Identification of Violence in Twitter Using a Custom Lexicon and NLP

Jonathan Adkins
Norwich University, Northfield, USA
jadkins@norwich.edu

Abstract: Information warfare is no longer a denizen purely of the political domain. It is a phenomenon that permeates other domains, especially those of mass communications and cybersecurity. Deepfakes, sock puppets, and microtargeted political advertising on social media are some examples of techniques that have been employed by threat actors to exert influence over consumers of mass media. Social Network Analysis (SNA) is an aggregation of tools and techniques used to research and analyze the nature of relationships between entities. SNA makes use of such tools as text mining, sentiment analysis, and machine learning algorithms to identify and measure aspects of human behavior in certain defined conditions. One area of interest in SNA is the ability to identify and measure levels of strong emotions in groups of people. In particular, we have developed a technique in which the potential for increased violence within a community can be identified and measured using a combination of text mining, sentiment analysis, and graph theory. We have compiled a custom lexicon of terms used commonly in discussions relating to acts of violence. Each term in the lexicon has a numerical weight associated with it, indicating how violent the term is. We will take samples of online community discussions from Twitter and use the R and Python programming languages to cross-reference the samples with our lexicon. The results will be displayed in a Twitter discussion graph where the user nodes are color-coded according to the overall level of violence that is inherent in the Tweet. This methodology will demonstrate which communities within an online social network discussion are more at risk for potentially violent behavior. We assert that when this approach is used in association with other NLP techniques such as word embeddings and sentiment analysis, it will provide cybersecurity and homeland security analysts with actionable threat intelligence.

Keywords: social network analysis, natural language processing, sentiment analysis, text mining

1. Introduction

The Internet has brought with it several immeasurable benefits, such as the ability to instantly share information between multiple entities over great distances. It has also ushered in several negative variables that appeal to some of the basest aspects of the human psyche. These include anonymity, aggression, criminal opportunism, and appeals to engage with others in violent acts. The most ubiquitous venue on the internet where these antisocial behaviors are explored is social media. Twitter has been used by terrorists to recruit as well as disseminate information (Oh, Agrawal and Rao, 2011). Facebook accounts have been used for countless cases of cyberbullying (Sharif & Hoque, 2021). A niche microblogging site called *Parler* served as a staging platform for the January 6 rioters in Washington DC (Prabhu et al, 2021). There have been many studies in the academic literature that seek to use Natural Language Processing (NLP) techniques to identify and measure aspects of aggression and violence in social media (Lytos et al, 2019). Ni et al (2020) used interview data with analytics to predict a student's risk of violence in a school setting. Bigrams and trigrams were used with an unsupervised classifier to identify hate speech in a polarized political environment in Nigeria (Udanor and Anyanwu, 2019). Studies such as these only capture a small piece of the puzzle. Bigrams and trigrams by themselves do not provide a full context of what we seek to know from an online population. Based on empirical data, many sentiment lexicons used in the literature do not provide the granularity that is needed to isolate the variables that are needed to identify people who are at risk for violence (Rekik, Jamoussi and Hamadou, 2019). Lexicons such as the Bing and NRC are either too narrow or too broad to succinctly identify the elements that precipitate violence in a community. These limitations have created a gap in the literature concerning studies of online aggression and violence. To this end, we propose a custom niche lexicon containing terms specifically related to violence in the English Language (Beigi and Moattar, 2021). Our lexicon is not confined to specific domains of politics or deviant psychology. Our lexicon will contain terms from politics, popular culture slang, hacking terms, and criminal justice terms. Each term is weighted for severity, ranging from one to three. Our custom lexicon will identify terms from Twitter and will aggregate the weighted values to form a *violence score*. Our Violence Score will be represented in graphs on the x-axis compared with sentiment, subjectivity, and emotion. When the violence score is evaluated against these three variables, we will provide more insight into the context of violence that is inherent within an online community (Akhtar, Ekbal and Cambria, 2020). The remainder of this paper will be divided into eight sections. Sections two and three will discuss the violence lexicon and the Twitter dataset. Sections four, five, and six will discuss the sentiment, subjectivity, and emotion graphs juxtaposed to

Proceedings of the 21st European Conference on Cyber Warfare and Security

the violence score. Section seven will discuss our color-coded social network graph. In the final section, we will discuss our conclusions and future work.

2. Violence lexicon

Beigi and Moattar (2021) developed a framework for creating a domain-specific lexicon for positive and negative sentiment. Our approach differs from their technique in two fundamental ways. First, we include multiple domains under the larger umbrella of violence. Under our paradigm, terms from different domains are relevant if they in some way suggest or directly invoke the construct of violence. Second, we don't assign *positive* or *negative* valence to our terms. We assign a numerical weight to a term depending upon its perceived severity. In its current proof-of-concept version, a word such as warlord suggests the potential for violence but does not directly invoke it, therefore it gets a score of one. A term such as eviscerate directly describes a severely violent act, therefore it gets a score of three (the highest). In later research, we plan to expand the number of terms. We also will re-evaluate the scoring metric to see if a larger range of values is warranted. A weight of two is assigned to words that are perceived as violent but not as severe as a level three term. The weights assigned to individual terms will also be re-evaluated in later research since manual labeling is a subjective task and requires additional consideration. Table 1 below displays a small sample from the lexicon. The term associated with violence is displayed in the left column. The weight, ranging from 1 to 3, is displayed in the right column.

Table 1: Sample from the Violence lexicon

word	weight
abduct	2
abuse	2
aggressor	2
aggression	2
agitated	2
agitator	2
airstrike	3
ambush	2
anarchism	1
anarchy	1
anguish	2
annihilate	3

3. Twitter dataset

Khader, Awajan and Al-Naymat (2018) stated the value of using social media data for sentiment analysis. According to their study, platforms such as Twitter provide large volumes of "high-velocity data" that contains valuable information that can be extracted for analysis. Kausar, Soosaimanickam and Nasar (2021) used Twitter data from the most infected countries in the world during the height of the COVID-19 pandemic to understand how the countries dealt with the crisis. Instead of applying sentiment analysis tools and social media data toward studies of public health, we intend to apply the "high-velocity" social media data from the Twittersphere toward the identification and metastasizing of violence in a given population. This is a proof-of-concept study and our custom lexicon currently only hosts 400 terms with associated weights. For the sake of reference, the National Research Council Canada (NRC) lexicon has approximately 20,000 terms to evaluate eight different emotions in text (Khoo & Johnkhan, 2018). Based on prior empirical studies, we decided upon a Twitter dataset of 30,000 tweets to evaluate for *violence*, *sentiment*, *subjectivity*, and *emotion*. We collected the dataset using the R programming language's "twitteR" library, which contains a plethora of tools for querying Twitter's public-facing Application Programming Interface (API). We queried the Twitter API for any tweets currently in circulation based on the keyword search "assault." We saved the query results to a comma-separated value (csv) file. For our evaluation, we used three columns from the csv file. These were *screenName*, *text*, and *isRetweet*. In the next section, we will discuss assessing the *violence score*.

3.1 Assessing the violence score

The first step in our study was to identify any tweets that contained terms that were listed in the violence lexicon and discard any tweets which did not contain any violent terms. A script was used in R which looped through all 30,000 rows of the dataset and compared the terms in the text field to the terms that were in the violence lexicon. Out of the original 30,000 rows of tweets from the dataset, approximately 15,000 contained one or more violent terms that were listed in our lexicon. For the tweets in the dataset that remained, the R script looped through each one row by row and identified one or more violent terms in the tweet. The weights for each term were added together and the resulting score was written to a new column called *ViolenceScore*. Depending on how many violent terms were in a tweet, the resulting score could range from a one to approximately thirteen. A score of *one* suggests a solitary word with mild violent content. A score of *five* or more suggests that the tweet was significantly more violent in content. The values in the ViolenceScore column were placed on the x-axis for the sentiment, subjectivity, and emotion graphs. By approaching the graphs in this manner, we were able to see the distribution of violence scores relative to the responses by users in the population.

4. Sentiment and violence

To evaluate the sentiment analysis of our assault dataset, we used an R library called *word2vec*. Word2vec uses word vectors (otherwise known as word embeddings) to mathematically define words and evaluate their context (Giatsoglou et al, 2017). A machine learning neural network is used to train a sample of tokenized words taken from the original tweets. Word2vec assigns a probability for each tweet as to *how negative* or *how positive* the tweet is. Each dot in this graph represents a tweet. The closer to zero a tweet is, the more negative it is. The closer to 1.0 the tweet is, the more positive it is. As seen in Figure 1 below, the ViolenceScore column was placed on the x-axis, so that the distribution of tweets can be seen juxtaposed to their sentiment valence (positive or negative). We made the following observations with regard to the sentiment analysis of the *assault* dataset. The majority of the tweets had a violence score of 3 (still mild). In this level of violence, there is a full range of sentiments, from extremely negative to extremely positive. Another observation we made was that the more violent the content became, the fewer in number the tweets were for those violence scores. The tweets with the highest violence scores were outliers and were highly negative.
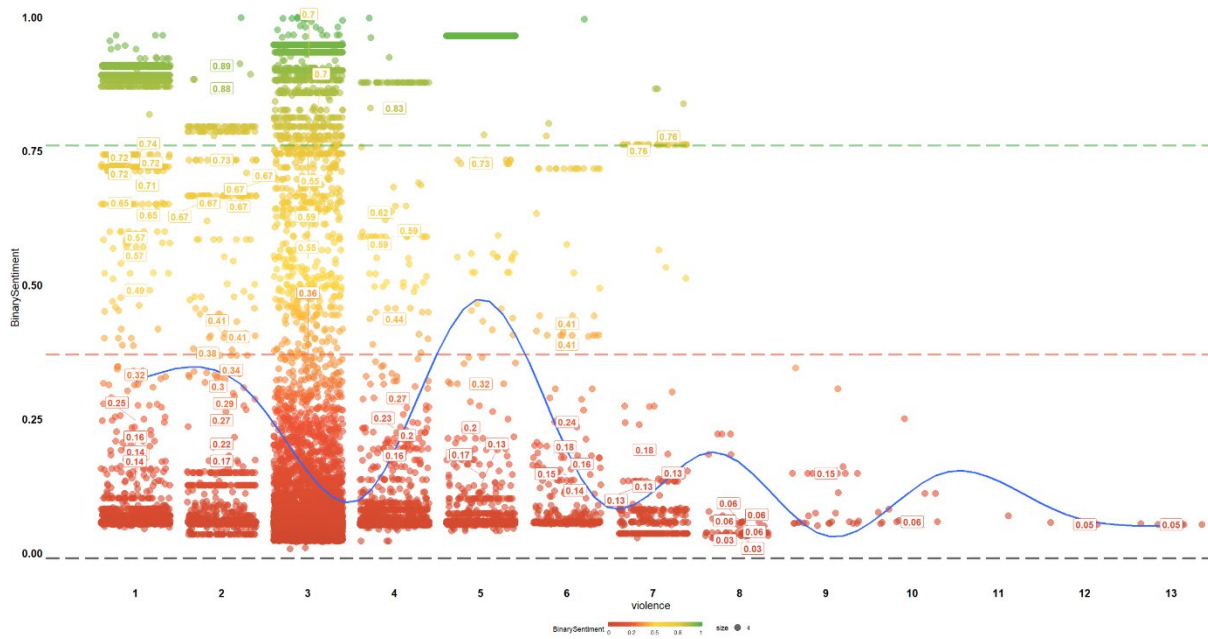

Figure 1: Sentiment analysis and violence

5. Subjectivity and violence

Subjectivity is a metric that is often included in the same class as *sentiment* and *emotion*. There are notable differences between these three constructs. *Sentiment* (which we discussed in the previous section) is a valence that exists between positive and negative. By itself, it does not articulate a specific emotion (Neogi et al, 2021). *Subjectivity* is a metric that seeks to evaluate how opinionated or factual a person's statements are (Yaqub et al,

2018). By combining the sentiment and subjectivity metrics, there is a more vivid context for a person's intent (Akhtar, Ekbal and Cambria, 2020). For example, a person can speak at length and have his speech qualify as highly negative based on sentiment analysis. However, if it was demonstrated that his subjectivity measured very low, it could suggest that the person was speaking academically and objectively about negative subject matter.

We used word2vec with the assault Twitter dataset in order to evaluate the level of subjectivity in the text. Even though this is the same approach we used for sentiment analysis, the technique differed in that we had to calculate the subjectivity scores for the assault dataset. We accomplished this task by using the Python programming language and a library called textblob. The textblob library assesses input text for words expressing opinion and feelings toward a topic (Saha, Yadav and Ranjan, 2017). For example, if a speaker says, "I think" or "this should be," he is conveying *opinion*. If statements lack qualifiers such as these, there is a higher probability that the statements are more *objective* or *neutral*. The textblob library was used in Python to score the assault Twitter dataset for subjectivity. Word2vec was then used to evaluate the overall subjectivity of the tweets with the violence score values on the x-axis. The results can be seen in the graph below in Figure 2.
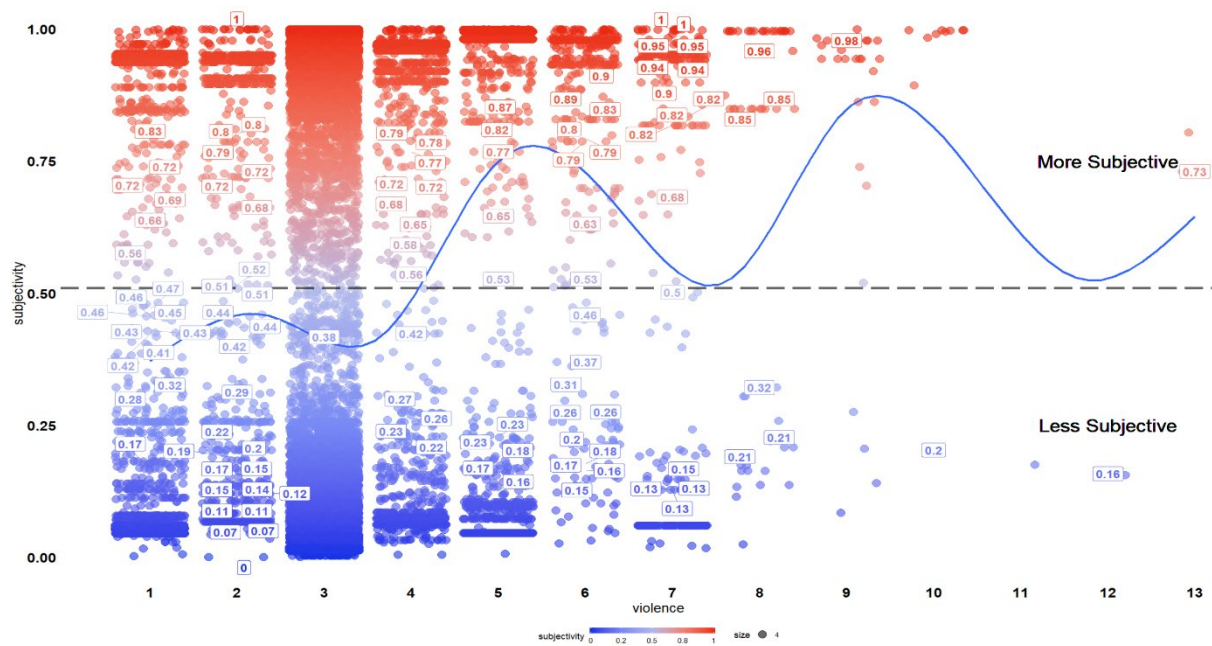

Figure 2: Subjectivity analysis and violence

Based on our findings, there was one significant difference between the subjectivity and sentiment graphs. We observed that as the violent content of the tweets increased (higher violence score), subjectivity also increased. This observation by itself if inconclusive. Several more samples would be needed to assess whether this was a recurring pattern. There is one variable in this approach that can be tweaked. Word2vec works on a probability scale. In both sentiment and subjectivity, the tweets were assessed as to how likely they were to be positive or negative, subjective or not subjective. For subjectivity, we set the threshold to .25 or 25%. This means that if a tweet was scored as 25% subjective, it was overall more subjective. The .25 value itself may be too high. The more words of personal opinion we add to a statement, the more subjective it becomes. We felt for this proof-of-concept experiment that 25% was a valid threshold.

6. Emotion and violence

In sections 4 and 5 we compared our Twitter sentiment and subjectivity distributions to their relative violence scores (Peng et al, 2021). We have observed thus far that the most violent content tends to be outliers that are highly negative and highly subjective. The integration of sentiment and subjectivity provides us with a sharper, more enhanced perspective into a speaker's intent. What we were missing up until this point was an emotional metric to provide a more complete answer to the question: *why* is the speaker more violent, more subjective, and more negative (Akhtar, Ekbal and Cambria, 2020)?

The National Research Council Canada (NRC) lexicon measures the amount of emotion in text. The NRC lexicon is actually an aggregation of ten different lexicons. Eight of the ten lexicons are emotions: anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. It also includes lexicons for positive and negative sentiments (Agarwal and Toshniwal, 2018). Based on empirical findings, some of the emotion lexicons do not provide useful feedback. For example, there are more words in the anticipation lexicon than in the other lexicons. This frequently causes a skewed result where the anticipation observations are higher than any other emotion (Zad, Jimenez and Finlayson, 2021). Based on this empirical observation, we removed the anticipation, trust, and disgust lexicons when we run assessments. We assessed the assault Twitter dataset using a truncated version of the NRC lexicon: anger, fear, joy, sadness, and surprise. The results can be seen in the graph below in Figure 3. Violence scores were placed on the x-axis and the percentage of associated emotion words on the y-axis. We tried several different types of plots, but most did not provide us with insightful observations. We decided to create our emotion graph using violin plots. The violin plots allowed us to adequately measure the levels of violence juxtaposed to the percentage of emotion-related words from the Twitter dataset (Sinha et al, 2021).
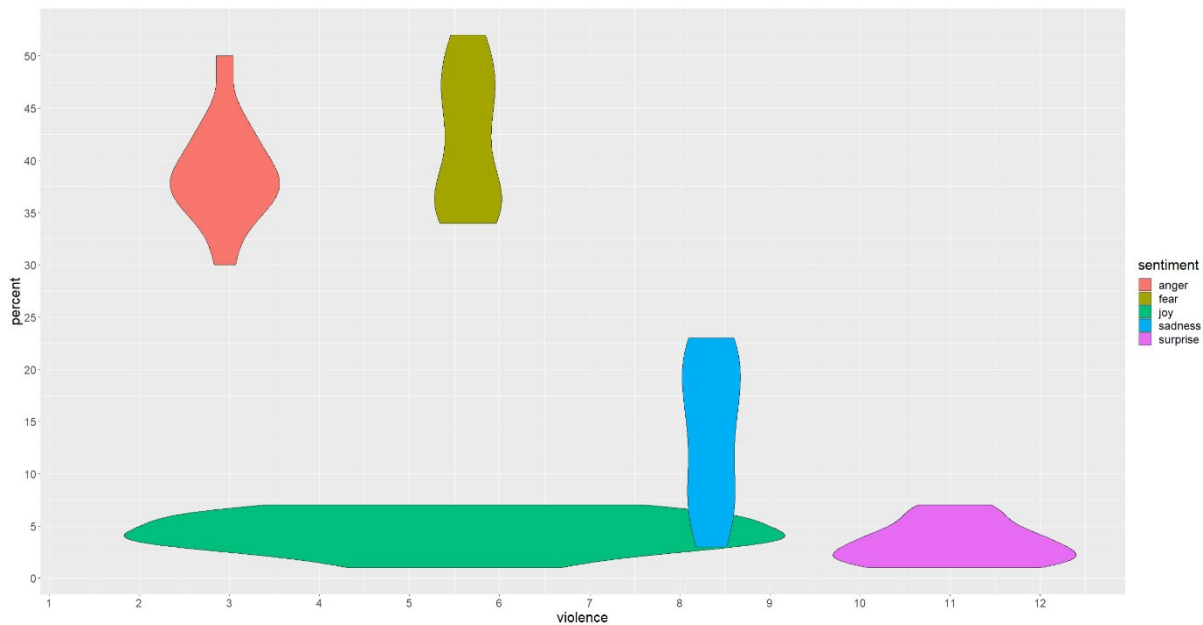

Figure 3: NRC emotion lexicon graph

After reviewing the violin plots of anger, fear, joy, sadness, and surprise we made the following observations. As violence scores increased, the prevailing emotion was surprise. Tweets with violence scores ranging from 8 to 9 tended to convey sadness. Violence scores that ranged between 5 and 6 conveyed fear. The majority of tweets in the assault dataset scored in the 3 range, which suggested that they were predominantly angry. From these scores it could be extrapolated that the average Twitter user from this dataset was only mildly violent and angry. The sentiment scores for this level of violence were evenly distributed between positive and negative. The subjectivity score for 3 range was also more objective. Overall, these observations suggest that the Twitter users were debating more objective facts from ideologically opposing viewpoints. In section 7 we will further demonstrate the utility of our violence lexicon by creating a social network graph using the violence scores as attributes.

7. Social network graph using the violence scores as attributes

Social Network Analysis (SNA) is an interdisciplinary domain of research that has been around since the early 1990's (Min et al, 2021). One of the many tools used by SNA researchers is the social network graph. One of the primary purposes of a SNA graph is to model relationships between entities in a population. SNA graphs are composed by two principal features called edges and nodes. Nodes represent specific entities and edges represent relationships between entities (Tabassum et al, 2018). There are several implementations of SNA graphs that range from public health and law enforcement to search engine optimization. The application of the graph depends on which aspect of the graph researchers wish to focus (nodes or edges). Homeland Security and other law enforcement agencies use SNA graphs to model the relationships between criminal organizations such as terrorist networks or organized crime (Min et al, 2021). Graphs have also been used by epidemiology

researchers to model the spread of COVID-19 in a population (So et al, 2020). Another implementation of SNA graph analysis is called community detection. A graph of an online discussion taken from Twitter is represented by a composite set of nodes and edges. Within this larger composite collective of points and lines, there are smaller subnetworks where the connections between certain nodes are denser. These subnetworks that exist within a larger online discussion are referred to as *communities*. The area of SNA research that endeavors to identify graph subcommunities is called *community detection* (Du et al., 2007). Community detection can provide researchers with an abundance of information concerning group identities, ideological differences, and the flow of information (Kanavos et al, 2018). A good example of community detection being implemented is Campan, Cuzzocrea and Truta (2017). In their study, community detection was used to detect fake news as it spread through online communities.

For our proof-of-concept, we created a SNA graph (seen in Figure 4) which displays the retweet relationships between Twitter users in an online conversation concerning assault. The composite violence score for this dataset ranged from one to thirteen. In order to efficiently depict this information, we scaled the sizes of each node in the graph to correspond to their violence score. The larger a node is, the higher its violence score. In addition, we color-coded the nodes using thirteen shades of red. The nodes with the highest violence scores have the deepest shade of red. Nodes with lower violence scores are smaller and have lighter shades of red. If we look at the graph below in Figure 4, we see that several subnetworks or communities have formed. In each community, there is at least one node or Twitter user that is more violent than the others. The benefit of this type of modeling is that researchers can isolate individual subnetworks in this discussion graph and identify which communities have a larger number of larger scaled nodes that are deeper red in color.
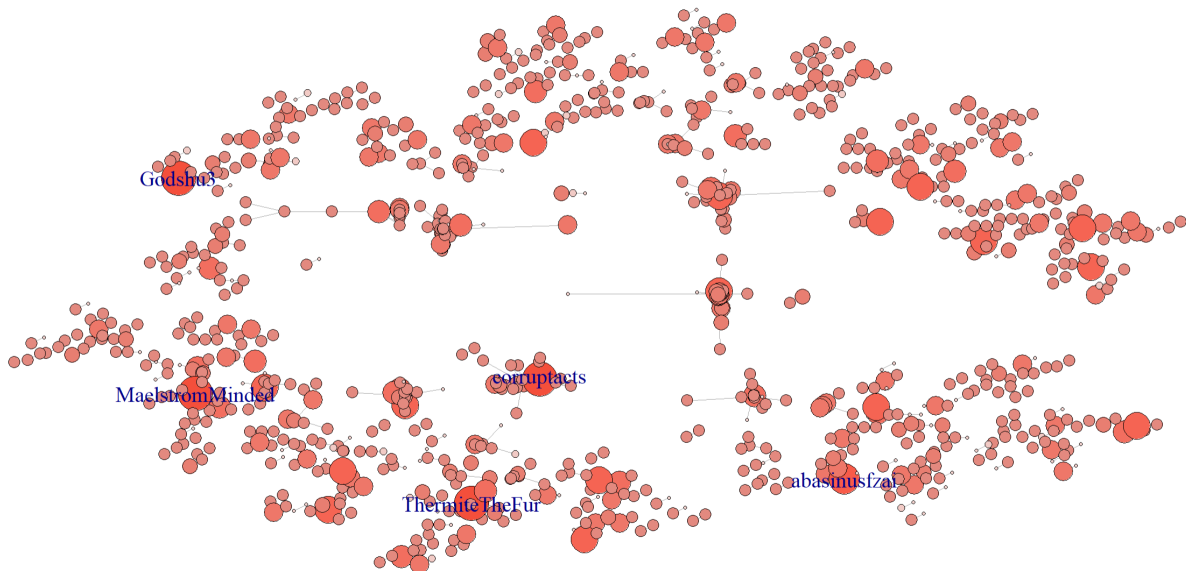

Figure 4: Social network graph of Twitter *assault* online discussion

7.1 Bigrams and trigrams found in the Twitter discussion network

In many studies using SNA and NLP, unigrams (keywords), bigrams (two-word phrases), and trigrams (three-word phrases) are used to capture principal themes that exist in a body of text (Arts, Hou and Gomez, 2021). We extracted the top 10 bigrams and trigrams from the assault Twitter dataset. All three n-gram combinations are necessary for extracting themes. Through empirical analysis, we have found that some themes will not become apparent until larger n-gram combinations are extracted. Table 2 and Table 3 display the themes that were most prevalent in the network seen in Figure 4. From our list of bigrams, the most vivid themes we found were "assault weapons" and "background checks." If we extrapolate the larger discussion based on the phrases in this table, the evidence suggests that Twitter users were discussing themes relating to assault weapons and firearm legislation.

The most insightful trigrams from Table 3 are "high capacity magazines," "ban assault weapons," and "universal background checks." These themes further validate what was found using bigrams. Based on the frequency of

the bigrams and trigrams, topics relating to automatic firearms and gun control legislation are the most frequently discussed themes among Twitter users in this online conversation.

Table 2: Top 10 bigrams from *assault* Twitter dataset

bigram	n
assault weapons	1996
ban assault	1822
high capacity	1820
capacity magazines	1800
background checks	1541
universal background	1537
checks ban	1532
pass universal	1531
magazines repeal	1474
occupied jerusalem	1189

Table 3: Top 10 trigrams from *assault* Twitter dataset

trigram	n
high capacity magazines	1800
ban assault weapons	1790
universal background checks	1538
background checks ban	1532
pass universal background	1531
checks ban assault	1530
capacity magazines repeal	1474
sheikh jarrah neighborhood	849
u.s rightly condemns	849
also under threat	846

8. Conclusion and future work

In this study, we proposed a proof-of-concept algorithm for the identification and evaluation of violence within an online social network. For this paper we used Twitter as our case study, however, the framework can be implemented to suit any online social media platform. The novel contribution that we offer in this study is the creation of a violence lexicon. The lexicon is highly focused on areas that articulate concepts of violence, therefore it is more unique than existing lexicons such as the NRC, Bing, or AFINN lexicons. To maximize the utility of this lexicon, we have included numerical weights for each term in the lexicon. Currently, the weights range from scores of one to three. We plan to revisit the scoring mechanism and perhaps expand the range after further consideration. In this study, we also demonstrated an ensemble of SNA metrics that were used in concert with our violence score. Specifically, we used word embedding-based sentiment analysis, subjectivity analysis, and emotion analysis in plots with violence scores on the x-axis. By plotting SNA graphs with violence along the x-axis, we could quickly identify the distribution of sentiment, subjectivity, and emotions per unit of aggregated violence. This technique allows us to identify trends concerning human responses to certain inds of subject matter in online discussions. The aggregation of sentiment, subjectivity, and emotion allowed us to put user responses into a more vivid context. For future research, we plan to integrate sentiment, subjectivity, emotion, and violence into a single composite score that incorporates time series into its algorithm. By integrating these features with a probabilistic function, we will be able to predict *how violent* an online community will get in the future.

References

Agarwal, A. and Toshniwal, D., 2018, June. Application of lexicon based approach in sentiment analysis for short tweets. In *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)* (pp. 189-193). IEEE.

Akhtar, M.S., Ekbal, A. and Cambria, E., 2020. How intense are you? Predicting intensities of emotions and sentiments using stacked ensemble [application notes]. *IEEE Computational Intelligence Magazine*, *15*(1), pp.64-75.

Arts, S., Hou, J. and Gomez, J.C., 2021. Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy*, *50*(2), p.104144.

Beigi, O.M. and Moattar, M.H., 2021. Automatic construction of domain-specific sentiment lexicon for unsupervised domain adaptation and sentiment classification. *Knowledge-Based Systems*, *213*, p.106423.

Campan, A., Cuzzocrea, A. and Truta, T.M., 2017, December. Fighting fake news spread in online social networks: Actual trends and future research directions. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 4453-4457). IEEE.

Du, N., Wu, B., Pei, X., Wang, B. and Xu, L., 2007, August. Community detection in large-scale social networks. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 16-25).

Giatsoglou, M., Vozalis, M.G., Diamantaras, K., Vakali, A., Sarigiannidis, G. and Chatzisavvas, K.C., 2017. Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications*, *69*, pp.214-224.

Kanavos, A., Perikos, I., Hatzilygeroudis, I. and Tsakalidis, A., 2018. Emotional community detection in social networks. *Computers & Electrical Engineering*, *65*, pp.449-460.

Kausar, M.A., Soosaimanickam, A. and Nasar, M., 2021. Public sentiment analysis on Twitter data during COVID-19 outbreak. *Int. J. Adv. Comput. Sci. Appl*, *12*(2), pp.415-422.

Khader, M., Awajan, A. and Al-Naymat, G., 2018, November. The effects of natural language processing on big data analysis: Sentiment analysis case study. In *2018 International Arab Conference on Information Technology (ACIT)* (pp. 1-7). IEEE.

Khoo, C.S. and Johnkhan, S.B., 2018. Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, *44*(4), pp.491-511.

Lytos, A., Lagkas, T., Sarigiannidis, P. and Bontcheva, K., 2019. The evolution of argumentation mining: From models to social media and emerging tools. *Information Processing & Management*, *56*(6), p.102055.

Min, S., Gao, Z., Peng, J., Wang, L., Qin, K. and Fang, B., 2021. STGSN—A Spatial–Temporal Graph Neural Network framework for time-evolving social networks. *Knowledge-Based Systems*, *214*, p.106746.

Neogi, A.S., Garg, K.A., Mishra, R.K. and Dwivedi, Y.K., 2021. Sentiment analysis and classification of Indian farmers' protest using twitter data. *International Journal of Information Management Data Insights*, *1*(2), p.100019.

Ni, Y., Barzman, D., Bachtel, A., Griffey, M., Osborn, A. and Sorter, M., 2020. Finding warning markers: leveraging natural language processing and machine learning technologies to detect risk of school violence. *International journal of medical informatics*, *139*, p.104137.

Obembe, D., Kolade, O., Obembe, F., Owoseni, A. and Mafimisebi, O., 2021. Covid-19 and the tourism industry: An early stage sentiment analysis of the impact of social media and stakeholder communication. *International Journal of Information Management Data Insights*, *1*(2), p.100040.

Oh, O., Agrawal, M. and Rao, H.R., 2011. Information control and terrorism: Tracking the Mumbai terrorist attack through twitter. *Information Systems Frontiers*, *13*(1), pp.33-43.

Peng, S., Cao, L., Zhou, Y., Ouyang, Z., Yang, A., Li, X., Jia, W. and Yu, S., 2021. A survey on deep learning for textual emotion analysis in social networks. *Digital Communications and Networks*.

Prabhu, A., Guhathakurta, D., Subramanian, M., Reddy, M., Sehgal, S., Karandikar, T., Gulati, A., Arora, U., Shah, R.R. and Kumaraguru, P., 2021. Capitol (Pat) riots: A comparative study of Twitter and Parler. *arXiv preprint arXiv:2101.06914*.

Rekik, A., Jamoussi, S. and Hamadou, A.B., 2019, September. Violent vocabulary extraction methodology: Application to the radicalism detection on social media. In *International Conference on Computational Collective Intelligence* (pp. 97-109). Springer, Cham.

Rezaeinia, S.M., Rahmani, R., Ghodsi, A. and Veisi, H., 2019. Sentiment analysis based on improved pre-trained word embeddings. *Expert Systems with Applications*, *117*, pp.139-147.

Saha, S., Yadav, J. and Ranjan, P., 2017. Proposed approach for sarcasm detection in twitter. *Indian Journal of Science and Technology*, *10*(25), pp.1-8.

Sharif, O. and Hoque, M.M., 2021. Tackling Cyber-Aggression: Identification and Fine-Grained Categorization of Aggressive Texts on Social Media using Weighted Ensemble of Transformers. *Neurocomputing*.

Sinha, S., Chen, H., Sekhon, A., Ji, Y. and Qi, Y., 2021. Perturbing inputs for fragile interpretations in deep natural language processing. *arXiv preprint arXiv:2108.04990*.

So, M.K., Tiwari, A., Chu, A.M., Tsang, J.T. and Chan, J.N., 2020. Visualizing COVID-19 pandemic risk through network connectedness. *International Journal of Infectious Diseases*, *96*, pp.558-561.

Sullivan, Bebe. 2022. Violence Lexicon [Unpublished lexicon]. Peace and War Center, Norwich University.

Tabassum, S., Pereira, F.S., Fernandes, S. and Gama, J., 2018. Social network analysis: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(5), p.e1256.

Udanor, C. and Anyanwu, C.C., 2019. Combating the challenges of social media hate speech in a polarized society: A Twitter ego lexalytics approach. *Data Technologies and Applications*.

Yaqub, U., Sharma, N., Pabreja, R., Chun, S.A., Atluri, V. and Vaidya, J., 2018, May. Analysis and visualization of subjectivity and polarity of Twitter location data. In *Proceedings of the 19th annual international conference on digital government research: governance in the data age* (pp. 1-10).

Zad, S., Jimenez, J. and Finlayson, M., 2021, August. Hell hath no fury? correcting bias in the nrc emotion lexicon. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)* (pp. 102-113).