a distinct population or populations. Indeed, the large difference in distribution of mask-wearing attitudes between the mask hashtag Tweet set and the random sample of mask Tweets seem to suggest this.

Expanding on this last point, we found a distinct homogeneity in mask-wearing attitude among hashtags; among manually-coded Tweets there were nearly twice as many pro-mask than anti-mask posts, and among the hashtag-containing Tweets, the ratio was almost 18:1. Such a low degree of variation in attitude towards mask-wearing among hashtagging users, in addition to highlighting the possibility that it may not be indicative of users in general, would also have the tendency to drown out any potential experimental effect on the anti-mask hashtag group.

Our network analysis was restricted to a handful of hashtags containing the word mask and the 162 Tweets in which they were contained. Expanding the hashtag network to include all other hashtags contained within those Tweets (as depicted in Figure 5) or through recursive searches into adjacent Tweets, we would conceivably gain greater understanding of how hashtags were being used semantically. While such recursion would increase the complexity of the network, the trade off would be a greater contextual account of the mask hashtag network, and presumably more heuristic information with which to ascertain proper meaning.

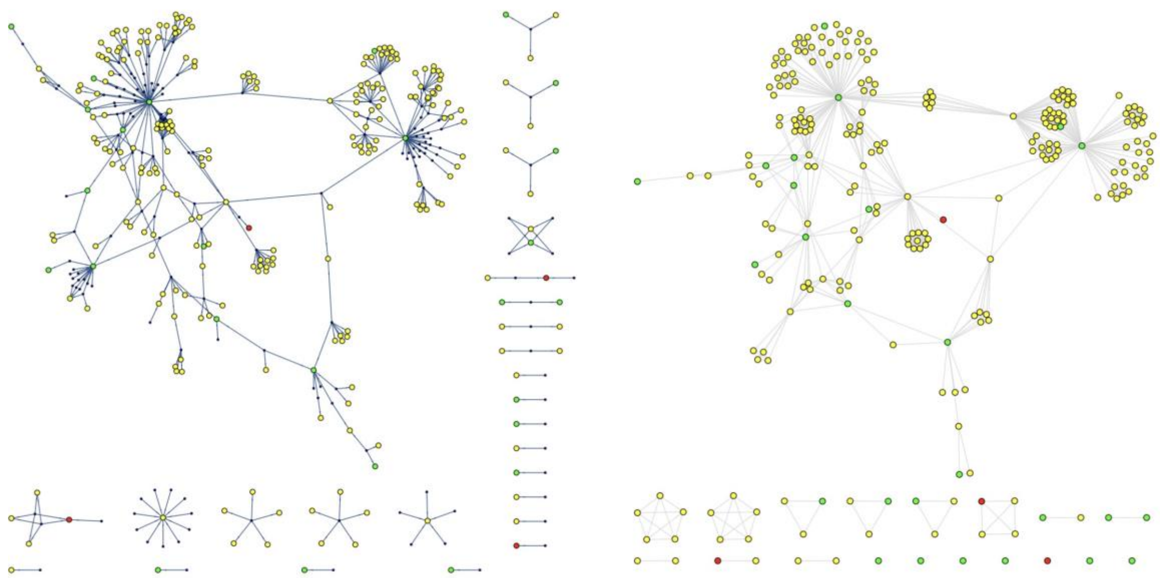

Figure 5. Expanded mask hashtag co-occurrence networks

Study Limitations and Future Work

There are several limitations to this study that may prevent its generalizability to other locations or datasets. First, it should be noted that all Tweets used in this study were associated with a geographic query - all Tweets contained explicit geographic information in some structured form. Geographic Tweets make up a fraction of available Tweets and are thus not necessarily representative of that area as a whole. Second, our study focuses on one particular locale that may experience unique demographic and social media patterns. Another limitation is that the majority of our analysis focuses on random subsets of the dataset in order to feasibly manually code data for comparative purposes. Finally, manually coding may be subject to coder bias even though multiple coders were used to reduce the impact of this limitation. In the following paragraphs, we explore directions for the future of this work.

From our observation, N-grams alone were not particularly useful in our current method of attempting to identify mask-wearing attitude, however, we are not prepared to abandon them just yet. Middleton et al. for example demonstrated that in crisis research, n-grams could be useful markers for inferring the geolocation of Tweets when explicit geodata was otherwise absent, through a simple statistical model. Perhaps using a similar approach but from the standpoint of attitude-assessment might yield output more usable for comparison.

While hashtags and textual content reside side-by-side in a Tweet, they are often evaluated as distinct types of entities. In a variation of observing hashtag co-occurrence, we ponder whether enriching our co-occurrence network with the inclusion of n-gram word tokens might increase the contextual value of both hashtags and the word tokens. By treating hashtags and n-grams as the same time of vertex, we may very well discover new avenues of centrality and meaning-transmission.

Our ultimate goal is provide an automated or semi-automated pathway through which mask-wearing attitude (and potentially other event-specific types of attitude) can be derived from a corpus of Tweets. If and when this is achieved, Tweet data will lend itself more readily to machine learning and other statistical models of language prediction and analysis, and ultimately contribute to increased situational awareness during crisis situations. This could potentially enable more efficient and impactful public messaging, especially during protracted crises such as we have seen with COVID-19.

CONCLUSION

In this paper, we have examined local attitudes towards mask-wearing on using a variety of social media analysis methods. Our results highlight several factors that help in selecting social media analysis methods using for situational awareness of local attitudes, namely the reliability of various methods in correctly identifying local attitude and methods to identify hashtag analysis starting points quickly. Overall, our results indicate that sentiment analysis and n-gram filtering are not reliable tools to gauge local attitudes on this topic and warrants future work to refine methods that allow researchers to identify a baseline metric that establishes the percent of all Tweets on a topic that are either for or against an issue. One important contribution from this study is the use of hashtag analysis to better understand the variety of ways that a community has discussed an issue as a method to quickly select hashtags that co-occur in Tweets, as we observed with the network analysis that was performed on hashtag co-occurrence networks.

REFERENCES

A.M. Association (2020). Mask Up: Stop the spread of COVID-19 | American Medical Association. https://www.ama-assn.org/delivering-care/coronavirus/mask-stop-spread-covid-19. 2020.

Alshaabi, T., Arnold, M. V., Minot, J. R., Adams, J. L., Dewhurst, D. R., Reagan, A. J., Muhamad, R., Danforth, C. M., and Dodds, P. S. (2021). "How the world's collective attention is being paid to a pandemic: COVID-19 related n-gram time series for 24 languages on Twitter". In: PLoS ONE 1 (2021).

Avvenuti, M., Cresci, S., Marchetti, A., Meletti, C., and Tesconi, M. (2014). "Ears (earthquake alert and report system) a real time decision support system for earthquake crisis management". In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. 2014, pp. 1749–1758.

Broder, A. Z., Glassman, S. C., Manasse, M. S., and Zweig, G. (1997). "Syntactic clustering of the Web". In: Computer Networks and ISDN Systems 29.8 (1997), pp. 1157–1166.

Brown, K., Kellner, A., and Raggett, D. (2001). Stochastic language models n-Gram specification. W3C. 2001. URL: https://www.w3.org/TR/ngram-spec/.

Bruns, A., and Stieglitz, S. (2013). "Towards more systematic Twitter analysis: metrics for tweeting activities." In: International Journal of Social Research Methodology 16.2 (2013), pp. 91–108.

Bruns, A., Burgess, J., Crawford, K., and Shaw, F. (2012). # qldfloods and@ qpsmedia: Crisis communication on twitter in the 2011 south east queensland floods. 2012.

Chatfield, A., and Brajawidagda, U. (2012). "Twitter tsunami early warning network: a social network analysis of Twitter information flows". In: (2012).

Damerau, F. J. (1966). "The role of Markov models in linguistic theory". English. PhD thesis. 1966, p. 240.

Elrod, N., Hall, H., Mahajan, P., Grace, R., and Kropczynski, J. (2021). "Data census of a geographically-bounded tweet set for enhancing common operational picture tools". In: Proceedings of the Hawaii International Conference on Systems Sciences. HICSS-54. Kauai, Hawaii, USA, 2021, in press.

Endsley, M. R. (1995). "Toward a theory of situation awareness in dynamic systems". In: Human factors 37.1 (1995), pp.

32–64.

Gilbert, C., and Hutto, E. (2014). "Vader: A parsimonious rule-based model for sentiment analysis of social media text". In: Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf. Vol. 81. 2014, p. 82.

GitHub - cjhutto/vaderSentiment: VADER Sentiment Analysis. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, and works well on texts from other domains. https://github.com/ cjhutto/vaderSentiment.

Graves, L., and Anderson, C. (2020). "Discipline and promote: Building infrastructure and managing algorithms in a 'structured journalism' project by professional fact-checking groups". In: New Media and Society 22.2 (2020), pp. 342–360.

Hachaj, T., and Ogiela, M. R. (2017). "Clustering of trending topics in microblogging posts: A graph-based approach". In: Future Generation Computer Systems 67 (2017), pp. 297–304.

Halse, S. E., Tapia, A. Squicciarini, A., and Caragea, C. (2018). "An emotional step toward automated trust detection in crisis social media". In: Information, communication & society 21.2 (2018), pp. 288–305.

Hassan, A., Abbasi, A., and Zheng, D. (2013). "Twitter sentiment analysis: A bootstrap ensemble framework". In: Proceedings of the ASE/IEEE International Conference on Social Computing. 2013, pp. 357–364.

Hiltz, S. R. and Plotnick, L. (2013). "Dealing with information overload when using social media for emergency management: Emerging solutions." In: ISCRAM. 2013.

Hiltz, S. R., Gonzalez, J. J., and Turoff, M. (2013). "ICT support and the effectiveness of decision making in disasters: A preliminary system dynamics model." In: ISCRAM. 2013.

I. D. S. of America (2020). Experts in Infectious Diseases Team Up With Ad Council To Encourage People To Wear Face Masks In Public. https://www.idsociety.org/news--publications-new/articles/2020/ experts- in- infectious- diseases- team- up- with- ad- council- to- encourage- people- to- wear-face-masks-in-public/. Oct. 2020.

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., and Meier, P. (2013). "Extracting information nuggets from disaster-Related messages in social media." In: Iscram. 2013.

Lavrakas, P. (2008). Encyclopedia of Survey Research Methods. Sage, 2008. Chap. Intercoder Reliability.

Lee, H., Abdar, M., and Yen, N. (2018). "Event-based trend factor analysis based on hashtag correlation and temporal information mining". In: Applied Soft Computing 71 (2018), pp. 1204–1215.

Liu, B. (2012). "Sentiment analysis and opinion mining". In: Synthesis lectures on human language technologies 5.1 (2012), pp. 1–167.

Mendonca, D., Beroggi, G. E., and Wallace, W. A. (2001). "Decision support for improvisation during emergency response operations". In: International journal of emergency management 1.1 (2001), pp. 30–38.

Middleton, S. E., Middleton, L., and Modafferi, S. (). "Real-Time Crisis Mapping of Natural Disasters Using Social Media". In: IEEE Intelligent Systems 29.2 (), pp. 9–17.

Mishra, P., Danneman, N., Heimann, R., and Makhabel, B. (2017). R: Mining Spatial, Text, Web, and Social Media Data. Packt Publishing, 2017.

Neppalli, V. K., Caragea, C., Squicciarini, A., Tapia, A., and Stehle, S. (2017). "Sentiment analysis during Hurricane Sandy in emergency response". In: International journal of disaster risk reduction 21 (2017), pp. 213–222.

Nguyen, D., Ali Al Mannai, K., Joty, S., Sajjad, H., Imran, M., and Mitra, P. (2017). "Robust classification of crisis-related data on social networks using convolutional neural networks". In: Proceedings of the International AAAI Conference on Web and Social Media 11.1 (May 2017).

Prather, K. A., Wang, C. C., and Schooley, R. T. (2020). "Reducing transmission of SARS-CoV-2". In: Science (2020).

Rabiner, L., and Juang, B. (1993). Fundamentals of Speech Recognition. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

Rao, L. (2009).. Twitter makes hashtags more #useful. TechCrunch.com https://techcrunch.com/2009/07/ 02/twitter-makes-hashtags-more-useful/. July 2009.

Seward, Z. (2013). The first-ever hashtag, -reply and retweet, as Twitter users invented them. Quartz.com https:/ / qz . com / 135149 / the - first - ever - hashtag - reply - and - retweet - as - twitter - users - invented-them/. Oct. 2013.

Singh, N., Roy, N., and Gangopadhyay, A. (2018). "Analyzing the sentiment of crowd for improving the emergency response services". In: 2018 IEEE International Conference on Smart Computing (SMARTCOMP). IEEE. 2018, pp. 1–8.

Son, J., Lee, H. K., Jin, S. and Lee, J. (2019). "Content features of tweets for effective communication during disasters: A media synchronicity theory perspective". In: International Journal of Information Management 45 (2019), pp. 56–68.

Source code for nltk.tokenize.casual. NLTK 3.5 documentation https://www.nltk.org/_modules/

nltk/tokenize/casual.html#TweetTokenizer. 2020.

Structured journalism. Duke Reporter's Lab. 2014. URL: https://reporterslab.org/structured- journalism/.

Toosinezhad, Z., Mohamadpoor, M., and Malazi, H. T. (2019). "Dynamic windowing mechanism to combine sentiment and N-gram analysis in detecting events from social media". English. In: Knowledge and Information Systems 60.1 (July 2019), pp. 179–196.

Van de Walle, B., and Turoff, M. (2008). "Decision support for emergency situations". In: Handbook on decision support systems 2. Springer, 2008, pp. 39–63.

van den Berg, J. A. (2014). "The story of the hashtag(#): A practical theological tracing of the hashtag(#) symbol on Twitter". en. In: HTS Theological Studies 70 (Jan. 2014), pp. 1–6.

Wilson, L. (2018). Semantic network analysis of ideological communities on twitter. Northwestern University:MSiA Student Research https://sites.northwestern.edu/msia/2018/12/04/semantic- network- analysis-of-ideological-communities-on-twitter/. Dec. 2018.

Wukich, C. and Steinberg, A. (2013). "Nonprofit and public sector participation in self-organizing information networks: Twitter hashtag and trending topic use during disasters". In: Risk, Hazards & Crisis in Public Policy 4.2 (2013), pp. 83–109.

Zappavigna, M. (2015). "Searchable talk: the linguistic functions of hashtags". In: Social Semiotics 25.3 (2015), pp. 274–291.