Identifying Changes in Discourse on Mass-Shootings Using Reddit Comments

Neha Chanu
Information Science Program, University of Pittsburgh

Jing Pang
Information Science Program, University of Pittsburgh

Yang Wang
Department of Civil Engineering, University of Pittsburgh

Anthony Sicilia
Intelligent Systems Program, University of Pittsburgh
anthonysicilia@pitt.edu

ABSTRACT

Mass shootings, unfortunately, have become a commonplace tragedy in our day-to-day lives. In fact, in this year alone, there have been more than 380 shootings – more shootings than there have been days. Public opinion on this topic has of course seen substantial growth as well with a wealth of data found across popular social media sites. Hidden in this data is information that is valuable for politicians, other public figures, and businesses, who wish to craft topic-sensitive narratives and advertisements that engage their audience in the *correct* way. Yet, there have not been many studies addressing how we as a community hold discourse on this topic. To this end, we investigate how public discourse changes *over time* in the immediate aftermath of a shooting event. We use state-of-the-art topic modeling techniques on textual data from social media to define categories of discourse, and further, we present case studies on 4 recent shootings. We use these case studies to draw general conclusions of interest, for example, to advertising companies who may wish to avoid displaying sensitive content during polarizing times.

1 INTRODUCTION

According to the "Gun Violence Archive (GVA), which tracks every mass shooting in the country, a mass shooting [is defined] as any incident in which at least four people are shot, excluding the shooter" [11]. Unfortunately, mass shootings have become all too common in the U.S. In fact, "as of December 1, which is the 335th day of the year, there have been 385 mass shootings in the U.S." [11]. This means that there have been more mass shootings than the number of days in a year!

As social communities online have become more prevalent over the years, so has the discussion about mass shootings on social media. Social media platforms offer a rich resource for data about how people think and relate to mass shootings. Since there is a wealth of data available on various social media platforms on this topic, there is tremendous potential for competitive insights to be gained. The value of these insights would be beneficial for various stakeholders such as politicians and other public figures, large businesses such as gun manufacturers, small businesses such as those offering personal defence classes, and nonprofits such as Sandy Hook Promise or Protect our Schools. For politicians and other public figures, this insight is valuable to craft better-informed campaigns, diplomatic answers and topic sensitive narratives. For advertisers, these insights are valuable to craft advertisements that engage their audience in the *correct* way.

Yet, despite how common place mass shootings have become and the value of the information available on social media, there are not a lot of studies published about how we as a community discuss mass shootings and how this discussion changes over time. Thus, the main question of interest is *how does public discourse change over time in the immediate aftermath of a shooting event*? In this project, we narrowed the scope of social media data to one platform – Reddit, a network of communities based on people's interests where members can share just about anything (images, text posts, links etc). This content can be voted up or down by other members. Moreover, we analyzed this text data from Reddit using state-of-the-art topic modeling techniques to define categories of discourse and honed in on four case studies on four recent shootings to draw general conclusions.

2 RELATED WORKS

Not much in the research community has been done to investigate the public discourse on mass-shootings. Often, in social science, the primary concern is explanation and investigation of the cause of such events in our society. In particular, many works focus on the sudden increase of these events. Towers et al. [12], for example, fall back on a comparison to epidemiological reasoning, describing the surge of mass-shootings akin to a contagion. Meindl et al. [5] challenge this notion by instead describing the recent surge of mass-shootings as a by-product of "copy-cat" like behavior; specifically, they describe the surge in mass-shootings via generalized imitation. Yet, neither of these is particularly concerned with analyzing the public discourse on mass-shootings.

Discussion of mass-shootings in the media has, of course, been staggering. While we will not list all such investigations by reputable news outlets, we will focus on a few which are related to our study. Many new outlets, for example [9], look at how the public's views on gun-rights have shifted over time. Journalist investigations such as these share with our study a focus on public discourse as a function of time, but our study holds a more fine-grained lens (i.e. in the weeks following a shooting, not years). The investigation of Simonton [10], with USA Today, is particularly related to our study because it focuses on how discourse changes in the weeks immediately following a shooting. The investigation, in fact, identifies a decrease in the volume of discussion on social media about the shooting after only 10 days (with respect to the El Paso 2019 shooting, amongst others). While this fine-grained perspective is similar to our study, Simonton's investigation does not address the *what* of the conversation. In the context of a politician or business interested in employing analysis of public discourse

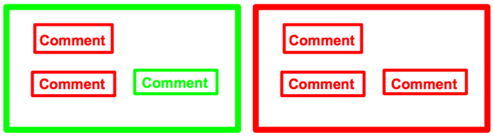

Figure 1: In the figure above, the left post is considered *about* the shooting, while the right is not. The left post contains a comment with some word which has the lexical stem *shoot* or *shooter*. All of the comments from the left post will be added to the dataset.

on mass-shootings for content-creation, this aspect is certainly of importance. It may not be the case that all topics encompassed within the public discourse on mass-shootings degrade in volume identically. Our work addresses the *what* of the conversation by clustering comments into topics or categories and viewing how these change over-time.

In more data-driven research communities such as computer and information science, research on analysis of public discourse on mass-shootings has only recently surfaced. Demszky et al. [3] provide a thorough analysis of twitter data surrounding 21 mass shooting events. Similar to our work, they cluster tweets into topics and analyze this grouping as well as performing a number of additional linguistic analyses on the twitter data. The important difference between our work and theirs is that Demszky et al. focus on discourse as a function of political affiliation, not time. Another recent study by Zhang et al. [17] does focus on this aspect of time, but limits their analyses to only two categories: sympathy and gun-policy. Our study uses unsupervised clustering techniques similar to Demszky et al. which effectively places no limit on the categories of discourse which we may consider; the categories are instead defined by what exists in the data.

In summary, in this review, we highlight three key components of our study that differentiate our work from the existing. We analyze the different topics of public discourse on mass-shootings as a function of time without imposing limitations on what those topics should be. While a number of works have investigated varying subsets of these aspects, to the best of our knowledge, ours is the first to consider all aspects jointly.

3 MODELING DISCOURSE OVER TIME

In the following, we discuss our approach to the problem of identifying categories and exploring their change over time. We describe the data used as well as our topic modeling approach with which we define clusters. We further describe how we define the semantic meaning of each of our clusters; i.e., the general, semantically meaningful category which the comments in the clusters represent. Finally, we present statistical analysis which verifies that topics do change over time.

While this section presents, in some detail, the methodology used in this study, the reader is directed to our code made openly available at github: https://github.com/knchanu/SCProject.

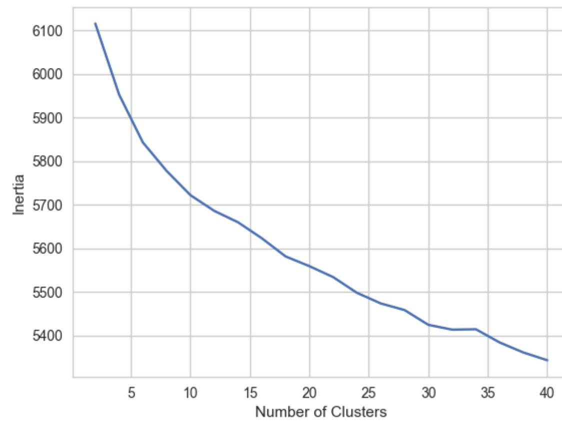

Figure 2: We use the elbow method to select the number of clusters. As can be seen in the above figure, there is a change in the slope at about k = 10. Since the Inertia measures cluster separation, we see a decreased rate of increase for cluster separation after k = 10. This is the *elbow*. We select k = 8 after some manual inspection of values near the elbow.

3.1 Dataset

Our dataset consists of Reddit comments from 4 different city-subreddits during the month immediately following a shooting-event. We restrict the comments considered in all of our analyses to be those contained in posts about the shooting-event of interest. Here, we define a post to be about the shooting if it contains any comment whose words have a lexical stem *shoot* or *shooter*. For example, the lexical stem of *shooting* is *shoot*. This process is further described in Figure 1. We remark that this approach may be overzealous, capturing comments which are not directly about the shooting, but we aim for high recall in order to cover as wide a range of categories as possible. In all analyses, as is common, we omit out-of-vocabulary words[1] and stop-words such as *the*, *an*, etc.[2]. We further remove URLs and text-formatting.

3.2 Topic Modelling

Our topic modeling approach is similar to that described in the work of Demzsky et al. [3] who, as mentioned, cluster tweets from social media to identify topics of public discourse. Our approach is based on the recent success of semantically meaningful word-vectors in Natural Language Processing research. The general idea is to encode each of our comments in a feature-rich vector space that effectively captures key aspects of semantic-meaning for each comment. Thus, standard machine learning algorithms (which operate on such vector spaces) can be applied to analyze these comments (i.e.; in our case, this ML algo. will be K-Means). To do this, we must

[1] Our vocabulary consists of those words for which there is a pre-trained Google News word vector [2]

[2] The full list of such words is composed of a predefined list which is contained in our GitHub repository as well as the list of words provided by NLTK a commonly used package for natural language processing [4]

Polarized	Neutral	Opinion (non-polarizing)	Locations/Facts	"Live" Updates	Law/Politics	Hostile	Positive

Table 1: Table of comment clusters found during topic-modelling.

chi2test	p-values
pittsburgh	4.204691809316447e-105
lasvegas	2.106010069285103e-10
orlando	8.686949657662132e-44
elpaso	1.800065118875468e-08

Table 2: p-values for χ^2-test of each city. Categories in contingency table are time-window and comment cluster. Significant p-values (across all cities) indicate a likely association between category and time. p-values are extremely low so we do not correct for multiple testing.

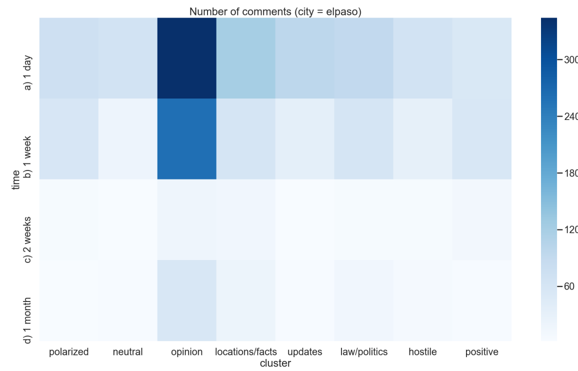

Figure 3: Raw number of comments over time for EL Paso Shooting

first translate the words in each comment into their respective word-vectors which we discuss next.

One of the first successful procedures for learning word-vectors was Skip-Gram with Negative Sampling proposed by Mikolov et al. [6]. We use pre-trained word-vectors learned via this technique and made publicly available by Google [2]. With each comment now represented by a set of semantically meaningful word-vectors, we combine these word-vectors with the method of Aurora et al. [1]. In their work, they show that their simple weighted average paired with subspace removal performs on par with state-of-the-art deep learning models. This produces a single vector which we use to represent each comment in the clustering.

We use a simple clustering method (K-Means) which attempts to minimize the sum of the squared-distances from each of K cluster centers[3]. We explored a large number of different K and chose 8 based on the elbow method [15] and manual inspection. As can be seen in Figure 2, there is a change in the magnitude of the slope around the point K=10. This indicates less substantial information gain as the number of clusters increase beyond this point. We further inspected manually to identify that for K > 8, the appearance of linguistically (but not semantically) meaningful clusters appeared; e.g., one such cluster consisted mainly of single word comments. As a result, we limited our number of clusters to 8.

Next, we define a semantic label for each cluster given by the above process; i.e., we assign some word or group of words that captures the general trend in meaning for the comments in each cluster. To do this, we inspect word-clouds restricted to comments contained in each individual cluster[4]. Using this visualization as a prior, we then manually inspect a random subset of comments from each cluster to verify our intuition and select a few key-words. The word-cloud visualizations for each cluster are provided in Appendix A. The semantic cluster labels we identify are given in Table 1.

3.3 Discourse Over Time

With topics/clusters assigned to all comments, we would like to identify if the topics people discuss are truly changing over time. To do this we define 4 time-windows: 1 day, 1 week, 2 weeks, and 1 month. Respectively, this corresponds to comments which took place: in the first 24 hours, between days 1 and 7, between days 7 and 14, and during the last 2 weeks. Now, we have two ways to categorize the comments: their clusters and their time-window. A common way to determine the association between two-categories is by constructing a contingency table and performing a χ^2-test [14]. If the p-value is significant, then we reject the null hypothesis that there is not an association between the categories. In performing this analysis for all 4 cities, we found significant p-values which are shown in Table 2. This indicates that discourse does in fact change over time and, further, supports our intuition that analysis of social media in the context of public discourse on mass-shootings requires consideration of *what* people are talking about.

While this statistical test does verify that topics are changing over-time, it is still important to identify how they are changing. To do this, we present case-studies across all 4 cites as well as comparisons of the cities with each other.

4 CASE STUDIES

This section will focus on finding potential relationships by analyzing several specific mass shootings in different cities. The reason for selecting these cities is based on several principles. The first principle is time range, the second principle is the degree of occurrence of events, and the third principle is the degree of impact of events. For example, Las Vegas was chosen because of its bad nature and serious impact, and Pittsburgh was chosen because the

[3]Our clustering algorithm was implemented by employing the commonly used data analysis package Scikit-Learn [8]. In addition to Scikit-Learn, we use the general ecosystem of data manipulation packages provided by SciPy [13]

[4]Our word clouds were generated using the Wordcloud package [7].

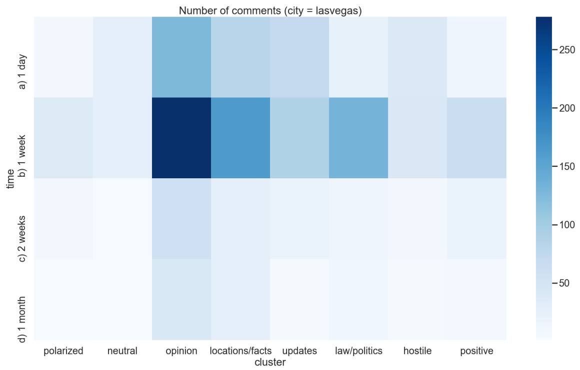

Figure 4: Raw number of comments over time for Las Vegas Shooting

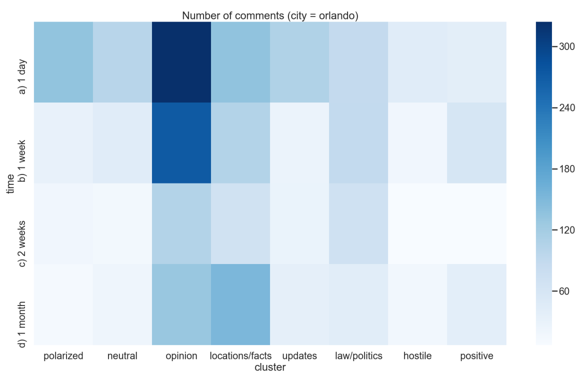

Figure 5: Raw number of comments over time for Orlando Shooting

Figure 6: Word clouds for different time windows; these clouds are specific to the Orlando shooting.

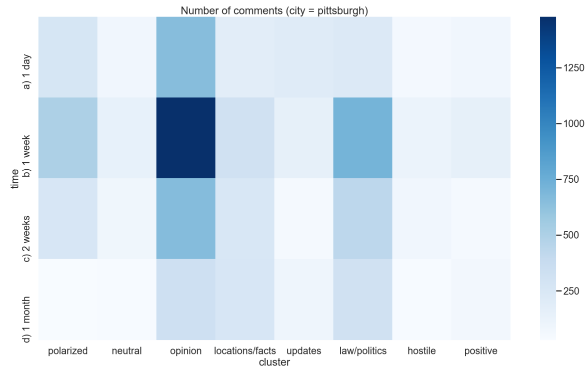

Figure 7: Raw number of comments over time for Pittsburgh Shooting

event was defined as a typical hatred case. In all cases, we provide some background on the shootings. The predominant source for background information was Wikipedia [16].

4.1 El Paso

The mass shooting happened in El Paso, Texas, on August 3 is the deadliest one in 2019. A 21-year-old gunman killed 22 people and injured 24 others at a Walmart store. After the investigation of this shooter's background and identity, the Federal Bureau of Investigation identified it as an act of domestic terrorism and hate crime.

A visual showing raw comment counts over time is available in Figure 3. In general, the duration of the comments lasted about one week. The popularity of the event hits the most at the beginning of the first day. However, there are a few categories that lasted more than one month. The comments on the opinion category reached a new peak in a one-month point, after comments reduced on the two-week point. Other comments on categories, such as polarized, normally follow the general conclusion. However, the comments on the location/facts and law/politics categories continuously posted for a month at a low-frequency level. Comments on the positive category only lasted for a week.

4.2 Las Vegas

The mass shooting happened in Las Vegas, Nevada, which is the deadliest by an individual in US history. The shooter caused enormous casualties, which killed 58 people and wounded 413 others. A bump stock modified a semi-automatic weapon to the deadliest weapon.

A visual showing raw comment counts over time is available in Figure 4. Similar to the El Paso case, the frequency of the comments lasted a week in all categories. However, the popularity of the topic hits the most after one week. Most of the comments were posted about the opinion category at all times. The frequency of the comments reached a maximum peak at the one-week point. A possible explanation is the discussion of a gun modification named bump stock. Currently, our data did not cover over four months. There would be an interesting point that if we have enough data to analyze the trend of comments on the four-month period because president Trump announced a ban on this modification after four months of this event. The comments on location/fact were continued for a month. The topic of the reason and place of this shooting drew much discussion during the time.

4.3 Orlando

Orlando is a city of leisure in Central Florida. At 2 o'clock am on June 12th 2016, a mass shooting happened in a nightclub in Orlando, leaving 49 deaths and 44 injuries. Born in New York, the killer Omar Mateen was a 29-year-old Afghan American. We chose the Orlando Shooting as a study case for the following reasons. Firstly, it is the second serious single mass shooting, following the 2017 Las Vegas shooting, in the American history. This shooting has a great influence on society. Secondly, the Orlando Shooting happened in 2016, in which the United States presidential election was held. Shootings are closely associated with political topics that all political parties and the public are concerned about in the presidential election, especially the gun control act. By analyzing this shooting, we want to explore the influence of the year of presidential election on comments on the shooting. At last, the loose gun control in Orlando is always an issue that the public actively talk about. For example, gun trade between individuals doesn't require background investigation in Orlando. We want to study the influences of different gun control policies on the public topic of shooting.

A visual showing raw comment counts over time is available in Figure 5. Generally speaking, its overall trend is similar to that of the previous two cities. Comments from the category of Opinion are more than those from other categories. However, comments in week 2 don't show an obvious advantage compared with comments in other time windows. In addition, polarized, hostile and positive comments reduce after week 1.

To better analyze changes in comments in different time windows, we built word clouds for different time windows as shown in Figure 6. Then, we obtained some interesting findings. Firstly, the word "gun" only appears in the period of "day". Later, it doesn't frequently appear. We analyzed the original comments and found that the public tended to discuss the shooting itself shortly after the incident. However, as time went by, the public were not concerned about the shooting itself anymore and shifted their focus to other relevant topics like gun control-related policy. Secondly, we found a lot of comments that contained the word "want" in week 2. For example, these comments express what a life they want or what policies they want. The surge in the number of similar topics is an interesting phenomenon.

As this incident happened in the year of presidential election, we wanted to find comments related to the government and election. Thus, we checked all time windows and found some relevant keywords like "vote" and "government" in the category of law in week 2. The occurrence of these words seemed a little odd, because those who cared about law and policy didn't discuss them until week 2. Given worlds like "want" in week 2, we believe this phenomenon may be a result of the election. As time went by after this incident, political parties might make use of this incident to attack their opponent's election policies or support their own election policies. Like this, words related to election sprang up in week 2.

4.4 Pittsburgh

On October 27th, 2018, a mass shooting happened in Pittsburgh, Pennsylvania, the United States, leaving 11 deaths and 6 injuries. It is said that this is the most serious attack on the Jewish People in American history.

A visual showing raw comment counts over time is available in Figure 7. Compared with the aforesaid three shootings, changes in comments on this shooting are different in some aspects. Firstly, we found that the number of comments in the category of "location" was bigger than that in the category of "updates". Secondly, compared with the aforesaid three cities, the number of comments on the Pittsburgh shooting in the category of "law" is the greatest in all time windows. In particular, the number of comments for the Pittsburgh shooting is much higher than that for the other three cities in week 2 and month. There are two reasons for such a phenomenon: one is the considerable number of comments on this shooting, and the other is the debate between the republican party and the democratic party. At last, we found that comments in the category of "polarized" lasted until week 2 and the comment duration for the Pittsburgh shooting was the longest among the four cities. As Pittsburgh is one of top 25 city subreddits, the great number of comments on this shooting may be a contributor to this phenomenon.

5 DEEP DIVE: COMPARISON ACROSS ALL CITIES

One common takeaway from the case-studies was the dissipation of comments among certain clusters or certain cities within the first two weeks after the shooting-event. To explore this further, we decided to investigate trends across all cities for each cluster. In particular, we look at what percentage of the cluster occurs during each of our time-windows (recall from Section 3: 1 day, 1 week, 2 weeks, 1 month). For a selection of clusters, this is visualized in Figure 8. What is easily identifiable in this visualization is the common peak at the 1 week time-window for Las Vegas and Pittsburgh. Some topics which exhibit this peak (*Polarized* comments, *Opinion* (non-polarized), *Law/Politics*, *Hostile*, *Positive*) would be of interest to politicians and advertisers navigating social media during the tumultuous time following a mass-shooting event.

Politicians who seek to engage their supporters on sensitive issues under the category of Law/Politics could potentially capitalize on this peak during the 1 week window by increasing presence on social media to strengthen their support-base. Likewise, they may instead be knowledgeable that certain topics are divisive within their support-base, promoting them to reduce their presence during this time period. Advertisers may seek to engage customers with content that is either targeted, sensitive to the situation, or even positions them for good publicity. For example, a guns-rights organization may capitalize on advertising during highly polarized times, while a company whose product is unrelated to the shooting-event may instead wish to donate during the most active conversation time to increase the spread of their viral publicity stunt. For both of these cases, our study indicates this active period may be within the first week.

While there does appear to be a trend that the first week is the most active time for a variety of important topics in Pittsburgh and

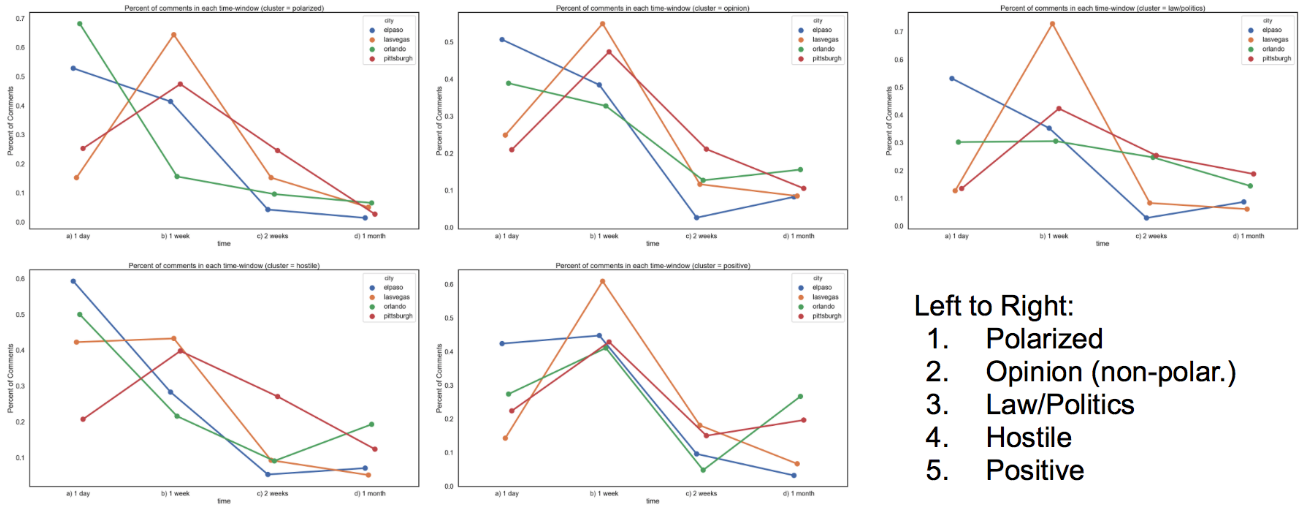

Figure 8: Plots show the proportion of each cluster in each time window for different cities. y-axes are percent of cluster. x-axes correspond to time-windows (1 day, 1 week, 2 weeks, 1 month). Blue is El Paso, green is Orlando, orange is Las Vegas, Red is Pittsburgh. Pittsburgh and Las Vegas share peaks around the one week mark for this selection of comment clusters.

Las Vegas, the subreddits of El Paso and Orlando to not quite follow this trend. To explain this difference we looked into contrasting differences between the cities (e.g. demographically, politically, etc.). In a way though, this is a red-herring. While it is true the predominant users of a city-subreddit will be those living in the city, this is only a subpopulation of the city. To avoid bias in our interpretation, we decided to instead look at only this subpopulation. One way to interpret features about the subpopulation within a city-subreddit is by constructing a subreddit-network visualization. The implementation we used (available at https://anvaka.github.io/sayit/) bases links in the network on the principle "users who posted to this subreddit also post to…". We present the resulting networks for each city subreddit in Appendix A. While these networks are indeed interesting, no features are clearly distinguished, so we refrain from speculating on any reasons why we might see the mentioned differences in topic distributions over time. We acknowledge that this inability to generalize to an entire city and interpret differences across cities is a limitation in this study. We discuss this and additional limitations in the next section.

6 LIMITATIONS

While the main question of this project was to identify how public discourse changes over time, in this project the focus was only on one social media platform – Reddit. Narrowing the scope to Reddit naturally excludes members of the community who do not use Reddit or may prefer other forms of social media such as Twitter or Facebook. Members of the community who don't prefer to post their opinions on online communities are also naturally excluded and thus, the opinions on Reddit may not have been as diverse. Another point is that the subreddit communities in two of the four cities analyzed as case studies, Pittsburgh and Orlando are much more active in general, than the subreddit communities in Las Vegas and El Paso. This means that it is likely that more opinions were captured from the subreddit communities of Pittsburgh and Orlando.

7 CONCLUSION

To conclude, the main goal of this project was to identify how public discourse changes over time in the immediate aftermath of a shooting event. To this end, state-of-the-art topic modeling techniques were applied on textual data from Reddit to define categories of discourse. Furthermore, case studies on four recent shootings were analyzed to draw general conclusions. The four recent shootings included the El Paso Shooting in 2019, Las Vegas Shooting in 2017, Orlando Shooting in 2016, and the Pittsburgh Shooting in 2018. From the case studies, a noticeable trend was around the 1 week mark after the shooting incident. At or around the 1 week mark, activity on Reddit was the highest. This insight is valuable for politicians & other public figures, businesses, nonprofits because showing advertisements around the 1 week mark is the best time to get the most bang for the buck. While the content of advertisements is important, the timing of when the advertisement is aired is also a crucial part of delivering the message across the audience.

REFERENCES

[1] Sanjeev Arora, Yingyu Liang, and Tengyu Ma. 2016. A simple but tough-to-beat baseline for sentence embeddings. (2016).

[2] Google Code. 2013. word2vec. https://code.google.com/archive/p/word2vec/. (2013).

[3] Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. Analyzing Polarization in Social Media: Method and Application to Tweets on 21 Mass Shootings. *arXiv preprint arXiv:1904.01596* (2019).

[4] Edward Loper and Steven Bird. 2002. NLTK: The Natural Language Toolkit. In *Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics - Volume 1 (ETMTNLP '02)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 63–70. https://doi.org/10.3115/1118108.1118117

[5] James N Meindl and Jonathan W Ivy. 2018. Reducing media-induced mass killings: Lessons from suicide prevention. *American behavioral scientist* 62, 2

(2018), 242–259.

[6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[7] A. Mueller. 2019. Wordcloud. https://github.com/amueller/word_cloud. (2019).

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[9] Nathaniel Rakich. 2019. How Views On Gun Control Have Changed In The Last 30 Years. https://fivethirtyeight.com/features/how-views-on-gun-control-have-changed-in-the-last-30-years/. (2019).

[10] Teghan Simonton. 2019. 31 people died in El Paso, Dayton mass shootings. Twitter data shows talk faded in just 10 days. https://www.usatoday.com/story/news/investigations/2019/08/15/twitter-data-shows-we-quickly-stop-talking-about-mass-shootings-like-dayton-el-paso/1984989001/. (2019).

[11] Jason Sliverstein. 2019. There have been more mass shootings than days this year. https://www.cbsnews.com/news/mass-shootings-2019-more-mass-shootings-than-days-so-far-this-year/. (2019).

[12] Sherry Towers, Andres Gomez-Lievano, Maryam Khan, Anuj Mubayi, and Carlos Castillo-Chavez. 2015. Contagion in mass killings and school shootings. *PLoS one* 10, 7 (2015), e0117259.

[13] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake Vand erPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors. 2019. SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python. *arXiv e-prints*, Article arXiv:1907.10121 (Jul 2019), arXiv:1907.10121 pages. arXiv:cs.MS/1907.10121

[14] Wikipedia. 2019. Chi-squared test. https://en.wikipedia.org/wiki/Chi-squared_test. (2019).

[15] Wikipedia. 2019. Elbow method (clustering). https://en.wikipedia.org/wiki/Elbow_method_(clustering). (2019).

[16] Wikipedia. 2019. Wikipedia Articles on Mass Shooting. https://en.wikipedia.org/. (2019).

[17] Yini Zhang, Dhavan Shah, Jordan Foley, Aman Abhishek, Josephine Lukito, Jiyoun Suk, Sang Jung Kim, Zhongkai Sun, Jon Pevehouse, and Christine Garlough. 2019. Whose Lives Matter? Mass Shootings and Social Media Discourses of Sympathy and Policy, 2012–2014. *Journal of Computer-Mediated Communication* 24, 4 (2019), 182–202.

A WORD CLOUDS AND CITY SUBREDDIT NETWORKS

Figure 9: Word cloud visualization for *Polarized* comments cluster

Figure 10: Word cloud visualization for *Neutral* comments cluster

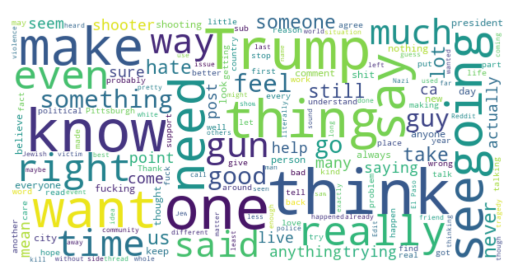

Figure 11: Word cloud visualization for *Opinion* (non-polarizing) comments cluster

Figure 15: Word cloud visualization for *Hostile* comments cluster

Figure 16: Word cloud visualization for *Positive* comments cluster

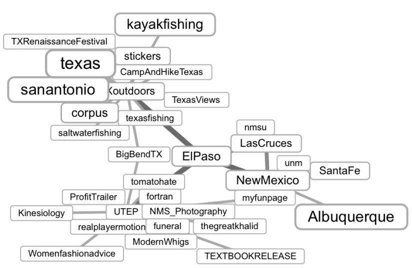

Figure 17: City subreddit visualization for El Paso

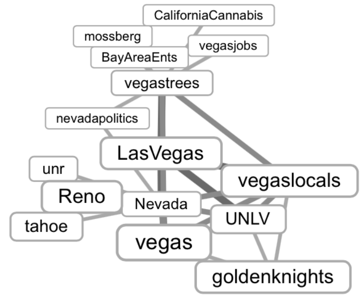

Figure 18: City subreddit visualization for Las Vegas

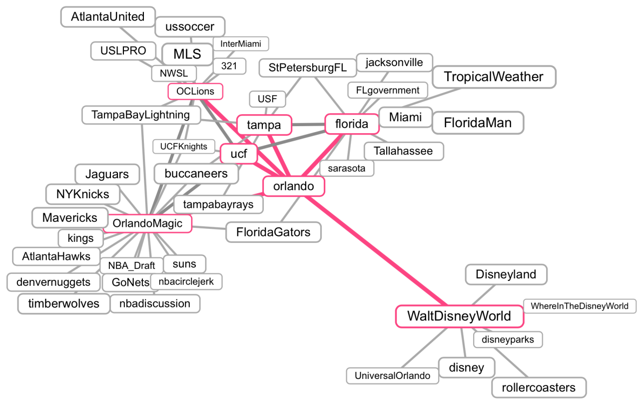

Figure 19: City subreddit visualization for Orlando

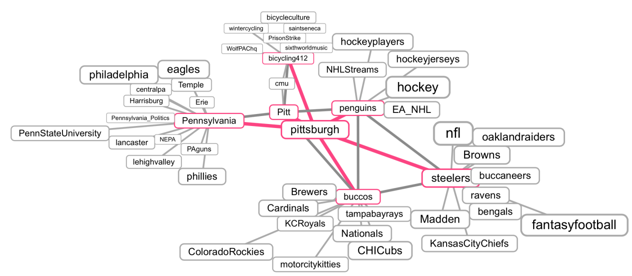

Figure 20: City subreddit visualization for Pittsburgh