Analyzing Polarization in Social Media:
Method and Application to Tweets on 21 Mass Shootings

Dorottya Demszky1 Nikhil Garg1 Rob Voigt1 James Zou1
Matthew Gentzkow1 Jesse Shapiro2 Dan Jurafsky1
1Stanford University 2Brown University
{ddemszky, nkgarg, robvoigt, jamesz, gentzkow, jurafsky}@stanford.edu
jesse_shapiro_1@brown.edu

Abstract

We provide an NLP framework to uncover four linguistic dimensions of political polarization in social media: topic choice, framing, affect and illocutionary force. We quantify these aspects with existing lexical methods, and propose clustering of tweet embeddings as a means to identify salient topics for analysis across events; human evaluations show that our approach generates more cohesive topics than traditional LDA-based models. We apply our methods to study 4.4M tweets on 21 mass shootings. We provide evidence that the discussion of these events is highly polarized politically and that this polarization is primarily driven by partisan differences in framing rather than topic choice. We identify framing devices, such as grounding and the contrasting use of the terms “terrorist” and “crazy”, that contribute to polarization. Results pertaining to topic choice, affect and illocutionary force suggest that Republicans focus more on the shooter and event-specific facts (news) while Democrats focus more on the victims and call for policy changes. Our work contributes to a deeper understanding of the way group divisions manifest in language and to computational methods for studying them.1

1 Introduction

Elites, political parties, and the media in the US are increasingly polarized (Layman et al., 2010; Prior, 2013; Gentzkow et al., forthcoming), and the propagation of partisan frames can influence public opinion (Chong and Druckman, 2007) and party identification (Fiorina and Abrams, 2008).

Americans increasingly get their news from internet-based sources (Mitchell et al., 2016), and political information-sharing is highly ideologically segregated on platforms like Twitter (Conover et al., 2011; Halberstam and Knight, 2016) and Facebook (Bakshy et al., 2015). Prior NLP work has shown, e.g., that polarized messages are more likely to be shared (Zafar et al., 2016) and that certain topics are more polarizing (Balasubramanyan et al., 2012); however, we lack a more broad understanding of the many ways that polarization can be instantiated linguistically.

This work builds a more comprehensive framework for studying linguistic aspects of polarization in social media, by looking at topic choice, framing, affect, and illocutionary force.

1.1 Mass Shootings

We explore these aspects of polarization by studying a sample of more than 4.4M tweets about 21 mass shooting events, analyzing polarization within and across events.

Framing and polarization in the context of mass shootings is well-studied, though much of the literature studies the role of media (Chyi and McCombs, 2004; Schildkraut and Elsass, 2016) and politicians (Johnson et al., 2017). Several works find that frames have changed over time and between such events (Muschert and Carr, 2006; Schildkraut and Muschert, 2014), and that frames influence opinions on gun policies (HaiderMarkel and Joslyn, 2001). Prior NLP work in this area has considered how to extract factual information on gun violence from news (Pavlick et al., 2016) as well as quantify stance and public opinion on Twitter (Benton et al., 2016) and across the web (Ayers et al., 2016); here we advance NLP approaches to the public discourse surrounding gun violence by introducing methods to analyze other linguistic manifestations of polarization.

1.2 The Role of the Shooter’s Race

We are particularly interested in the role of the shooter’s race in shaping polarized responses to these events. Implicit or explicit racial biases can be central in people’s understanding of social
problems (Drakulich, 2015); in the mass shooting context, race is a factor in an event’s newsworthiness (Schildkraut et al., 2018) and is often mentioned prominently in media coverage, particularly when the shooter is non-white (Mingus and Zopf, 2010; Park et al., 2012). Duxbury et al. (2018) find that media representations of white shooters disproportionately divert blame by framing them as mentally ill while representations of non-white shooters are more frequently criminalized, highlighting histories of violent behavior.

The important question remains as to how polarized ideologies surrounding race take shape on forums such as Twitter. Therefore, in all of the analyses throughout this paper we consider the race of the shooter as a potential factor. We note that in the 21 shooting events we study, shootings in schools and places of worship are overwhelmingly carried out by white perpetrators, so we cannot fully disentangle the effect of race from other factors.

2 Data: Tweets on Mass Shootings

Data collection. We compiled a list of mass shootings between 2015 and 2018 from the Gun Violence Archive.\textsuperscript{2} For each, we identified a list of keywords representative of their location (see Appendix A). Given Twitter search API’s limitations on past tweets, we retrieved data from a Stanford lab’s archived intake of the Twitter firehose.\textsuperscript{3}

For each event, we built a list of relevant tweets for the two weeks following the event. A tweet is relevant if it contained at least one of the event’s location-based representative keywords and at least one lemma from the following list: “shoot”, “gun”, “kill”, “attack”, “massacre”, “victim”. We filtered out retweets and tweets from users who have since been deactivated. We kept those 21 events with more than 10,000 tweets remaining. For more details see Appendix A.

Partisan assignment. We estimate the party affiliation of users in the dataset from the political accounts they follow using a method similar to that of Volkova et al. (2014), which takes advantage of homophily in the following behavior of users on twitter (Halberstam and Knight, 2016). We compile a list of Twitter handles of US Congress members in 2017, the 2016 presidential and vice presidential candidates, and other party-affiliated pages.\textsuperscript{4} We label a user as a Democrat if they followed more Democratic than Republican politicians in November 2017, and as a Republican if the reverse is true. For each event, 51–72% of users can be assigned partisanship in this way; to validate our method we compare state averages of these inferred partisan labels to state two-party vote shares, finding a high correlation (Figure 1).\textsuperscript{5}

3 Quantifying Overall Polarization

We begin by quantifying polarization (equivalently, partisanship) between the language of users labeled Democrats and Republicans after mass shooting events. We establish that there is substantial polarization, and that the polarization increases over time within most events.

3.1 Methods

Pre-processing. We first build a vocabulary for each event as follows. Each vocabulary contains unigrams and bigrams that occur in a given event’s tweets at least 50 times, counted after stemming via NLTK’s SnowballStemmer and stopword removal.\textsuperscript{6} We refer to these unigrams and bigrams collectively as tokens.

\textsuperscript{2}https://www.gunviolencearchive.org
\textsuperscript{3}With the exception of the two most recent shootings in Pittsburgh and Thousand Oaks, for which we collected tweets real time via the Twitter API.
\textsuperscript{4}See Appendix B.1 for the complete list.
\textsuperscript{5}We performed the sanity check for all partisan users with a valid US state as part of their geo-location (∼350k users).
\textsuperscript{6}Stopword list is provided in Appendix A.1
Measure of partisanship. We apply the leave-out estimator of phrase partisanship from Gentzkow et al. (forthcoming). Partisanship is defined as the expected posterior probability that an observer with a neutral prior would assign to a tweeter’s true party after observing a single token drawn at random from the tweets produced by the tweeter. If there is no difference in token usage between the two parties, then this probability is .5, i.e. we cannot guess the user’s party any better after observing a token.

The leave-out estimator consistently estimates partisanship under the assumption that a user’s tokens are drawn from a multinomial logit model. The estimator is robust to corpus size. The leave-out estimate for Fort Lauderdale is

\[
\hat{\pi}_{LO} = \frac{1}{2} \left( \frac{1}{|D|} \sum_{i \in D} \hat{q}_i \cdot \hat{\rho}_{-i} + \frac{1}{|R|} \sum_{i \in R} \hat{q}_i \cdot (1 - \hat{\rho}_{-i}) \right)
\]

where \( \hat{q}_i = c_i / m_i \) is the vector of empirical token frequencies for tweeter \( i \), with \( c_i \) being the vector of token counts for tweeter \( i \) and \( m_i \) the sum of token counts for tweeter \( i \); \( \hat{\rho}_{-i} = (\hat{q}^D \otimes (\hat{q}^D + \hat{q}^R)) \) is a vector of empirical posterior probabilities, excluding speaker \( i \) and any token that is not used by at least two speakers. Here we let \( \otimes \) denote element-wise division and \( \hat{q}^G = \sum_{i \in G} c_i / \sum_{i \in G} m_i \) denote the empirical token frequency of tweeters in group \( G \). The estimator thus captures two intuitive components of partisanship: between-group difference (posterior probability for each feature), and within-group similarity (dot-product between the feature vector of each speaker and that of their group).

User-level measures. As the above leave-out estimator represents the average of user-level polarization values, we take the user-level dot product \( (\hat{q}_i \cdot \hat{\rho}_{-i}) \) as an estimate of the polarization of user \( i \)'s language. We consider the correlation of this value and the number of politicians a user follows in total and from their preferred party.

3.2 Results and Discussion

Overall polarization. As Figure 2 shows, the discussion of each event\(^3\) is highly polarized: values range from .517 to .547. For comparison, this three-dimensional scatter plot shows the leave-out estimate of phrase partisanship for users from Fort Lauderdale.

Post-event polarization. To see how polarization changes at the event level, we computed the leave-out estimate for each of the first 10 days following the events (see Figure 3). An event-day level regression of partisanship on days since the event suggests a slight increase in post-event polarization across events (slope = .002, \( p < 0.05 \)). Fitting separate regressions, we find that the five events with the steepest increase in polarization are Burlington (slope = .03, \( p < 0.05 \)), Orlando (slope = .006, \( p < 0.001 \)), Las Vegas (slope = .003, \( p < 0.001 \)), Chattanooga (slope = .003, \( p < 0.05 \)) and Roseburg (slope = .003, \( p < 0.05 \)).

Are the changes in the leave-out score due to different users tweeting at different times or due to the same users becoming more or less political? We found that while on average only \(~ 10\%\) of users tweeted on multiple days (SD = 5%) across the events, these users contribute \(~ 28\%\) of the tweets (SD = 15%). After removing these users from the leave-out estimation, we found that the temporal patterns remain with the same statistical significance, providing one piece of evidence

\(^3\)For all the experiments using the leave-out estimator, we exclude Fort Lauderdale for which we only have tweets for the first day after the event; polarization is most dominant a few days after each event, making it incomparable. For reference, the leave-out estimate for Fort Lauderdale is .51.
that changes in polarization are not due to changes within users who tweet on multiple days.

**User-level polarization.** We estimated a linear regression of the leave-out score on the total number of followed politicians and the number from the user’s preferred party, with controls for event indicators. The estimates imply that, fixing the total number of followed politicians, one more followed politician from one’s preferred party is associated with a decrease of .009 SD in the leave-out. Fixing the number of followed politicians from the user’s preferred party, one more followed politician is associated with an increase of .009 SD in the leave-out.

### 4 Topics and Framing

Topic choice can be a tool for agenda-setting by establishing what an author or institution deems worthy of discussion (McCombs, 2002), and works in NLP have used topic modeling as an approach to measure this effect (Tsur et al., 2015; Field et al., 2018). The strategy of highlighting particular aspects within topics as a means of framing (Entman, 2007) has also been quantified in the NLP literature (Boydston et al., 2013; Card et al., 2015; Naderi and Hirst, 2017).

Previous work largely focuses on the relation between topic and framing in the news media; we study social media, proposing methods to identify general, non-event-specific topics and to quantify between- and within-topic polarization.

#### 4.1 Methods

**Topic assignment.** Our goal is to induce topics that are salient in our narrow domain and comparable across events. This presents a challenge for traditional topic modeling approaches, since the discourse surrounding these events is inherently tied to concrete aspects of the events that tend to covary with topic usage, like location, setting, and demographics of the shooter and victims.

We build on the ability of vector space models to represent higher-level semantics to develop our own embedding-based topic assignment approach, comparing it with two traditional LDA-based methods: MALLET and the Biterm Topic Model (BTM) (Yan et al., 2013); BTM was developed specifically for tweets. For all of these methods, we first randomly sample 10k tweets from each event forming our subset $S$ of all tweets $T$; then, we create a vocabulary $V$ of word stems that occur at least ten times in at least three events within $S$ (~2000 word stems) and remove all stems from $T$ not part of $V$. Sampling is crucial for encouraging event-independent topics given the large disparity among event-level tweet counts (the largest event, Orlando, has 225× more tweets than the smallest event, Burlington).

For the embedding-based approach, we:

1. Train GloVe embeddings (Pennington et al., 2014) on $V$ based on 11-50k random samples of tweets from each event.
2. Create sentence embeddings $e_t$, $\forall t \in T$ using Arora et al. (2017)’s method, by computing the weighted average $v_t$ of the embeddings of stems within $t$ and removing $v_t$’s projection onto the first principal component of the matrix the rows of which are $v_t$, $\forall t \in S$. Stem weights are set to be inversely proportional to their frequencies in $S$.
3. Jointly cluster the embeddings $e_t$, $\forall t \in S$ via k-means using cosine distance and assign all tweet embeddings $e_t$, $\forall t \in T$ to the centroids to which they are closest.

We also trained MALLET and BTM on $S$ and used the resulting models to infer topics for all tweets in $T$, assigning each tweet to its highest probability topic. Henceforth, we use $d$ to mean cosine distance for k-means and probabilities for MALLET and BTM.

A manual inspection found that about 25% of the tweets are either difficult to assign to any topic or they represent multiple topics equally. To filter out such tweets, for each tweet we looked at the ratio of $d$ to its closest and second closest topic and removed tweets that have ratios higher than the

---

*http://mallet.cs.umass.edu/topics.php

*This sample is different from $S$ as it includes more tweets to increase data size, which is important for training the embeddings, where the slightly disproportional representation of events is less problematic.
Table 1: Our eight topics (with their average proportions across events) and nearest-neighbor stem embeddings to the cluster centroids. Topic names were manually assigned based on inspecting the tweets.

75th percentile (calculated at the model-level).

To compare the models, we ran two MTurk experiments: a word intrusion task (Chang et al., 2009) and our own, analogically defined tweet intrusion task, with the number of topics \( k \) ranging between 6-10. Turkers were presented with either a set of 6 words (for word intrusion) or a set of 4 tweets (for tweet intrusion), all except one of which was close (in terms of \( d \)) to a randomly chosen topic and one that was far from that topic but close to another topic. Then, Turkers were asked to pick the odd one out among the set of words/tweets. More details in Appendix D.

We find that our model outperforms the LDA-based methods with respect to both tasks, particularly tweet intrusion — see Figure 4. This suggests that our model both provides more cohesive topics at the word level and more cohesive groupings by topic assignment. The choice of \( k \) does not yield a significant difference among model-level accuracies. However, since \( k = 8 \) slightly outperforms other \( k \)-s in tweet intrusion, we use it for further analysis. See Table 1 for nearest neighbor stems to each topic and Appendix C.2 for example tweets.

Measuring within-topic and between-topic partisanship. Recall that the leave-out estimator from Section 3.1 provides a measure of partisanship. The information in a tweet, and thus partisanship, can be decomposed into which topic is discussed, and how it’s discussed.

To measure within-topic partisanship for a particular event, i.e. how a user discusses a given topic, we re-apply the leave-out estimator. For each topic, we calculate the partisanship using only tweets categorized to that topic. Then, overall within-topic partisanship for the event is the weighted mean of these values, with weights given by the proportion of tweets categorized to each topic within each event.

Between-topic partisanship is defined as the expected posterior that an observer with a neutral prior would assign to a user’s true party after learning only the topic — but not the words — of a user’s tweet. We estimate this value by replacing each tweet with its assigned topic and applying the leave-out estimator to this data.

4.2 Results

Figure 5 shows that for most events within-topic is higher than between-topic partisanship, suggesting that while topic choice does play a role in phrase partisanship (its values are meaningfully higher than .5), within-topic phrase usage is significantly more polarized. Linear estimates of the relationship between within and between topic partisanship and time show that while within-topic polarization has increased over time, between-topic polarization remains stable.

Partisanship also differs by topic, and within
Figure 6: Las Vegas within-topic polarization in the days after the event. The bar charts show the proportion of each topic in the data at a given time.

days after a given event. Figure 6 shows polarization within topics for 9 days after Las Vegas. We find that solidarity has the lowest and shooter’s identity & ideology the highest polarization throughout; polarization in most topics increases over time and news has the steepest increase. Similar patterns are present after Orlando (Figure 17 in Appendix I). Measuring polarization of topics for other events over time is noisy, given the sparsity of the data, but overall within-topic polarization is consistent: the most polarized topics on average across events are shooter’s identity & ideology (.55) and laws & policy (.54), where people are apparently polarized about both why an event happened and what to do about it. Fact- and sympathy-based topics display less polarization: news (.51), victims & location (.52), solidarity (.52) and remembrance (.52).

As shown in Figure 7, investigation, news, and shooter’s identity & ideology are more likely to be discussed by Republicans and laws & policy and solidarity more likely to be discussed by Democrats across events.11 Topics preferred by Republicans seem to relate more to the shooter than to the victims, while topics preferred by Democrats seem to relate more closely to the victims. The shooter’s race appears to play a role in topic preference: if the shooter is white, Democrats become more likely to focus on shooter’s identity & ideology and laws & policy and Republicans on news and investigation than if the shooter is a person of color.

5 Specific Framing Devices

In the previous section, we show that topic-level framing and topic choice are different dimensions of polarization. We now look at the specific terms and types of grounding that are used as partisan framing devices, contributing to polarization.

5.1 Methods

Partisan tokens. We estimate the partisanship of tokens via their event-level log odds ratio of Democrats relative to Republicans (based on the vocabularies we create in Section 3.1). We compare these estimates across events.12

Grounding. We study whether there is polarization in which prior tragic events are referenced in the context of a particular mass shooting. We compile a list of keywords representing major events of mass violence in the US in the past two decades and kept those that were mentioned at least 100 times by Democrat or Republican users. For all tweets for each event in our dataset, we counted the mentions of past context events. For example, in the following tweet posted after Las Vegas: “Dozens of preventable deaths should not be the cost of living in America. Stand up to the #NRA. #LasVegasShooting #SandyHook #Charleston”, Sandy Hook and Charleston are the context events. Finally, we calculated the partisan log odds ratio of each context event.

11 p-values are calculated using a one sample t-test, comparing with zero: shooter’s identity & ideology ($p < 0.05$), investigation ($p < 0.001$), laws & policy ($p < 0.1$), news ($p < 0.05$), solidarity ($p < 0.001$).

12 To compare the partisanship of tokens at the event- and topic-level, we also z-score the log odds ratios (Monroe et al., 2008) within events and topics; the most partisan tokens are reported in Appendix F. The reason why we do not z-score for the between-event comparison is because data verbosity disproportionately affects the range of the values’ magnitudes. Note that the signs of values — which we focus on for the cross-event comparison — are not affected by z-scoring.
5.2 Results

We focus on the partisanship of the term “terrorist” and “crazy”, which exhibit differential patterns across events based on the shooter’s race.13 “Terrorist” is always more likely to be used by Democrats than Republicans in events where the shooter is white, and the opposite is true when the shooter is a person of color (Figure 8); “crazy” is more likely used by Republicans if the shooter is white than if they are a person of color and the opposite is true (although the pattern is weaker) when a shooter is white.

These findings support related work (Perea, 1997; Delgado and Stefancic, 2017) discussing binary conceptualization of race in the US, and its influence in determining whether a shooter’s mental health or aspects of their identity are discussed. However, the fact that the influence of race flips completely for Democrats and Republicans is a striking result that calls for further exploration.

The partisanship of contextual grounding also corroborates our finding that the shooter’s race influences how people conceptualize a certain event. Our results in Figure 9 suggest a few key takeaways: the two most frequently employed context events are both highly partisan (Sandy Hook for Democrats and 9/11 for Republicans); shootings at schools and places of worship are more likely to be brought up by Democrats; Democrats are more likely to reference events with white shooters, while Republicans are more likely to reference those with shooters who are people of color.

6 Affect

Affect is intimately tied to ideological reasoning (Redlawsk, 2002; Taber et al., 2009), and so emotional expression represents another semantic layer relevant to polarization (Iyengar et al., 2012; Suhay, 2015). Others have shown that emotion words can help detect political ideology on Twitter (Preotiuc-Pietro et al., 2017) and that emotive political tweets are more likely to be shared (Stieglitz and Dang-Xuan, 2012). Here, we employ a lexicon-based approach to measure valence (positive and negative) and five basic emotion categories (disgust, fear, trust, anger, and sadness).

6.1 Methods

Since word-affect associations are highly domain dependent, we tailored an existing affect lexicon, the NRC Emotion Lexicon (Mohammad and Turney, 2013), to our domain via label propagation (Hamilton et al., 2016).

Specifically, we stem all the words in the lexicon and select 8-10 representative stems per emotion category that have an association with that emotion in the context of mass shootings. For each emotion category, we compute pairwise cosine distances between the GloVe embedding of each in-vocabulary stem and the representative stems for that emotion, and include the 30 stems with the lowest mean cosine distances. The resulting lexicons can be found in Appendix E.

We use these lexicons to measure the partisanship of each affect category. For each event and each party we aggregate stem frequencies per emotion category. We then calculate the partisan log odds ratio of each category for each event.

6.2 Results

The log odds ratio of each affect category is shown in Figure 10. These findings suggest that positive sentiment, sadness and trust are more likely to be expressed by Democrats across events, while fear and disgust are more likely to be expressed by Republicans, particularly when the shooter is a person of color. Anger, trust and negative sentiment is similarly likely to be expressed by both parties.14

Our results about fear and disgust accord with existing literature on emotion and political ideology: conservatives score higher than liberals on subjective measures of fear (e.g. Jost et al., 2017; Federico et al., 2009; Hibbing et al., 2014) and disgust sensitivity is also associated with political conservatism (e.g. Inbar et al., 2009, 2012).

\[ p\] -values are calculated using a one sample t-test, comparing to zero: anger (\( p \approx 0.43 \)), disgust (\( p \approx 0.06 \)), fear (\( p < 0.001 \)), negative (\( p \approx 0.2 \)), positive (\( p < 0.001 \)), sadness (\( p < 0.02 \)), trust (\( p \approx 0.07 \)).
Figure 9: The partisanship of events of mass violence when used as a context for a given mass shooting. The position of the events on the line represents their partisan log odds ratio (Democrat < 0 < neutral < Republican). The pie charts indicate the proportion of Democrat and Republican users’ tweets that involve this “context” event.

Figure 10: The log odds of each emotion category in our lexicon (one observation represents one event).

7 Modality and Illocutionary Force

Modality is a lexical category concerned with necessity and possibility (Kratzer, 2002; Fintel, 2006). In the aftermath of a tragic event, people seek solutions, a process that often involves reflecting on what should have happened or should happen now or in the future (e.g. to prevent such events). We hypothesize that the use of modals in our data gives insight into the kind of (illocutionary) acts (Austin, 1962) the users are performing via their tweets, such as calling for action, assigning blame, expressing emotions, and stating facts.

7.1 Methods

We work with all forms of the four most frequent necessity modals in our data — should, must, have to and need to. For each, we quantify its partisanship via its partisan log odds ratio. We also annotate a random sample of 200 tweets containing modals to see whether they are indeed used in contexts that imply calls for change / action (e.g. ‘We must have gun control!’) and / or to express the user’s mental state about the event, such as despair or disbelief (e.g. ‘Why do people have to die?’).

Table 2: Random sample of tweets with modals. Only one of the eight (Ex. 6) tweets expresses ideas traditionally associated with conservative ideology.

7.2 Results

Table 2 shows a random sample of tweets containing some form of either should, must, have to, or need to. More collocations, as well as their parti- sanship, can be found in Appendix H. These examples, as well as our annotation, support the hypothesis that these modals are primarily used to call for action. Of the 200 modal uses, ~78% express calls for change/action, ~40% express the user’s mental state.

We also compute the representation $p^m_x$ of each modal $m$ in each topic $x \in X$ via $(f^m_x / \sum_{x' \in X} f^m_{x'}) / (f_x / \sum_{x' \in X} f_{x'})$, where $f_x$ is the number of tweets from topic $x$, and $f^m_x$ the number of those also containing $m$. We find that that modals are over-represented in the laws & policy topic (see Figure 11). This evidence suggests that calls for policy change — especially gun control, based on annotated samples — are a dominant subset of calls for action.

$^{15}$Other uses are primarily epistemic ones (e.g. ‘The suspect must be mentally ill’).
Conclusion

We show that inspecting polarization on social media from various angles can shed light on salient phenomena pertinent to group divisions. Applying the leave-out estimator of phrase partisanship to data on mass shootings, we find that reactions to these events are highly polarized politically.

To disentangle topic choice and topic-level framing — two phenomena that contribute to polarization — we introduce a tweet-clustering approach. By sampling, requiring words in the vocabulary to appear in multiple events and relying on the abstraction of a vector space model, we generate cohesive topic representations that are robust to disparities among event-level vocabularies and tweet counts. Human evaluation shows that our method outperforms LDA-based approaches.

Our induced topics suggest that Republicans preferentially discuss topics about the shooter’s identity and ideology, investigation and news, while Democrats preferentially discuss solidarity and policy-related topics. We also find that the setting and the shooter’s race interact with polarization. For example, Democrats are more likely to contextualize any mass shooting among school shootings and call white shooters “terrorists” than are Republicans, who in turn are more likely to liken any shooting to other violent events perpetrated by people of color — whom they are more likely to call “terrorist” than are Democrats. Moreover, Democrats are more likely to frame the shooter as mentally ill when they are a person of color and Republicans when they are white.

We also demonstrate that looking at affect and illocutionary force can help us understand users’ polarized responses to these tragic events: Republicans are more likely to express fear and disgust than are Democrats, while Democrats are more likely to express sadness and positive sentiment, to make calls for action and assign blame.

Polarization is a multi-faceted phenomenon: in this paper we present a set of measures to study these different facets through the lens of language. We show that these measures provide convergent evidence, creating a clearer picture of the complex ideological division permeating public life.

Acknowledgements. We thank Jure Leskovec and Adrijan Bradaschia for data, and Cleo Condoravdi, Chris Potts, Linda Ouyang, David Ritzwoller and Frank Yang for helpful feedback. We are grateful for the support of the Stanford Cyber Initiative, the Melvin and Joan Lane Stanford Graduate Fellowship (to D.D.), NSF GRF DGE-114747 (to N.G.), the Michelle and Kevin Douglas Stanford Interdisciplinary Graduate Fellowship (to R.V.), NSF grant CRII 1657155 (to J.Z.), the Stanford Institute for Economic Policy Research and the Knight Foundation (to M.G.) and the Brown University Population Studies and Training Center (to J.S.).
References


A Data

Table 3 contains properties of the data. Figure 12 contains the distribution of partisan tweets for each event.

**Event-specific keywords.** We use the following location-specific keywords (case insensitive) to find tweets on the events:

- Chattanooga: chattanooga, military recruitment center
- Roseburg: umpqua, roseburg
- Colorado Springs: colorado springs, coloradosprings, planned parenthood, planned-parenthood
- San Bernardino: san bernardino, san-bernardino
- Kalamazoo: kalamazoo
- Orlando: orlando, pulse nightclub
- Dallas: dallas
- Baton Rouge: baton rouge, batonrouge
- Burlington: burlington, cascade mall
- Fort Lauderdale: lauderdale
- Fresno: fresno
- San Francisco: ups, san francisco
- Las Vegas: vegas, route91, harvest festival, harvestfestival, mandalay bay
- Thornton: thornton, walmart, denver
- Sutherland Springs: sutherland springs, sutherlandsprings
- Parkland: parkland, marjory stoneman
- Nashville: nashville, waffle house
- Santa Fe: santa fe, santafe
- Annapolis: annapolis, capital gazette
- Pittsburgh: pittsburgh, treeoflife, tree of life
- Thousand Oaks: thousand oaks, thousand-oaks
<table>
<thead>
<tr>
<th>Event city / town</th>
<th>State</th>
<th>Specific location</th>
<th>Date</th>
<th>No. victims</th>
<th>Race / ethnicity of shooter</th>
<th>No. tweets</th>
<th>No. partisan tweets</th>
<th>No. Dem tweets</th>
<th>No. Rep tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chattanooga</td>
<td>TN</td>
<td>Military Recruitment Center</td>
<td>7/16/15</td>
<td>7</td>
<td>Middle Eastern</td>
<td>29575</td>
<td>20709</td>
<td>5925</td>
<td>14794</td>
</tr>
<tr>
<td>Roseburg</td>
<td>OR</td>
<td>Umpqua Community College</td>
<td>10/1/15</td>
<td>18</td>
<td>Mixed</td>
<td>18078</td>
<td>11503</td>
<td>6419</td>
<td>5086</td>
</tr>
<tr>
<td>Colorado Springs</td>
<td>CO</td>
<td>Planned Parenthood clinic</td>
<td>11/29/15</td>
<td>12</td>
<td>White</td>
<td>95843</td>
<td>39119</td>
<td>66814</td>
<td>13105</td>
</tr>
<tr>
<td>San Bernardino</td>
<td>CA</td>
<td>Inland Regional Center</td>
<td>12/2/15</td>
<td>35</td>
<td>Middle Eastern</td>
<td>70491</td>
<td>45819</td>
<td>20798</td>
<td>25021</td>
</tr>
<tr>
<td>Kalamazoo</td>
<td>MI</td>
<td>Multiple</td>
<td>2/20/16</td>
<td>8</td>
<td>White</td>
<td>10996</td>
<td>6807</td>
<td>4350</td>
<td>2457</td>
</tr>
<tr>
<td>Orlando</td>
<td>FL</td>
<td>Pulse nightclub</td>
<td>2/18/16</td>
<td>102</td>
<td>Middle Eastern</td>
<td>1831082</td>
<td>872022</td>
<td>450784</td>
<td>421238</td>
</tr>
<tr>
<td>Dallas</td>
<td>TX</td>
<td>Black Lives Matter protest</td>
<td>7/7/16</td>
<td>16</td>
<td>Black</td>
<td>260373</td>
<td>144205</td>
<td>64628</td>
<td>79577</td>
</tr>
<tr>
<td>Roseburg</td>
<td>OR</td>
<td>Umpqua Community College</td>
<td>10/1/15</td>
<td>18</td>
<td>Mixed</td>
<td>46126</td>
<td>29015</td>
<td>12010</td>
<td>16996</td>
</tr>
<tr>
<td>Burlington</td>
<td>WA</td>
<td>Cascade Mall</td>
<td>9/23/16</td>
<td>5</td>
<td>Middle Eastern</td>
<td>10986</td>
<td>6807</td>
<td>4350</td>
<td>2457</td>
</tr>
<tr>
<td>Fort Lauderdale</td>
<td>FL</td>
<td>Fort Lauderdale airport</td>
<td>1/6/17</td>
<td>11</td>
<td>Hispanic</td>
<td>12525</td>
<td>7154</td>
<td>6073</td>
<td>4124</td>
</tr>
<tr>
<td>Fresno</td>
<td>CA</td>
<td>downtown</td>
<td>4/18/17</td>
<td>3</td>
<td>Black</td>
<td>8868</td>
<td>6128</td>
<td>1377</td>
<td>4751</td>
</tr>
<tr>
<td>San Francisco</td>
<td>CA</td>
<td>UPS store</td>
<td>6/14/18</td>
<td>3</td>
<td>Asian</td>
<td>10287</td>
<td>6623</td>
<td>4350</td>
<td>2293</td>
</tr>
<tr>
<td>Nevada</td>
<td>NV</td>
<td>Route 91 Harvest Festival</td>
<td>10/1/17</td>
<td>604</td>
<td>White</td>
<td>1286399</td>
<td>726739</td>
<td>315343</td>
<td>411396</td>
</tr>
<tr>
<td>Thornton</td>
<td>CO</td>
<td>Walmart</td>
<td>11/1/17</td>
<td>3</td>
<td>White</td>
<td>14341</td>
<td>9170</td>
<td>5527</td>
<td>3641</td>
</tr>
<tr>
<td>Sutherland Springs</td>
<td>TX</td>
<td>Texas First Baptist Church</td>
<td>11/2/17</td>
<td>46</td>
<td>White</td>
<td>154076</td>
<td>106220</td>
<td>32515</td>
<td>53560</td>
</tr>
<tr>
<td>Parkland</td>
<td>FL</td>
<td>Marjory Stoneman Douglas High School</td>
<td>2/14/18</td>
<td>31</td>
<td>White</td>
<td>272499</td>
<td>186570</td>
<td>113856</td>
<td>72714</td>
</tr>
<tr>
<td>Nashville</td>
<td>TN</td>
<td>Waffle House</td>
<td>2/22/18</td>
<td>8</td>
<td>White</td>
<td>30600</td>
<td>24328</td>
<td>14886</td>
<td>9720</td>
</tr>
<tr>
<td>Santa Fe</td>
<td>CA</td>
<td>Santa Fe High School</td>
<td>7/18/18</td>
<td>23</td>
<td>White</td>
<td>73621</td>
<td>43968</td>
<td>26784</td>
<td>16184</td>
</tr>
<tr>
<td>Annapolis</td>
<td>MD</td>
<td>Capital Gazette</td>
<td>6/28/18</td>
<td>7</td>
<td>White</td>
<td>27715</td>
<td>18464</td>
<td>11363</td>
<td>6605</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>PA</td>
<td>Tree of Life Synagogue</td>
<td>10/27/18</td>
<td>18</td>
<td>White</td>
<td>59925</td>
<td>30920</td>
<td>22735</td>
<td>14385</td>
</tr>
<tr>
<td>Thousand Oaks</td>
<td>CA</td>
<td>Borderline Bar and Grill</td>
<td>11/19/18</td>
<td>23</td>
<td>White</td>
<td>117935</td>
<td>62819</td>
<td>40328</td>
<td>22484</td>
</tr>
</tbody>
</table>

Table 3: Data properties.

![Figure 12: Distribution of partisan tweets for each event.](image-url)
A.1 Stopwords

no, noone, nobody, nowhere, nothing, nor, not, none, non, a, able, about, above, according, accordingly, across, actually, after, afterwards, again, against, all, allow, allows, almost, alone, along, already, also, although, always, am, among, amongst, an, and, another, any, anybody, anyhow, anyone, anything, anyway, anyways, anywhere, apart, appear, appreciate, appropriate, are, aren, around, as, aside, ask, asking, associated, at, available, away, awfully, b, be, became, because, become, becomes, becoming, been, before, beforehand, behind, being, believe, below, beside, besides, best, better, between, beyond, both, brief, but, by, c, came, can, cannot, cant, cause, causes, certain, certainly, changes, clearly, co, com, come, comes, concerning, consequently, consider, considering, contain, containing, contains, corresponding, could, couldnt, couldnt, couldve, course, currently, d, definitely, described, despite, did, didnt, different, do, dont, don, does, doesnt, doing, done, down, downwards, during, e, each, edu, eg, eight, either, else, elsewhere, enough, entirely, especially, et, etc, even, ever, every, everybody, everyone, everything, everywhere, ex, exactly, example, except, f, far, few, fifth, first, five, followed, following, follows, for, former, formerly, forth, four, from, further, furthermore, g, get, gets, getting, given, gives, go, goes, going, gone, got, gotten, greetings, h, had, hadnt, happens, hardly, has, hasnt, hasn, have, havent, haven, having, he, hes, hell, hello, help, hence, her, here, hereafter, hereby, herein, hereupon, hers, herself, hi, him, himself, his, hither, hopefully, how, howbeit, however, i, im, ive, ie, if, i, ignored, immediate, in, inasmuch, inc, indeed, indicate, indicated, indicates, inner, insofar, instead, into, inward, is, isn, it, its, itself, j, just, k, keep, keeps, kept, know, knows, known, l, last, lately, later, latterly, least, less, lest, let, like, liked, likely, little, ll, look, looking, looks, ltd, m, mainly, many, may, maybe, me, mean, meanwhile, merely, might, mightve, more, moreover, most, mostly, much, must, mustn, mustnt, mustve, my, myself, n, name, namely, nd, near, nearly, necessary, need, neednt, needs, neither, never, nevertheless, new, next, nine, nor, normally, novel, now, o, obviously, of, off, often, oh, ok, okay, old, on, once, one, ones, only, onto, or, other, others, otherwise, ought, our, ours, ourselves, out, outside, over, overall, own, p, particular, particularly, per, perhaps, placed, please, plus, possible, presumably, probably, provides, q, que, quite, qv, r, rather, rd, re, really, reasonably, regarding, regardless, regards, relatively, respectively, right, s, said, same, saw, say, saying, says, second, secondly, see, seeing, seem, seemed, seeming, seems, self, selves, sensible, sent, serious, seriously, seven, several, shall, she, shell, shes, should, shouldnt, shouldd, shouldve, since, six, so, some, somebody, somehow, someone, something, sometime, sometimes, somewhat, somewhere, soon, sorry, specified, specify, specifying, still, sub, such, sup, sure, t, take, taken, tell, tends, th, than, thank, thanks, thanx, that, thats, the, their, theirs, them, themselves, then, thence, there, thereafter, thereby, therefore, therein, theres, thereupon, these, they, theyre, theyve, think, third, this, thorough, thoroughly, those, though, three, through, through-out, thru, thus, to, together, too, took, toward, towards, tried, tries, truly, try, trying, twice, two, u, un, under, unfortunately, unless, unlikely, until, unto, up, upon, us, use, used, useful, uses, using, usually, uucp, v, value, various, ve, very, via, viz, vs, w, want, wants, was, way, we, welcome, well, went, were, what, whatever, when, whence, whenever, where, whereafter, whereas, whereby, wherein, whereupon, wherever, whether, which, while, whither, who, whoever, whole, whom, whose, why, will, willing, wish, with, within, without, wonder, would, wouldnt, wouldve, x, y, yes, yet, you, your, youre, youve, your, yours, yourself, yourselves, z, zero
B Partisanship Assignment

B.1 Political Twitter Handles

B.2 Partisan Assignment Sanity Check

In our sanity check, we exclude DC, as 1) it is not an official US state and 2) there the population of users (including politicians and media) is expected to differ from the voting population. Figure 13 gives DC values.

![Figure 13: This plot was generated the same way as Figure 1, except that it also includes DC, which shows that it is an outlier in our data due to the fact that there are many Republican politicians and media outlets who are affiliated on Twitter with DC, while DC’s voting population tends to be Democratic.]

B.3 Russian account presence

We find no substantial presence of Russian accounts in the tweet set that we used for analysis, after all pre-processing. We use the list of Russian accounts identified by Twitter and banned in November 2017.\textsuperscript{18} This list is likely to be an underestimate of the true presence of foreign influence, but it nevertheless provides some indication of such activity. Table 4 contains a breakdown by event; we exclude the events that occurred after the accounts were banned. Orlando had one account that tweeted 115 times, and Vegas had 4 that tweeted a total of 70 times.

\textsuperscript{18} Available at https://www.recode.net/2017/11/2/16598312/russia-twitter-trump-twitter-deactivated-handle-list
## Table 4: Tweets by Russian accounts

<table>
<thead>
<tr>
<th>Event</th>
<th>Number of Russian accounts</th>
<th>Number of tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baton Rouge</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Burlington</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Chattanooga</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Colorado Springs</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dallas</td>
<td>4</td>
<td>39</td>
</tr>
<tr>
<td>Fort Lauderdale</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fresno</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Kalamazoo</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Orlando</td>
<td>1</td>
<td>115</td>
</tr>
<tr>
<td>Roseburg</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>San Bernardino</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>San Francisco</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sutherland Springs</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Thornton</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vegas</td>
<td>4</td>
<td>70</td>
</tr>
</tbody>
</table>
Table 5: The highest probability stems per topic (for $k = 8$) for MALLET.

<table>
<thead>
<tr>
<th>Topic 10 Highest Probability Stems</th>
<th>Topic 10 Highest Probability Stems</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. gun, shoot, law, control, people, shooter, church, house, stop, school</td>
<td>1. victim, familial, prayer, today, thought, life, communities, violent, pray, heart</td>
</tr>
<tr>
<td>2. school, shoot, high, gun, student, parkland, kid, shooter, texa, kill</td>
<td>2. police, report, suspect, shooter, dead, rifle, shot, cap hit, break, multiple</td>
</tr>
<tr>
<td>3. shoot, victim, family, prayer, thought, pray, today, kill, gun, heart</td>
<td>3. shoot, mass, vegas, news, gunman, thousand, las, dead, california, die</td>
</tr>
<tr>
<td>4. shoot, police, dead, shooter, report, suspect, people, shot, kill, airport</td>
<td>4. attack, san, trump, orlando, terrorist, plan, bernardino, call, terror, obama</td>
</tr>
<tr>
<td>5. shoot, shooter, attack, terrorist, gun, terrorist, san, trump, plan, call</td>
<td>5. shoot, church, shot, live, airport, time, day, texa, fire, talk</td>
</tr>
<tr>
<td>6. shoot, vegas, mass, las, thousand, victim, kill, california, bar, church</td>
<td>6. kill, people, house, shot, shoot, shooter, shot, black, police, man</td>
</tr>
<tr>
<td>7. victim, shoot, trump, house, flag, cap hit, honor, half, president, staff</td>
<td>7. shoot, mass, vegas, news, gunman, thousand, las, dead, california, die</td>
</tr>
<tr>
<td>8. kill, people, white, house, shoot, shooter, shot, black, police, man</td>
<td>8. victim, family, prayer, today, thought, life, communities, violent, pray, heart</td>
</tr>
</tbody>
</table>

Table 6: The highest probability stems per topic (for $k = 8$) for BTM.

<table>
<thead>
<tr>
<th>Topic 10 Highest Probability Stems</th>
<th>Topic 10 Highest Probability Stems</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. victim, family, prayer, today, thought, life, communities, violent, pray, heart</td>
<td>1. victim, family, prayer, today, thought, life, communities, violent, pray, heart</td>
</tr>
<tr>
<td>2. police, report, suspect, shooter, dead, rifle, shot, cap hit, break, multiple</td>
<td>2. police, report, suspect, shooter, dead, rifle, shot, cap hit, break, multiple</td>
</tr>
<tr>
<td>3. shoot, mass, vegas, news, gunman, thousand, las, dead, california, die</td>
<td>3. shoot, mass, vegas, news, gunman, thousand, las, dead, california, die</td>
</tr>
<tr>
<td>4. attack, san, trump, orlando, terrorist, plan, bernardino, call, terror, obama</td>
<td>4. attack, san, trump, orlando, terrorist, plan, bernardino, call, terror, obama</td>
</tr>
<tr>
<td>5. shoot, church, shot, live, airport, time, day, texa, fire, talk</td>
<td>5. shoot, church, shot, live, airport, time, day, texa, fire, talk</td>
</tr>
<tr>
<td>6. kill, people, house, shot, stop, man, guy, black, murder, colorado</td>
<td>6. kill, people, house, shot, stop, man, guy, black, murder, colorado</td>
</tr>
<tr>
<td>7. shooter, dada, cop, killer, media, blame, make, show, cnn, post</td>
<td>7. shooter, dada, cop, killer, media, blame, make, show, cnn, post</td>
</tr>
<tr>
<td>8. gun, school, high, control, law, parkland, student, nra, kid, arm</td>
<td>8. gun, school, high, control, law, parkland, student, nra, kid, arm</td>
</tr>
</tbody>
</table>

### C Topic Model Outputs

#### C.1 Topic Words

Table 5 and Table 6 show the highest probability words per topic (for $k = 8$) for MALLET and BTM, respectively.

#### C.2 Topic Tweets

We present a sample of tweets belonging to one of the 8 topics assigned by our model.

**News.**

- HAPPENING NOW: Multiple people wounded in shooting at Colorado Walmart
- 3 people are reported dead in #FtLauderdale shooting
- UPDATE: Baton Rouge 2 Officers confirmed dead 7 officers have been shot in total. Details are slowly coming in.
- San Francisco police responding to reports of a shooting at a UPS facility, according to multiple reports.
- BREAKING: Police confirm several fatalities in #Annapolis newsroom shooting and multiple people seriously injured. The suspect is in custody. @SkyNewsAust

**Investigation.**

- Alleged synagogue shooter charged in deaths of 11 people
- Michigan prosecutor: Suspect in Kalamazoo rampage that killed 6 admitted to shootings.
- #COShooting #PlannedParenthood It’s over...suspect walked out and they apprehended him-no altercation in arrest. Suspect turned himself in
- Capital Gazette shooter has been identified as Jarrod W. Ramos, 38, who had previously filed a defamation lawsuit against the paper and a columnist in 2012.
- Waffle House gunman’s father facing charges for GIVING him gun used to kill four

**Shooter’s identity & ideology.**

- HATE CRIME : Fresno Islamic Killer Referred to White People as “Devils”
- To say that extremist Muslims represent all Muslims is like saying the gunmen in Colorado springs represents all Christians
- So who has #BatonRouge blamed for this shooting Imperialist Obama BLM Movement The Black Man Or the isolated mentally ill white lone wolf
- Again, the Lafayette white killer is a “lone wolf” but the Chattanooga Arab killer is an entire religion. Aggressively insane troll logic.
- Who is surprised that the San Bernardino shooters were radicalized Muslims?

**Victims & location.**

- 12 Killed in California Shooting; Gunman Targeted Bar in Thousand Oaks
- Synagogue shooting in Pittsburgh: what we know
- Navy Veteran Who Survived Afghanistan Dies in Las Vegas Shooting
- Aaron Feis, who died in the mass shooting at Marjory Stoneman Douglas High, was praised by college recruiters and former players for helping dozens of high school athletes land a chance to play in college.
- Texas church, site of deadly massacre, to be demolished (Via WLTX 19)
Laws & policy.

- This Parkland Tragedy was completely a security meltdown in or security officials. 100% preventable. Gun laws had nothing to do with this massacre. But gun control could have diminish the carnage. Two different things...

- NRA allowed acts like #Chattanooga to become commonplace. Their lobbying permits people on the terror watch list to buy guns. Remember that.

- I will not just #prayforsutherlandsprings, today I vote for @PhilMurphyNJ and stronger gun control laws in NJ #ElectionDay #GunControlNow

- Again the mental health flags were waving about shooter in Santa Fe no one acted. By now, if ones in charge cared and wanted to protect students, Every School would have had a security assessment and have hardened access points. Can never stop these but can make it harder.

- This is a mental health issue, security issue AND a gun issue. Our government has taken action on NONE of these to protect our students. So clearly we are not being heard, and our kids are being executed. What do we do now? #EnoughIsEnough #DoSomething #SantaFe #WeMustBeHeard

Solidarity.

- Our thoughts and prayers go out the victims and their families involved in the #Chattanooga tragedy 2nite.

- Praying for the loved ones of the victims of the shooting yesterday at the synagogue in Pittsburgh.

- Our prayers go out to the victims, families, friends & everyone affected in #SanBernardino #BeSafe #BeAware

- My heart goes out to the friends and family of the victims in that parkland shooting :-(

- Our hearts goes to #UmpquaCommunityCollege in the #oregonshooting #ChristCentredEnt is praying 4 u #AllLivesMatter

Remembrance.

- @HarfordCC will honor victims @UmpquaCC TODAY by observing a National Moment of Silence 1:45PM in Quad. Students should attend #IamUCC

- Photo: US flag flying at half-staff at White House for victims of Roseburg, Ore., school shooting

- More than 100 Romans gather in solidarity against hate, honor victims of Orlando shooting

- Trump Denies Request To Lower Flags To Honor Capital Gazette Shooting Victims #DouchebagDon #CrookedTrump @FLOTUS @realDonaldTrump

- Live from #ChattanoogaShootings memorial next on #11Alive

Other.

- I’m sure coming from you, this has nothing to do with the Fresno shooting?

- I realize the dude from the Waffle House shooting did a miraculous thing disarming the shooter, but if I see the gaping wound on his forearm one more time I’m gonna lose my mind.

- The only thing that stops a bad guy with a gun, is a good guy with a gun!!! #Kalamazoo

- The little clip I saw of what DT just said about #PittsburghShooting #PittsburghSynagogueShooting was ok (I guess)

- But when is someone gonna reassure me a planned parenthood attack won’t happen
D Topic Model Evaluation

In the next two subsections we describe the tasks which we crowdsourced to compare the three topic models: MALLET, BTM and our embedding-based model.

D.1 Word intrusion

Our word intrusion task is the same as is described in Chang et al. (2009). Our topic-word distance metric for MALLET and BTM is probability (we use the exact topic-word matrix that these models output) and for our model it is cosine distance. We created 2850 experimental items (i.e. sets of words) with the following procedure:

1. Sample a model \( M \), an \( k \) (between 6 – 10) and a topic \( x \) ranging between 1 – \( k \).
2. For a given choice of \( M \), \( k \) and \( x \),
   (a) sample 5 words among the closest 10 words to \( x \).
   (b) sample one word that is among the 5% of the furthest words from \( x \) but also among the 5% of the closest words to another topic.
3. Shuffle the words.

Turkers were asked to pick the odd word out among a set of words. Each MTurk HIT consisted of 6 sets of words.

D.2 Tweet intrusion

Since we only assign one topic to each tweet, we would like to evaluate how coherent the tweets are that are assigned to the same topic. Therefore, we define the tweet intrusion task analogously to word intrusion. Our distance metric in this case is the ratio of proximities (probability for MALLET and BTM and cosine for our model) between the closest topic and second closest topic — this value quantifies the proximity to the closest topic as well as how uniquely close that topic is (in contrast to the second topic). We created 1800 experimental items via the following procedure:

1. Sample a model \( M \), an \( k \) (between 6 – 10) and a topic \( x \) ranging between 1 – \( k \).
2. For a given choice of \( M \), \( k \) and \( x \),
   (a) sample 3 tweets among the closest 1% of tweets to \( x \).
   (b) sample one tweet that is among the 1% of the furthest tweets from \( x \) but also among the 1% of the closest tweets to another topic.
3. Shuffle the tweets.

Turkers were asked to pick the odd tweet out from these tweets. Each MTurk HIT consisted of three sets of tweets.
E Emotion Lexicon

The following words were the final stems in our emotional lexicon.

**positive** love, friend, pray, thought, affect, bless, god, pleas, communiti, hope, stand, thank, help, condol, will, comfort, time, strong, work, support, effect, strength, feel, peac, word, rest, give, great, action, good

**negative** hate, violenc, hatr, of, evil, tragedi, will, word, attack, sad, feel, anger, murder, shoot, massacr, want, need, pain, kill, grie, crime, ignor, victim, lost, grief, senseless, tragic, fear, loss, sick

**sadness** senseless, loss, tragedi, lost, devast, sad, love, grie, horrif, terribl, pain, violenc, condol, broken, hurt, feel, victim, mourn, horrifi, will, grieve, ach, suffer, sick, kill, aw, sicken, evil, massacr, mad

**disgust** disgust, sick, shame, ignor, wrong, blame, hell, ridicul, idiot, murder, evil, coward, sicken, feel, disgrac, slaughter, action, bad, insan, attack, pathet, outrag, polit, terrorist, mad, damn, lose, shit, lie, asshol

**anger** gun, will, murder, kill, violenc, wrong, shoot, bad, death, attack, feel, shot, action, arm, idiot, crazi, crimini, terrorist, mad, hell, crime, blame, fight, ridicul, insan, shit, die, threat, terror, hate

**fear** danger, threat, fear, arm, gun, still, shooter, attack, feel, fight, hide, murder, shot, shoot, bad, kill, chang, seriou, violenc, forc, risk, defend, warn, govern, concern, fail, polic, wrong, case, terrorist

**trust** school, like, good, real, secur, show, nation, don, protect, call, teacher, help, law, great, save, true, wonder, respons, sad, answer, person, feel, safe, thought, continu, love, guard, church, fact, support

The following words were used as seeds to generate this lexicon, as described in the main text.

**positive** love, donat, heart, thought, strength, bless, solidar

**negative** hatr, hate, grie, grief, wrong

**sadness** mourn, sadden, grie, grief, sad, suffer, affect, broken, senseless, loss, heartbroken
Most Partisan Phrases

F.1 Most Partisan Phrases Overall

We list the 20 most Democrat and Republican unigrams and bigrams that occur at least 100 times in tweets about a particular event. The number in brackets indicates the z-score of the log odds of these words (Monroe et al., 2008) — values with an absolute value greater than 2 are significantly partisan.

Chattanooga. Most Republican phrases: obama (7.41), gun free (5.89), zone (5.70), free (5.69), free zone (5.53), flag (5.33), #tcot (5.33), #chattanoogattack (5.13), #wakeupamerica (4.94), islam (4.69), #obama (4.01), #islam (3.83), attack (3.67), #gunfreezon (3.66), lower (3.53), arm (3.44), workplac (3.32), white hous (3.23), workplac violenc (3.21)

Most Democrat phrases: blame georg (-8.62), bush invad (-8.62), invad iraq (-8.60), base lie (-8.55), georg bush (-8.53), war base (-8.51), lie happen (-8.51), georg (-8.35), iraq war (-8.31), iraq (-7.77), lie (-7.45), charleston (-6.99), mass (-6.82), #lafayett (-6.48), happen (-6.19), charlestonshoot (-5.96), blame (-5.40), #gunsens (-5.09)

Roseburg. Most Republican phrases: obama (8.02), #2a (6.28), #obama (6.01), gun free (5.37), #tcot (5.32), free (5.22), christian (5.09), zone (5.08), chicago (5.06), shooter (4.78), free zone (4.75), #gunfreezon (4.04), agenda (3.95), religion (3.85), train (3.79), liber (3.73), #christian (3.71), secur (3.45), guard (3.40), skarlato (3.39)

Most Democrat phrases: #gunsens (-4.69), heart (-4.59), gun nut (-3.92), fuck (-3.83), mass (-3.82), gun violenc (-3.81), violenc (-3.80), nra (-3.54), thought (-3.47), nut (-3.28), school (-3.22), #gunviol (-3.10), countri (-3.08), chang (-3.04), congress (-3.02), love (-2.99), vigil (-2.86), mass shoot (-2.75), protest (-2.74), america (-2.63)

Colorado Springs. Most Republican phrases: babi (13.62), liber (11.39), kill babi (8.77), polic (8.46), kill (8.30), left (7.42), office (7.01), bank (6.97), babi kill (6.84), lib (6.42), obama (6.34), #activeshoot (6.23), #tcot (6.13), activ (6.12), parenthood kill (6.01), report (5.99), injur (5.90), #break (5.81), activ shooter (5.75), plan (5.72)

Most Democrat phrases: terrorist (-13.88), terror (-9.71), attack (-9.52), terrorist attack (-9.09), white (-8.33), #plannedparenthoodshoot (-7.63), #plannedparenthood (-7.29), gop (-7.04), domest (-6.67), candid (-6.40), #gopdeb (-6.32), #standwithpp (-6.27), attack #plannedparenthood (-6.23), women (-5.84), radic (-5.74), #gop (-5.39), defund (-5.05), christian (-5.02), call (-4.89), aliv (-4.73)

San Bernardino. Most Republican phrases: obama (12.59), attack (12.08), #tcot (9.33), terror (9.26), islam (9.20), terrorist attack (8.65), muslim (7.48), liber (6.94), gop (6.82), climat (6.62), climat chang (6.34), blame (5.91), #obama (5.79), islam terrorist (5.65), #pjnet (5.61), #wakeupamerica (5.35), workplac violenc (5.19), foxnew (5.05), call (5.02), vet (4.89)

Most Democrat phrases: mass (-13.70), mass shoot (-10.50), #gusens (-10.35), shoot (-7.97), disabl (-7.17), gop (-5.89), development (-5.82), fuck (-5.73), development disabl (-5.67), #gopdeb (-5.64), heart (-5.37), center (-5.34), thought (-5.29), day (-4.96), action (-4.93), domest (-4.78), #gunviol (-4.73), congress (-4.50), sick (-4.40), normal (-4.32)

Kalamazoo. Most Republican phrases: michigan (4.02), #break (3.39), ap (3.37), barrel (3.03), counti (2.98), cracker (2.98), cracker barrel (2.93), suspect (2.80), polic (2.71), area (2.56), dead (2.53), random shoot (2.49), mich (2.43), year girl (2.33), 14 year (2.24), victim (2.22), shoot (2.22), charg (2.20), 7 (2.20), counti michigan (2.18)

Most Democrat phrases: mass (-4.37), white (-4.11), mass shoot (-4.00), terrorist (-3.63), america (-3.39), #gunviol (-3.08), white man (-2.72), mental (-2.72), gun violenc (-2.60), coverag (-2.58), kalamazoo shoot (-2.52), ill (-2.42), mental ill (-2.38), unarm (-2.37), guy (-2.27), countri (-2.27), white male (-2.24), black (-2.20), violenc (-2.02), talk (-1.88)

Orlando. Most Republican phrases: islam (59.38), attack (53.07), terrorist (48.13), obama (47.01), #tcot (39.27), fbi (38.32), terror (37.21), blame (36.48), terror attack (35.32), terrorist attack (33.95), mateen (32.23), hillari (32.04), jihad (31.77), shooter (31.72), isi (31.63), radic (31.07), democrat (30.47), killer (29.49), liber (29.01), islam terror (28.69)

Most Democrat phrases: victim (-54.39), love (-40.34), heart (-34.47), hate (-34.40), violenc (-31.13), communiti (-29.11), #loveislov (-28.11), gun violenc (-27.06), lgbt (-26.84), lgbtq (-26.44), donat (-24.12), #prayforlando (-24.04), mass (-21.53).
Dallas. **Most Republican phrases:** obama (23.09), bluelivesmatt (17.78), offic (12.76), obama (11.52), white (10.33), polic offic (10.13), hillari (10.10), kill white (9.77), racist (9.09), foxnew (9.00), shot (8.98), offic shot (8.95), #tcot (8.81), hate crime (8.67), democrat (8.51), blame gun (8.49), crime (8.47), white cop (8.46), cop (8.30), 5 (8.03)

**Most Democrat phrases:** #altonsterl (-21.29), #philandocastil (-21.13), #altonsterl #philandocastil (-16.32), guy gun (-15.96), good guy (-15.26), violenc (-14.83), gun (-14.54), open carri (-13.00), guy (-12.63), carri (-12.47), open (-10.82), #gunviol (-10.54), good (-10.53), stop (-10.23), #philandocastil #altonsterl (-10.07), nra (-10.02), gun violenc (-9.80), #nra (-9.23), #disarmh (-8.64), answer (-8.54)

Baton Rouge. **Most Republican phrases:** #bluelivesmatt (8.69), obama (8.13), islam (5.87), nation islam (5.62), #obama (5.17), shot (5.06), cop killer (5.06), killer (5.05), terrorist (4.99), nation (4.74), offic shot (4.63), hillari (4.34), #backtheblu (4.25), #tcot (4.05), offic (3.99), thug (3.96), islam member (3.79), 3 (3.53), #trumppence16 (3.45), democrat (3.41)

**Most Democrat phrases:** gun (-8.88), violenc (-8.34), gun violenc (-6.65), #nra (-6.19), guy gun (-5.77), #altonsterl (-5.57), open carri (-5.48), assault (-5.48), weapon (-5.30), good guy (-5.14), blame (-4.88), citizen (-4.87), assault weapon (-4.84), carri (-4.74), race relat (-4.74), relat (-4.71), open (-4.61), guy (-4.58), civilian (-4.47), #enough (-4.46)

Burlington. **Most Republican phrases:** cetin (5.12), arcan cetin (5.06), arcan (5.01), muslim (4.72), turkish (4.57), vot (4.37), turkey (4.32), hispan (4.14), immigr (4.04), citizen (3.79), hillari (3.64), elect (3.45), immigr turkey (3.19), turkish muslim (3.14), id (3.11), ed arcans (3.05), id ed (3.03), shooter id (3.00), citizen vote (2.97), cetin immigr (2.96)

**Most Democrat phrases:** victim (-4.05), gun (-3.98), famili (-2.97), thought (-2.65), peopl (-2.48), dead (-2.36), heart (-2.32), seattl (-2.30), violenc (-2.23), larg (-2.11), mile (-2.10), shooter larg (-2.08), latest (-2.05), day (-2.04), tonight (-2.03), safe (-2.01), watch (-2.00), communiti (-1.98), shoot (-1.96), #break (-1.94)

Fort Lauderdale. **Most Republican phrases:** garag (4.54), shot fire (3.45), terrorist (3.14), fox (2.93), free zone (2.76), attack (2.71), gun free (2.70), fire (2.62), zone (2.61), free (2.43), muslim (2.36), shot (2.31), obama (2.09), park (2.08), terrorist attack (2.08), park garag (2.06), shooter (2.01), fox news (2.01), terror (1.93), report shot (1.81)

**Most Democrat phrases:** stop (-2.65), violenc (-2.52), custodi 9 (-2.25), gun (-2.14), mass (-2.12), fill (-2.10), 9 injur (-2.05), week (-2.00), multipl peopl (-1.91), 2017 (-1.86), airport suspect (-1.85), stay safe (-1.83), love (-1.78), heart (-1.70), safe (-1.70), report fire (-1.66), smh (-1.65), local (-1.64), msnbc (-1.64), thought (-1.61)

Fresno. **Most Republican phrases:** akbar (4.67), allahu (3.84), allahu akbar (3.76), yell (3.32), yell allahu (2.98), ap (2.61), suspect yell (2.38), al-lah (2.33), shout (2.25), terror attack (2.06), shout allahu (2.06), allah akbar (2.03), terrorist (1.98), terrorist attack (1.97), muslim (1.96), kill suspect (1.94), smm (1.83), islam (1.81), god (1.80), akbar hate (1.79)

**Most Democrat phrases:** chief (-3.97), victim (-3.89), polic (-3.74), famili (-3.65), fatal (-3.62), offic (-2.98), downtown (-2.92), dyer (-2.75), men (-2.71), polic chief (-2.70), gunman (-2.69), tuesday (-2.47), gun (-2.29), shoot (-2.22), suspect custodi (-2.17), kill california (-2.10), mass (-2.09), sad (-2.08), white men (-1.95), kill hate (-1.95)

San Francisco. **Most Republican phrases:** polic (4.85), multipl (4.41), report (4.32), pistol (4.18), assault pistol (4.12), 4 injur (3.41), shooter (3.33), assault (3.24), free (3.19), warehouse (3.05), injur (3.01), faci 4 (2.98), multipl victim (2.90), report faci (2.85), stolen (2.82), shot facili (2.76), hospit (2.68), law (2.60), gun law (2.59), compani (2.51)

**Most Democrat phrases:** today (-4.82), mass (-4.18), sf (-4.07), die (-3.47), mention (-3.29), cov-erag (-2.97), yesterday (-2.95), america (-2.75), #upsshoot (-2.49), #sf (-2.40), gun violenc (-2.40), mass shoot (-2.40), barclay (-2.31), virginia (-2.30), #up (-2.20), violenc (-2.17), morn (-2.17), shoot (-2.17), gob (-2.07), peopl kill (-2.04)

Vegas. **Most Republican phrases:** shooter (41.28), fbi (36.07), video (31.96), isi (31.08), democrat (26.78), paddock (26.34), liber (26.07),
Thornton. Most Republican phrases: multipl (6.04), suspect (5.24), parti (5.06), multipl parti (5.03), break (4.90), news (4.80), injur (4.36), dead (4.23), report (3.98), update (3.87), polic (3.84), suspect custodi (3.75), activ (3.71), chicago (3.57), detail (3.55), 2 (3.52), report multipl (3.28), #break (3.27), video (3.09), activ shooter (3.08)

Most Democrat phrases: white (-7.61), guy (-5.37), gun (-5.24), white guy (-3.92), bibl (-3.75), terrorist (-3.75), week (-3.61), white man (-3.48), good guy (-3.46), stack bibl (-3.33), live stack (-3.29), stack (-3.28), furnitur (-3.27), terror (-3.16), guy gun (-3.06), penalti (-3.04), bibl furnitur (-3.02), death penalti (-3.01), talk (-3.01), vega (-2.97)

Sutherland Springs. Most Republican phrases: shotgun (10.99), revolv (7.57), illeg (7.10), sheriff (6.20), liber (6.20), ban (6.14), metal detector (6.12), detector (6.00), secur (5.93), illeg (5.87), metal detector (5.87), metal (5.87), truck (5.79), rack (5.53), high school (5.45), bomb (5.43), high (5.33), leftist (5.30), gun rack (5.27), law stop (5.15), update (5.15), rifl (5.12)

Most Democrat phrases: black (-11.21), white (-9.41), trump (-8.06), terrorist (-7.67), tweet (-6.16), hero (-6.13), american (-6.05), shaw (-5.77), black man (-5.73), jame shaw (-5.48), jr (-5.44), jame (-5.42), jame shaw jr (-5.41), mention (-5.30), domest (-5.05), bond (-5.02), unarm (-4.74), domest terrorist (-4.69), black peopl (-4.63), man (-4.63)

Santa Fe. Most Republican phrases: shotgun (10.99), revolv (7.57), illeg (7.10), sheriff (6.20), liber (6.20), ban (6.14), metal detector (6.12), detector (6.00), secur (5.93), metal (5.87), truck (5.79), rack (5.53), high school (5.45), bomb (5.43), high (5.33), leftist (5.30), gun rack (5.27), law stop (5.15), update (5.15), rifl (5.12)

Most Democrat phrases: #guncontrolnow (-12.17), #gunreformnow (-10.36), #neveragain (-9.79), #enoughisennough (-8.90), nra (-8.57), children (-7.85), vote (-7.69), gun violenc (-7.59), america (-6.75), thought prayer (-6.72), violenc (-6.66), congress (-6.63), thought (-6.50), #enough (-6.21), fuck (-6.11), white (-6.01), #parkland (-5.74), gun (-5.64), republican (-5.44), gop (-5.40)

Annapolis. Most Republican phrases: blame (8.23), blame trump (7.82), liber (5.86), maryland (5.76), reuter (5.70), reuter editor (5.62), apolog (5.54), #break (5.23), editor apolog (5.22), disciplin (4.90), apolog disciplin (4.84), hat (4.84), dis-
ciplin blame (4.74), maga hat (4.60), claim shooter (4.48), fals (4.39), polic (4.38), #fakenew (4.36), democrat (4.32), wore maga (4.32)

Most Democrat phrases: journalist (-7.03), enemi (-6.10), press (-5.62), lower flag (-5.52), request lower (-5.36), request (-5.02), gazett victim (-4.55), declin request (-4.46), memori capitol (-4.42), lower (-4.40), enemi peopl (-4.39), press enemi (-4.37), white (-4.35), flag memori (-4.34), declin (-4.30), flag (-4.13), memori (-4.12), trump declin (-4.09), mass (-4.08), #guncontrolnow (-4.05)

Pittsburgh. Most Republican phrases: moment silenc (16.24), silenc (15.99), interrupt (14.74), interrupt moment (14.68), scream (13.09), moment (12.37), march (12.13), blackburn (11.85), leftist (11.73), silenc life (11.59), blame trump (11.28), silenc synagogu (11.20), protest (10.78), protest interrupt (10.56), rabbi blame (9.64), leftist interrupt (9.55), scream leftist (9.55), rage scream (9.52), horribl rage (9.43), scream insult (9.40)

Most Democrat phrases: violenc (-6.73), heart (-6.29), supremacist (-6.05), white supremacist (-6.05), muslim (-5.96), white (-5.94), mr (-5.88), trump vile (-5.86), result (-5.84), of (-5.80), result of (-5.80), inevit (-5.80), inevit result (-5.78), group (-5.77), synagogu inevit (-5.77), vile nation (-5.73), muslim group (-5.73), of trump (-5.71), roger (-5.69), massacr heart (-5.63)

Thousand Oaks. Most Republican phrases: california (16.51), zone (12.80), free (12.40), gun free (12.22), free zone (11.80), bar (9.98), california bar (9.61), strictest (9.01), strictest gun (8.52), men (7.63), #foxnew (7.33), killer ian (7.12), california strictest (7.11), fear resid (7.11), report killer (7.08), prayer massacr (7.04), communist (7.04), long mock (7.00), mock hope (6.80), ian long (6.77)

Most Democrat phrases: inact (-17.50), pattern (-17.46), pattern inact (-17.45), shoot pattern (-17.43), shoot (-16.92), januari (-13.09), inact januari (-12.92), #guncontrolnow (-8.09), mass shoot (-7.89), day (-7.34), fuck (-6.94), nra (-6.94), violenc (-6.71), mother (-6.64), mass (-6.59), thought (-6.44), high (-6.37), januar (-6.22), inact januar (-6.15), #potus (-6.05)

F.2 Most Partisan Phrases Per Topic

<table>
<thead>
<tr>
<th>Topic colors</th>
<th>shooter’s identity &amp; ideology</th>
<th>news</th>
<th>victims &amp; location</th>
<th>laws &amp; policy</th>
<th>investigation</th>
<th>solidarity</th>
<th>remembrance</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic colors</td>
<td>shooter’s identity &amp; ideology</td>
<td>news</td>
<td>victims &amp; location</td>
<td>laws &amp; policy</td>
<td>investigation</td>
<td>solidarity</td>
<td>remembrance</td>
<td>other</td>
</tr>
<tr>
<td>F.2 Most Partisan Phrases Per Topic</td>
<td>shooter’s identity &amp; ideology</td>
<td>news</td>
<td>victims &amp; location</td>
<td>laws &amp; policy</td>
<td>investigation</td>
<td>solidarity</td>
<td>remembrance</td>
<td>other</td>
</tr>
</tbody>
</table>
Table 7: Most partisan phrases per topic for Chattanooga. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Republican Phrases</th>
<th>Democrat Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalamazoo</td>
<td>#teardownthismosqu (0.42), #islamocar (0.42), #stoprush (0.42), #obama (0.42), #wakeupamerica (0.42)</td>
<td>#chattanoogaattack (0.33), #wakeupamerica (0.33), clinton (0.33), safe (0.33), muhammad (0.33)</td>
</tr>
<tr>
<td>Roseburg</td>
<td>decid (0.28), refus (0.28), command (0.28), troop (0.28), kill tennesse (0.28)</td>
<td>#chattanoogaattack (0.33), #wakeupamerica (0.33), clinton (0.33), safe (0.33), muhammad (0.33)</td>
</tr>
</tbody>
</table>

Table 8: Most partisan phrases per topic for Roseburg. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Republican Phrases</th>
<th>Democrat Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dallas</td>
<td>netanyahu releas (0.80), lgbt hop (0.80), obama contact (0.80), hop make (0.80), trump #nrabangunfreezon (0.80)</td>
<td>reaction tragic (0.92), wrong reaction (0.92), search massacr (0.92), kohn (0.92), sick muslim (0.92)</td>
</tr>
</tbody>
</table>

Table 9: Most partisan phrases per topic for Colorado Springs. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Republican Phrases</th>
<th>Democrat Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Bernardino</td>
<td>polic (0.66), motos (0.65), violenc (0.64), obama (0.63), peopl (0.63)</td>
<td>coverag (0.44), news (0.44), open (0.44), cover (0.44), white male (0.37)</td>
</tr>
</tbody>
</table>

Table 10: Most partisan phrases per topic for Kalamazoo. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Republican Phrases</th>
<th>Democrat Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orlando</td>
<td>#floridashoot #puls (-0.78), derail (-0.78), maryland (-0.78), #enough (-0.78), bbcworld (-0.78)</td>
<td>#stoprush #uniteblu (-0.92), #connecttheleft (-0.92), #p2 #connecttheleft (-0.92), muslim allianc (-0.92), lgbtq hate (-0.92)</td>
</tr>
</tbody>
</table>

Table 11: Most partisan phrases per topic for Orlando. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Republican Phrases</th>
<th>Democrat Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dallas</td>
<td>#homelandsecur (-0.39), #pjnet (-0.39), investig islam (-0.39), #trump2016 (-0.39), park spot (0.61)</td>
<td>#terror (-0.60), #plannedparenthoodshoot (-0.58), month (-0.57)</td>
</tr>
</tbody>
</table>

Table 12: Most partisan phrases per topic for Dallas. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Republican Phrases</th>
<th>Democrat Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalamazoo</td>
<td>#pjnet (0.64), mt prayer (0.64), #bluelivesmatt #pjnet (0.64), sam (0.64), #maga (0.64)</td>
<td>comey (0.67), recommend (0.67), #tcot (0.67), #obama (0.67), sex (0.67)</td>
</tr>
</tbody>
</table>

Table 13: Most partisan phrases per topic for Kalamazoo. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Republican Phrases</th>
<th>Democrat Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltimore</td>
<td>#pjnet (-0.74), profit (0.84), lib (0.81), chicago (0.76), bodi part (0.74)</td>
<td>#tcot (0.67), #pjnet (0.67), pro life (0.63), pro (0.55), #prolif (0.54)</td>
</tr>
</tbody>
</table>

Table 14: Most partisan phrases per topic for Baltimore. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Republican Phrases</th>
<th>Democrat Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>#tcot (-0.74), focus (-0.39), terrorist attack (-0.39), forc (-0.39), attack women (-0.39)</td>
<td>#terror (-0.60), #plannedparenthoodshoot (-0.58), month (-0.57)</td>
<td></td>
</tr>
</tbody>
</table>

Table 15: Most partisan phrases per topic for #tcot. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Republican Phrases</th>
<th>Democrat Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>#terror (-0.74), focus (-0.39), terrorist attack (-0.39), forc (-0.39), attack women (-0.39)</td>
<td>#terror (-0.60), #plannedparenthoodshoot (-0.58), month (-0.57)</td>
<td></td>
</tr>
</tbody>
</table>

Table 16: Most partisan phrases per topic for #terror. Brackets show the z-scores of the log odds of each phrase.
Table 14: Most partisan phrases per topic for Baton Rouge. Brackets show the z-scores of the log odds of each phrase.

Table 15: Most partisan phrases per topic for Burlington. Brackets show the z-scores of the log odds of each phrase.

Table 16: Most partisan phrases per topic for Fort Lauderdale. Brackets show the z-scores of the log odds of each phrase.

Table 17: Most partisan phrases per topic for Fresno. Brackets show the z-scores of the log odds of each phrase.

Table 18: Most partisan phrases per topic for San Francisco. Brackets show the z-scores of the log odds of each phrase.

Table 19: Most partisan phrases per topic for Vegas. Brackets show the z-scores of the log odds of each phrase.
Table 20: Most partisan phrases per for Thornton. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
</tr>
</tbody>
</table>

Table 21: Most partisan phrases per for Sutherland Springs. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
</tr>
</tbody>
</table>

Table 22: Most partisan phrases per for Parkland. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
</tr>
</tbody>
</table>

Table 23: Most partisan phrases per for Nashvile. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
</tr>
</tbody>
</table>

Table 24: Most partisan phrases per for San Fe. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
</tr>
</tbody>
</table>

Table 25: Most partisan phrases per for Annopolis. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
</tr>
</tbody>
</table>

Table 26: Most partisan phrases per for Pittsburgh. Brackets show the z-scores of the log odds of each phrase.

<table>
<thead>
<tr>
<th>Phrase</th>
<th>Democrat</th>
<th>Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
<td>school (0.54), wait (0.54), church (0.38), houos attack (0.34), 4 peopl (0.34)</td>
</tr>
<tr>
<td>Republican</td>
<td>Democrat</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Violent women (-0.55), republican (-0.49), women (-0.49), histor (-0.47), link (-0.47)</td>
<td>Mass 12 (-0.61), 12 victim (-0.61), gun violence (-0.47), hour (-0.42), violence (-0.42)</td>
<td></td>
</tr>
<tr>
<td>Strait (-0.61), mass die (-0.45), live gun (-0.45), gop (-0.45)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 27: Most partisan phrases per topic for Thousand Oaks. Brackets show the z-scores of the log odds of each phrase.
G  Personal Pronouns

Pronoun usage has been found to express different degrees of self-reflection, awareness, inclusivity, perspective-taking, among other psychological phenomena, both on Twitter (Qiu et al., 2012) and in various other domains (Hirsh and Peterson, 2009; Yarkoni, 2010; Pennebaker, 2011). We rely on these findings to treat pronouns as proxies to measure polarization in terms of users’ personalization of the experience (1st person sg.), inclusion (1st person pl.), deflection (3rd person).

G.1 Methods

To quantify the partisanship of personal pronouns, we take the five personal pronoun categories in LIWC (Pennebaker et al., 2001) (I, You, We, SheHe, They) and then calculate their partisan log-odds ratio.

G.2 Results

The log odds of pronouns suggest that first and second person pronouns are significantly more likely to be used by Democrats across the events, while They is somewhat more likely to be used by Democrats and SheHe is used similarly by the two parties: I (mean: −.26, p < .001), We (mean: −.26, p < .001), You (mean: −.13, p < .05), They (mean: −.06, p < .1), SheHe (mean: .05, p ≈ 0.36) (see Figure 15).

The use of two pronouns is significantly different based on the shooter’s race: SheHe and You are both more likely to be used by Democrats when the shooter is white and by Republicans if the shooter is a person of color (p < 0.01 for SheHe and p < 0.05 for You from two-tailed t test). The pattern pertaining to SheHe might be partly explained by differential mentions of the president (see Appendix F with the most partisan words), since it so happens that most of the events where the shooter was a person of color occurred under Obama’s presidency while the ones where the shooter was white predominantly occurred under Trump’s presidency.

To better understand this result, we are again interested to see if there is a link between pronoun usage and topic preferences — we use the same procedure to measure the representation of pronouns in topics as in the case of modals. Our findings (see Figure 15) show that first I predominantly occurs in solidarity and other, which, coupled with previous findings about these topics being preferred by Democrats and about affect, suggest that Democrats in our dataset are more likely to personalize their tweets and write about their own mental state and feelings towards the event and the victims. Similarly We is overrepresented in laws & policy, also a topic that is preferred by Democrats, which, building on our results about modals (Section 7.2), provide evidence that Democrats are more likely to call for collective action.

SheHe, on the other hand, are most frequent in investigation, shooter’s identity & ideology and victims & location — topics that are more likely to be discussed by Republicans. This result, supported by our finding about affect, suggests that Republicans in this domain are more likely to author tweets that focus on a third person (e.g. the shooter).
H Modal Collocations

We study the subjects and complements of modals and their partisanship to get a better view of how these modals are used. We calculate the z-scored partisan log odds ratio of the collocations (subject + modal + complement constructions) at the event- and modal-level. We keep all collocations that are more likely to be used by Democrats or Republicans by at least .5 SD of the event- and modal-level log odds ratios (in other words, whose z-score has an absolute value that is at least .5). In the list below, we show those collocations that are partisan in the same direction (Democrat or Republican) in at least three events.

Note that before calculating the log odds ratio, we replace contracted forms with their non-contracted form (e.g. “shouldn’t” with “should not”, “should’ve” with “should have”). The number following the collocations is the number of events for which a particular term is partisan towards a given party. Often when a collocation seems ungrammatical, it is because it is comes from a question (e.g. “long must we” → “how long must we”).

Note that the patterns in the collocations accord with the findings discussed in the paper. Democrats are more likely to use modals to call for collective and proactive action (e.g. “we must do”, “we have to do”, “something needs to be done”, “we need to act”) and also to express emotion (e.g. “[why do] people have to die”, “[why do] people need to die”) than Republicans. Republicans are more likely to use modals epistemically (e.g. “it must have been”) and in other, idiomatic, senses that do not imply necessity (e.g. “it has to do with”, “I have to say”, “I must say”). Republicans are less likely to use modals in contexts with first person plural subjects (“we”) than Democrats, but are more likely to use “we” in contexts where the modal’s prejacent implies “stopping” something rather than “doing” something (e.g. “we must ban”, “we must protect”, “we should ban”, “we need to protect”).

MUST

**Democrat**
we must do (14), this must stop (8), we must end (7), something must be done (7), we must act (7), long must we (7), congress must act (7), we must stand (6), we must stop (6), he must be white (6), we must address (6), violence must end (6), violence must stop (6), lives must be lost (6), killing must stop (6), we must make (6), times must this (5), it must stop (5), we must all (5), people must die (5), many must die (5), we must keep (5), and must do (5), we must pass (5), we must continue (5), times must we (5), this must end (5), we must take (5), we must honor (5), shooter must be white (5), shooter must have been (4), insanity must stop (4), we must treat (4), we must remember (4), shootings must stop (4), is a must (4), shootings must end (4), change must happen (4), congress must take (4), this must change (4), we must get (4), action must be taken (4), control must happen (4), we must push (4), we must find (4), we must have gun (4), we must change (4), we must work (4), we must combat (4), we must deal (4), lives must be taken (4), we must try (4), congress must pass (4), madness must end (4), we must hold (4), gunviolence must end (4), we must enact (4), rifles must be banned (4), they must know (3), i must say (3), we must ask (3), we must always (3), guncontrol must happen (3), we must reclaim (3), madness must stop (3), it must take (3), there must have been (3), action must follow (3), we must condemn (3), i must ask (3), we must stopgunviolence (3), innocents must die (3), nra must be so (3), why must we (3), we must come (3), we must fight (3), things must change (3), tragedies must end (3), violence must be stopped (3), we must confront (3), violence must be addressed (3), we must stay (3), you must be so (3), laws must change (3), we must (3), shootings must we (3), we must join (3), you must really (3), weapons must be banned (3), we must have change (3), love must prevail (3), shooting must have been (3), more must we (3), we must not become (3), we must create (3), we must allow (3), children must die (3), you must be proud (3), that must mean (3), we must fix (3), shooting must stop (3), we must look (3), must feel (3), we must as (3), more must die (3), there must be something (3), you must do (3), and must be stopped (3), we must talk (3), we must not forget (3), we must vote (3)

**Republican**
it must have been (6), they must be stopped (5), must watch (5), i must say (5), we must ban (5), they must know (4), you must know (4), we must protect (4), we must remain (4), i must have missed (4), he must resign (4), why must you (4), you must protect (4), we must be vigilant (4), attacks must stop (4), you must be happy (4), we
must reform (4), it must be stopped (4), this must be one (4), rhetoric must stop (4), we must understand (4), we must stand (4), we must return (4)

**SHOULD**

*Democrat*

this should not be normal (8), we should all (7), who should not have guns (6), who should not have them (6), this should not be happening (6), this should not happen (6), who should not have had (6), this should not have happened (5), we should just (5), you should be ashamed (5), congress should be forced (5), that should not be happening (5), people should be able (5), we should never (5), who should not have a gun (5), civilians should not have access (5), we should know (5), you should see (5), never should have had (5), everyone should be able (4), we should ban (4), we should start (4), you should (4), we should probably (4), should we (4), you should not be able (4), we should bring (4), one should die (4), there should be some (4), we should not do (4), we should wait (4), this should never (4), we should not need (4), it should not be this (4), one should be able (4), we should not be afraid (4), everyone should have a gun (4), civilians should have access (4), they should have had (4), we should now (4), people should not have access (4), one should be afraid (4), he should not have had (4), this should not be our (4), we should try (4), one should live (4), we should not allow (4), hook should have been (4), people should have access (4), everyone should have guns (4), everyone should own (4), it should not take (4), this should have never (4), guns should be allowed (4), there should be a ban (3), media should be ashamed (3), we should change (3), people should be more (3), you should ask (3), we should politicize (3), you should never (3), i should have known (3), shootings should not be the norm (3), we should do (3), we should be upset (3), we should be afraid (3), there should have been (3), there should never (3), it should not be that (3), we should be focused (3), one should have access (3), we should not care (3), you should also (3), you should tweet (3), something should be done (3), should not they (3), it should have been (3), we should not be surprised (3), they should know (3), there should be more (3), why should it (3), you should call (3), congress should do (3), he should go (3), why should anyone (3), people should not be shot (3), guns should be outlawed (3), one should feel (3), weapons should be outlawed (3), rifles should be banned (3), guns should be illegal (3), people should not be killed (3), never should have happened (3), one should have to fear (3), people should feel (3), we should leave (3), what should be a safe (3), guns should be legal (3), people should be allowed (3), anyone should be able (3), media should stop (3), we should remember (3), but should not we (3), that should be illegal (3), they should feel (3), you should all (3), one should lose (3), you should too (3), civilian should have access (3), we should focus (3), we should (3), person should not be able (3), you should donate (3), should i (3), one should be getting (3), weapons should not be sold (3), everyone should have the right (3), guns should never (3), civilians should not have assault (3), people should not have to worry (3), you should not be allowed (3), shooting should not happen (3), we should stand (3), you should be embarrassed (3), we should build (3), you should leave (3), you should be more (3), we should not be able (3), what should you (3), you should not own (3), civilians should be able (3), guns should not be allowed (3), one should have to worry (3), we should have gun (3), he should never (3), everyone should read (3), we should not rush (3), it should not be allowed (3), we should say (3), all should (3), this should not be a political (3), he should be able (3), we should be having (3), people should not be able (3), i should not feel (3), they should put (3), man should not have been (3), civilian should ever (3), it should not be lost (3), civilians should not own (3), weapons should be legal (3), i should never (3), people should have died (3), fbi should investigate (3), we should attack (3), when should we (3), parent should ever (3), we should name (3), shooting should of (3), u should get (3), civilians should own (3), we should arm (3), it should be for (3), teachers should have guns (3), nra should have to pay (3), you should not have the right (3), shooting should have never (3), they should have been (3), we should really (3), it should not matter (3), who should get (3), that should not have guns (3), can should do (3), we should discuss (3), it should (3), one should be shot (3), should people (3), we should treat (3), we should not talk (3), they should (3), rifles should not be available (3), someone should tell (3), you should be shot (3), that should have been (3), we
should not have laws (3), people should not be allowed (3), it should not even (3), or should we (3), civilian should own (3), they should call (3), you should be tweeting (3), one should have to go (3), that should be safe (3), that should never (3), we should definitely (3), child should ever (3), somebody should tell (3), nothing should be done (3), we should not let (3)

**Republican**

we should ban (8), we should make (7), and should be prosecuted (6), they should just (6), police should have guns (5), i should have been (5), you should know (5), they should of (5), you should go (5), he should have been (5), we should keep (5), he should be arrested (5), they should not be allowed (5), government should have guns (5), we should see (5), should we (5), they should give (4), they should be allowed (4), everyone should know (4), we should at (4), they should make (4), shooting should not we (4), obama should resign (4), someone should inform (4), we should outlaw (4), you should do (4), you should ask (4), he should be fired (4), us should take (4), you should be fired (4), what should be done (4), fbi should have been (4), they should (4), why should he (4), and should have been (4), we should be asking (4), we should have guns (4), parents should be held (4), he should have stopped (4), we should add (4), he should take (4), and should be fired (4), they should use (4), democrats should not be allowed (4), they should hang (4), you should be mad (4), we should work (4), what should they (4), they should take (4), we should be able (4), guns should be taken (4), you should stop (4), you should see (4), there should never (4), shooter should get (4), who should we (4), shooter should not have had (4)

**NEED TO**

**Democrat**

we need to take (8), more need to die (8), we need to act (8), we need to talk (8), we need to do (8), we need to stand (7), we need to stop (7), people need to die (7), shootings need to happen (6), laws need to change (6), we need to change (6), we need to fix (6), we need to vote (6), we need to make (6), we need to call (5), they need to carry (5), we need to remember (5), don’t need to see (5), all need to take (5), we need to hear (4), you need to rethink (4), we need to be better (4), we need to start (4), guns need to go (4), we need to figure (4), people need to stop (4), you need to change (4), we need to keep (4), we need to solve (4), laws need to be changed (4), we need to work (4), we need to end (4), we need to be able (4), we need to address (4), we need to help (4), we need to demand (4), weapons need to be banned (4), we need to have more (4), i need to know (4), lives need to be lost (3), acts need to stop (3), you need to know (3), we need to focus (3), don’t need to keep (3), many need to die (3), we need to have before (3), we need to look (3), we need to say (3), we need to see (3), we need to pass (3), massacres need to happen (3), you need to recognize (3), also need to make (3), this need to happen (3), lives need to be taken (3), we need to #endgunviolence (3), all need to stand (3), we need to acknowledge (3), we need to mourn (3), shootings need to stop (3), don’t need to know (3), violence need to end (3), we need to have better (3), we need to live (3), laws need to happen (3), violence need to stop (3), they need to know (3), people need to be shot (3), we need to wait (3), you need to talk (3), we need to listen (3), we need to honor (3), people need to stand (3), i need to stop (3), we need to be talking (3), we need to also (3), we need to speak (3), who need to get (3), we need to go (3), people need to hear (3), we need to wake (3), guns need to be restricted (3), really need to get (3), we need to have a real (3), felt the need to take (3), we need to understand (3), all need to see (3), we need to be outraged (3), people need to start (3), i need to add (3), they need to pay (3), don’t need to do (3), we need to ban (3), leaders need to do (3), you need to tell (3), we need to limit (3), we need to prevent (3), we need to really (3)

**Republican**

you need to know (8), killings need to stop (6), we need to protect (6), we need to get (6), we need to focus (6), we need to know (6), we need to be more (5), really need to look (5), people need to realize (5), people need to stop (5), people need to know (5), we need to pray (4), we need to wait (4), we need to allow (4), they need to kill (4), americans need to be armed (4), really need to stop (4), we need to return (4), we need to kill (4), you need to watch (4), seriously need to change (4), heads need to roll (4), you need to investigate (4), democrats need to stop (4), people need to get (4), you need to address (4), you need to go (4), we need to enforce (4), we need to find (4), guns need
to be banned (4), we need to look (4), you need to look (4), you need to ask (4), we need to punish (4)

NEEDS TO

Democrat
something needs to be done (8), this needs to stop (8), violence needs to stop (7), it needs to end (6), something needs to change (5), there needs to be stricter (5), this needs to end (5), violence needs to end (5), gop needs to stop (5), america needs to do (4), change needs to happen (4), that needs to be addressed (4), shit needs to stop (4), congress needs to stop (4), it needs to stop (4), america needs to end (3), else needs to happen (3), seriously needs to be done (3), control needs to happen (3), one needs to own (3), what needs to happen (3), reform needs to happen (3), control needs to be a thing (3), there needs to be more (3), there needs to be gun (3), hatred needs to stop (3), it needs to be said (3), world needs to change (3), work needs to be done (3), law needs to change (3), seriously needs to stop (3), control needs to be addressed (3), cnn needs to stop (3), congress needs to do (3), someone needs to tell (3), story needs to be told (3), he needs to take (3), it needs to happen (3), he needs to get (3), congress needs to act (3), this needs to be stopped (3), trump needs to stop (3)

Republican
she needs to go (4), someone needs to ask (4)

HAVE TO

Democrat
people have to die (10), we have to do (7), more have to die (6), many have to die (6), don’t have to live (5), should not have to fear (4), we have to say (4), don’t have to do (4), this have to happen (4), you have to say (4), things have to change (4), should have to fear (4), we have to change (4), we have to talk (4), we have to act (4), we have to keep (4), i have to worry (4), we have to read (3), we have to go (3), we have to have before (3), we have to endure (3), we have to lose (3), lives have to be lost (3), we have to accept (3), we have to hear (3), and not have to worry (3), you have to offer (3), we have to end (3), i have to see (3), shootings have to stop (3), we have to look (3), should have to go (3), we have to pay (3), children have to die (3)

 Republican
i have to wonder (7), that have to do (6), shooting have to do (6), you have to say (6), we have to ask (5), i have to say (5), will have to live (5), shootings have to stop (5), i have to go (5), you have to kill (5), i have to agree (5), you have to change (5), you have to be to shoot (5), does have to do (5), nra have to do (5), trump have to do (5), these have to happen (4), we have to listen (4), you have to go (4), you have to put (4), will have to wait (4), going have to work (4), you have to talk (4), we have to pray (4), we have to see (4), you have to know (4), you have to stop (4), this have to do (4), people have to stop (4), you have to give (4), going have to get (4), who have to deal (4), you have to ask (4), we have to remember (4), you have to be really (4)

HAS TO

Democrat
this has to stop (11), something has to change (6), it has to stop (5), violence has to end (5), shooting has to do (4), this has to end (4), something has to be done (4), more has to happen (4), killing has to stop (4), it has to end (4), what the has to say (4), shit has to stop (3), what has to happen (3), this has to change (3), madness has to end (3), simply has to stop (3), humanity has to offer (3), what has to be done (3), stuff has to stop (3), else has to happen (3), there has to be footage (3), killings has to stop (3)

Republican
it has to do (8), that has to do (6), obama has to say (5), he has to say (5), this has to end (4), there has to be some (4), what has to say (4), one has to ask (4), shit has to end (4)

 Republican

I Results: Additional Plots

Figure 16: The log odds ratio of necessity modals.

Figure 17: Topic polarization of Orlando over time, as measured by the leave-out estimate of phrase partisanship. The bar charts show the proportion of each topic in the data at a given time.