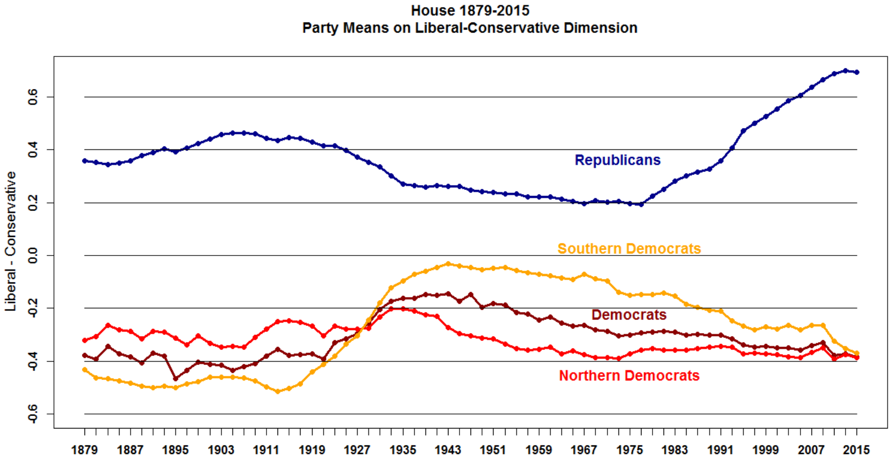
Building a Robot Judge: Data Science for the Law

14. Measuring Polarization in Text

Elliott Ash

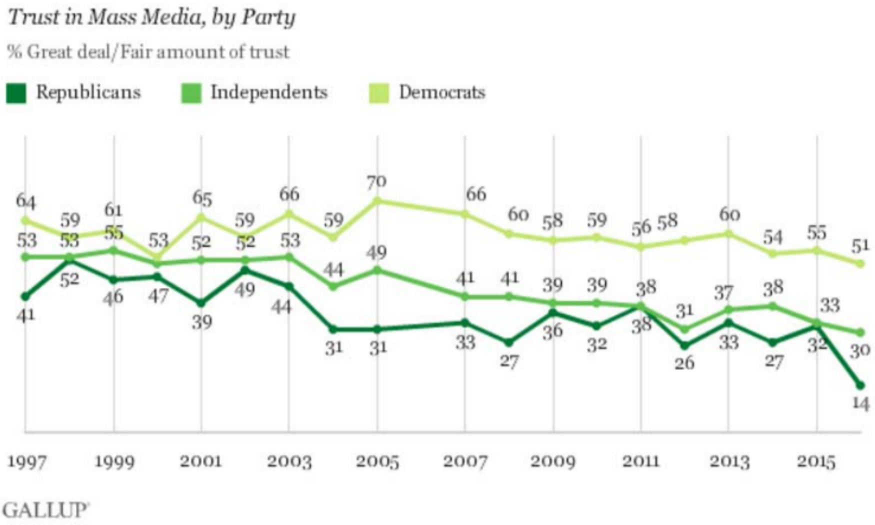
Polarization $\leftrightarrow$ Increasing Group Differences

- An important issue in public discourse across the world is **political polarization**.
  - That is, intensifying disagreement over the goals and means of government and society.
- In the U.S., Bernie Sanders on the Left an Donald Trump on the Right.
- In Europe, Five Star on the Left and Brexit on the Right.
- In Switzerland, the rise of the Swiss People's Party.



Source: Hans Noel.

Trust in Media



Gentzkow and Shapiro (2010)

- Corpora:
  - news text from large sample of US daily newspapers.
  - congressional text is 2005 Congressional Record.
- Pre-process text, stripping away prepositions, conjunctions, pronouns, and common words
  - get bigrams and trigrams
- Identify polarizing phrases using χ^2 metric. For each phrase w, let D_w be frequency for Democrats, R_w be frequency for Republicans. Let D_w^- and R_w^- be frequencies of *other* phrases.
- Then:

$$\chi_w^2 = \frac{(R_w D_w^- - D_w R_w^-)^2}{(D_w + R_w)(D_w + D_w^-)(R_w + R_w^-)(D_w^- + R_w^-)}$$

  - this is the test statistic for equality between parties of phrase use if they were both drawn from multinomial distributions.

TABLE I

MOST PARTISAN PHRASES FROM THE 2005 *CONGRESSIONAL RECORD*[a]

Panel A: Phrases Used More Often by Democrats		
Two-Word Phrases		
private accounts	Rosa Parks	workers rights
trade agreement	President budget	poor people
American people	Republican party	Republican leader
tax breaks	change the rules	Arctic refuge
trade deficit	minimum wage	cut funding
oil companies	budget deficit	American workers
credit card	Republican senators	living in poverty
nuclear option	privatization plan	Senate Republicans
war in Iraq	wildlife refuge	fuel efficiency
middle class	card companies	national wildlife
Three-Word Phrases		
veterans health care	corporation for public broadcasting	cut health care
congressional black caucus	additional tax cuts	civil rights movement
VA health care	pay for tax cuts	cuts to child support
billion in tax cuts	tax cuts for people	drilling in the Arctic National
credit card companies	oil and gas companies	victims of gun violence
security trust fund	prescription drug bill	solvency of social security
social security trust	caliber sniper rifles	Voting Rights Act
privatize social security	increase in the minimum wage	war in Iraq and Afghanistan
American free trade	system of checks and balances	civil rights protections
central American free	middle class families	credit card debt

TABLE I—*Continued*

Panel B: Phrases Used More Often by Republicans		
Two-Word Phrases		
stem cell	personal accounts	retirement accounts
natural gas	Saddam Hussein	government spending
death tax	pass the bill	national forest
illegal aliens	private property	minority leader
class action	border security	urge support
war on terror	President announces	cell lines
embryonic stem	human life	cord blood
tax relief	Chief Justice	action lawsuits
illegal immigration	human embryos	economic growth
date the time	increase taxes	food program
Three-Word Phrases		
embryonic stem cell	Circuit Court of Appeals	Tongass national forest
hate crimes legislation	death tax repeal	pluripotent stem cells
adult stem cells	housing and urban affairs	Supreme Court of Texas
oil for food program	million jobs created	Justice Priscilla Owen
personal retirement accounts	national flood insurance	Justice Janice Rogers
energy and natural resources	oil for food scandal	American Bar Association
global war on terror	private property rights	growth and job creation
hate crimes law	temporary worker program	natural gas natural
change hearts and minds	class action reform	Grand Ole Opry
global war on terrorism	Chief Justice Rehnquist	reform social security

[a]The top 60 Democratic and Republican phrases, respectively, are shown ranked by χ^2_{pl}. The phrases are classified as two or three word after dropping common "stopwords" such as "for" and "the." See Section 3 for details and see Appendix B (online) for a more extensive phrase list.

Consumers drive media slant (GS 2010)

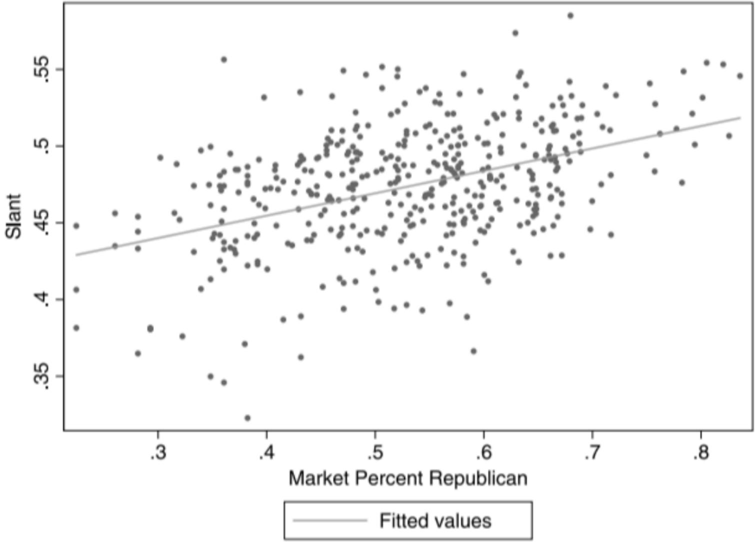


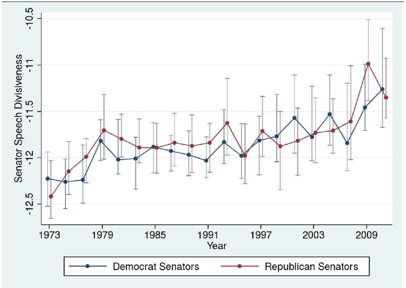
FIGURE 4.—Newspaper slant and consumer ideology. The newspaper slant index against Bush's share of the two-party vote in 2004 in the newspaper's market is shown.

Ash, Morelli, and Van Weelden (2017)

- Let f_{iwt} be the frequency of phrase w spoken by congressman i during session t, standardized within chamber-year to have mean zero and standard deviation one.
- Then the divisiveness of speech for congressman i at year t is defined as

$$Y_{it} = \log(\sum_w f_{iwt} \chi^2_{wt})$$

where the phrase-polarization measure χ^2_{wt} can vary by year.



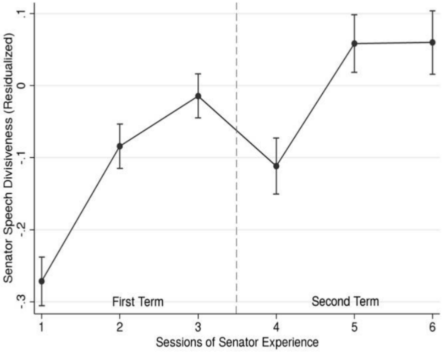


Figure 1. Senator speech divisiveness by election cohort. This figure plots average senator speech divisiveness over the course of the first two terms (six sessions, 12 years) of a senator's career. The values plotted are the mean residuals from a regression of senator speech divisiveness on a senator fixed effect, grouped by the first six sessions. This includes only senators who began their career in the first cohort (excluding senators appointed or elected to finish out an existing term). Error spikes indicate standard errors.

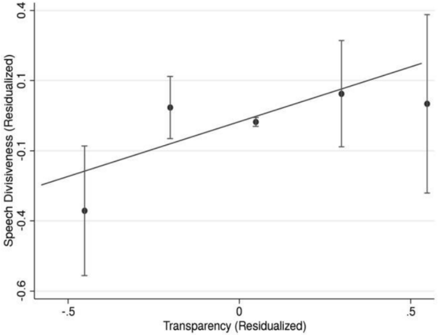
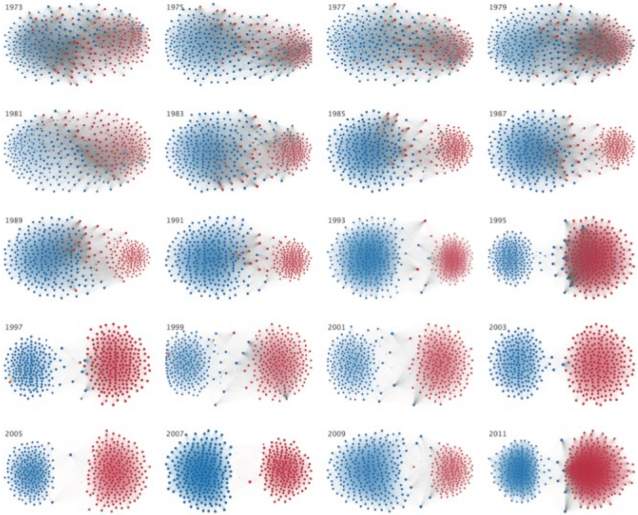


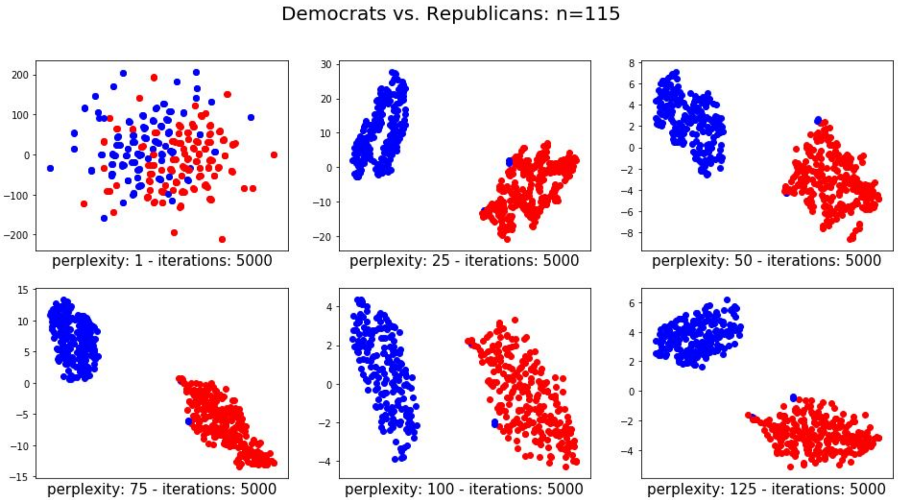
Figure 2. House member speech divisiveness by transparency level. This figure plots House member speech divisiveness against the transparency metric, after residualizing both on a party-year fixed effect, member fixed effect, and vote margin controls. Observations are grouped in bins of width .25. The trend line gives the linear fit. Error spikes indicate 95% confidence intervals.

Andris et al (2015)



"The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives." Plots of mutual agreement networks based on roll call votes, colored by party.

Congressman Embeddings



t-SNE plots of embedding vectors for each Congressman, trained to predict agreement on roll call votes. 115th Congress.

GST: Generative Model of Text

Gentzkow, Shapiro, and Taddy (*Econometrica* 2019)

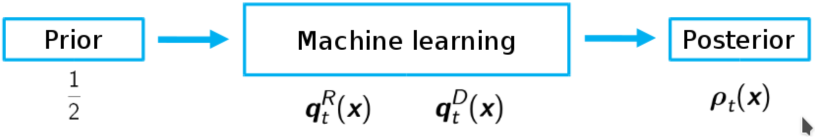
$\boldsymbol{c}_{it}^{p}$, vector of phrase frequences for speaker i at year t, by party $p \in D, R$, drawn from multinomial distribution

$$\boldsymbol{c}_{it}^{D} \sim \mathrm{MN}(\boldsymbol{q}_{t}^{D})$$
$$\boldsymbol{c}_{it}^{R} \sim \mathrm{MN}(\boldsymbol{q}_{t}^{R})$$

- $\boldsymbol{q}_{t}^{D}$ and $\boldsymbol{q}_{t}^{R}$ are party-specific vectors of probabilities

Bayesian Learning of Partisanship

Gentzkow, Shapiro, and Taddy (2019)



$\rho_{jt} = \frac{q_{jt}^R}{q_{jt}^R + q_{jt}^D}$, **posterior probability** that observer with neutral prior assigns to speaker being Republican if see phrase j in year t

- $\boldsymbol{\rho}_t$ is the vector of posteriors associated with each phrase

Define π_t = **partisanship** at time t:

$$\pi_t = \frac{1}{2}\boldsymbol{q}_t^R \cdot \boldsymbol{\rho}_t + \frac{1}{2}\boldsymbol{q}_t^D \cdot (1 - \boldsymbol{\rho}_t)$$

- Weighted average of posteriors – that is, is text informative of affiliation?

Language Choice Model

Gentzkow, Shapiro, and Taddy (2019)

- Let speaker i's "utility" from speaking phrase j at time t be

$$u_{ijt} = \alpha_{jt} + \boldsymbol{x}'_{it}\gamma_t + R_i\varphi_j$$

  - α_{jt}, baseline utility
  - γ_t, utility associated to speaker characteristics
  - R_i=Republican, so φ_j indexes party difference.

- If speaker chooses phrases to maximize utility u_{it} with respect to a choice-specific i.i.d. type 1 extreme value shock, then

$$q_{jt}(\boldsymbol{x}_{it}) = \frac{e^{u_{ijt}}}{\sum_l e^{u_{ilt}}}$$

Regularized cost function

Gentzkow, Shapiro, and Taddy (2019)

Learn $(\alpha_{jt}, \gamma_t, \varphi_j)$ to minimize

$$\sum_j \left\{ \sum_t \sum_i [\exp(\alpha_{jt} + \boldsymbol{x}'_{it}\gamma_t + R_i\varphi_j)m_{it} - (\alpha_{jt} + \boldsymbol{x}'_{it}\gamma_t + R_i\varphi_j)c_{ijt} + \lambda_j|\varphi_j|] \right\}$$

where m_{it}= number of phrases spoken; c_{ijt}= count for phrase j

- Approximate multinomial with Poisson model

$$c_{ijt} \sim \text{Pois}(\exp[\log(m_{it}) + u_{ijt}])$$

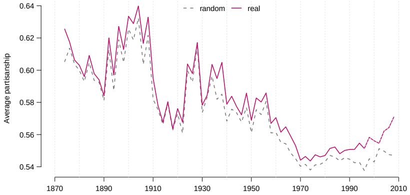
  allowing parallel computation across phrases.

- λ_j=phrase-specific lasso penalty, chosen to maximize information criterion.

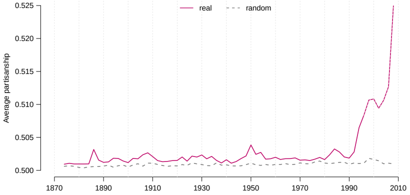
Regularization and Permutation

Gentzkow, Shapiro, and Taddy (2019)

Maximum Likelihood Estimator	Preferred (Penalized) Estimator



Usual method: Plug-in MLE w/ Congress speech



Regularized method w/ permutation inference

Figure 3: Informativeness of Speech by Speech Length and Session

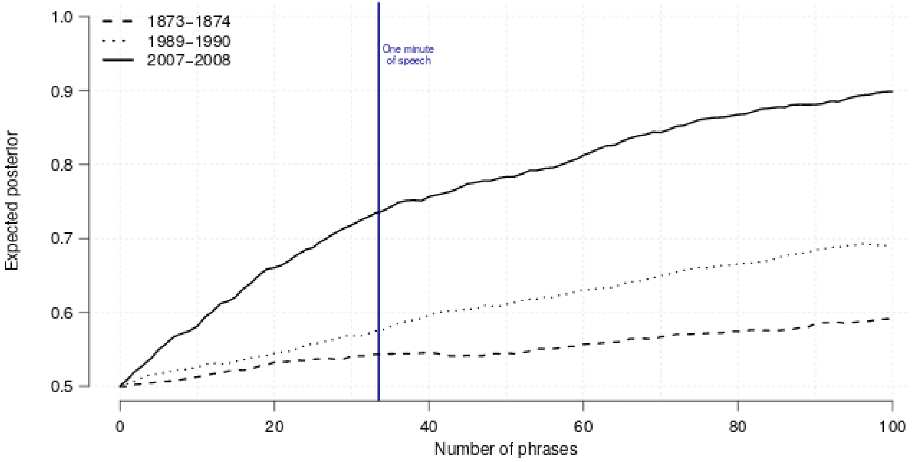
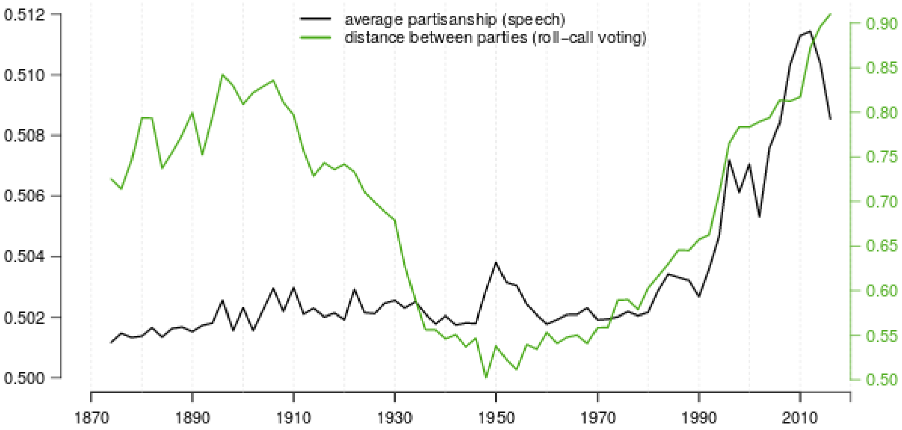


Figure 8: Partisanship vs. Roll-Call Voting

Panel A: Over Time

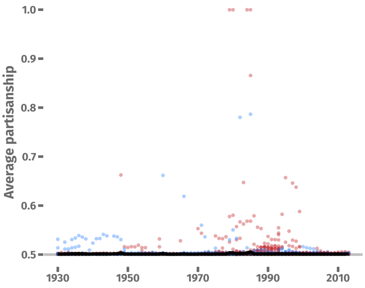


What about judges?

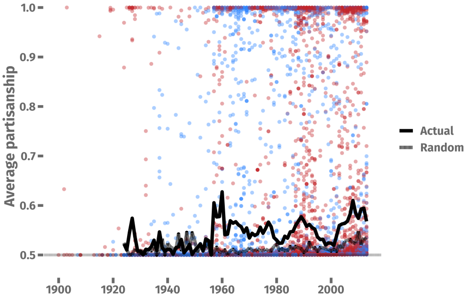
Ash, Chen, and Lu (2019)

- Apply the GST method to U.S. Circuit Courts, 1930-2013
- Look at polarization of language, as well as which previous cases are cited.

Polarization in Federal Judiciary



Phrases



Citations

Judicial prose (0.5) << Congress prose (0.515) << Precedent (0.6)

Analyzing polarization in social media: Method and application to tweets on 21 mass shootings

Demszky, Garg, Voigt, Zou, Gentzkow, Shapiro, and Jurafsky

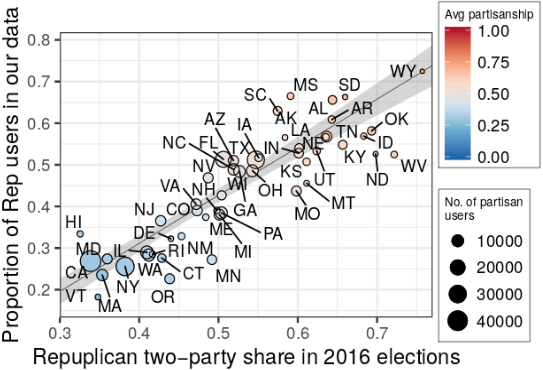
- Research Object:
  - use NLP to understand four “dimensions” of social media polarization: topic choice, framing, affect, modality.
- Context:
  - tweets in response to mass shooting events.
- Research question:
  - does political partisanship manifest in polarized responses to violent/polarizing events.

Dataset

- 21 mass shooting events, 2015-2018, from Gun Violence Archive
- tweets about those events, identified by:
  - location keywords (e.g. chattanooga, roseburg, san bernardino, fresno, etc.)
  - event keywords (lemmas): shoot, gun, kill, attack, massacre, victim
  - filter out retweets and tweets from deactivated accounts
  - $N = 10,000$ (out of 4.4 million tweets from the firehose archive).

Identifying party affiliation of Twitter users

- Party affiliation identified off of whether you follow more Democrats or Republicans, from a list of Twitter accounts associated with legislators, presidential candidates, and party organizations (Volkova et al 2014).
  - at least 51% of tweets for each event can be assigned partisanship this way.
- For geolocated users this matches up pretty well with party vote shares by state ($R^2 = .82$):



Pre-processing for partisanship

- Stemming and stopword removal.
- Event-specific vocabulary:
    - unigrams and bigrams
    - occur in event's tweets at least 50 times
    - must be used by at least two tweeters.

Partisanship

- Leave-one-out estimator from Gentzkow et al (2019), applied to each shooting event:

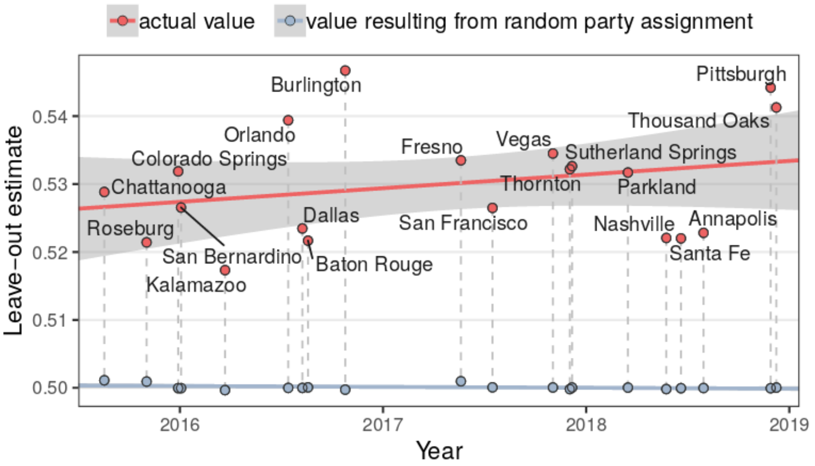
$$\pi = \frac{1}{2}(\frac{1}{|D|}\sum_{i \in D} \hat{\boldsymbol{q}}_i \cdot \hat{\boldsymbol{\rho}}_{-i} + \frac{1}{|R|}\sum_{i \in R} \hat{\boldsymbol{q}}_i \cdot (1 - \hat{\boldsymbol{\rho}}_{-i}))$$

  - $\hat{\boldsymbol{q}}_i$ = token frequencies for user i, drawn from set of democrats D and set of republicans R
  - $\hat{\boldsymbol{\rho}}_{-i}$ has elements

$$\rho_{-i} = \frac{q_i^D}{q_i^D + q_i^R}$$

    empirical poster probabilities computed from all other users.
- π is an estimate for expected posterior probability that a Bayesian observer would correctly predict party after observing one randomly sampled token.
  - consistency assumes tokens are drawn from multinomial logit.

Tweets about mass shootings are polarized



- comparable to $\pi = .53$ in Congressional speeches (GST 2019).
- The increase in polarization over time is not statistically significant.

Questions

- How polarized are tweets about other topics (not mass shootings)?
  - why not use a tweeter fixed effect and compare to their other tweets?
  - why not show pre-trends in polarization?
- Can show polarization separately by party?
- Validating π:
  - How accurate is π at the individual level?
  - Where is the binscatter of π versus actual party affiliation?

Sentence Embeddings for Topic Assignment

1. Make a new vocabulary:
   1.1 Sample 10,000 tweets from each event
   1.2 vocabulary of stemmed words occuring at least ten times in at least three events ($N = 2000$)
2. Train GloVe embeddings on random samples of tweets from each event (samples were different sizes, this is not explained)
3. Create Arora et al (2017) embeddings:
   3.1 for each tweet t, compute weighted average vectors v_t for each word, weighted by inverse frequency.
   3.2 take out first principal component of matrix whose rows are v_t

Topics = Embedding Clusters

1. Cluster the embeddings using *k*-means
2. Identify and drop hard-to-classify tweets:
   2.1 compute ratio of distance to closest topic and distance to second-closest topic.
   2.2 drop tweets above the 75th percentile.

- Validation using Amazon Mechanical Turk:
  - Identify word intruder: five from one cluster, one from another cluster.
  - Identify tweet intruder: three from one cluster, and one from another cluster.

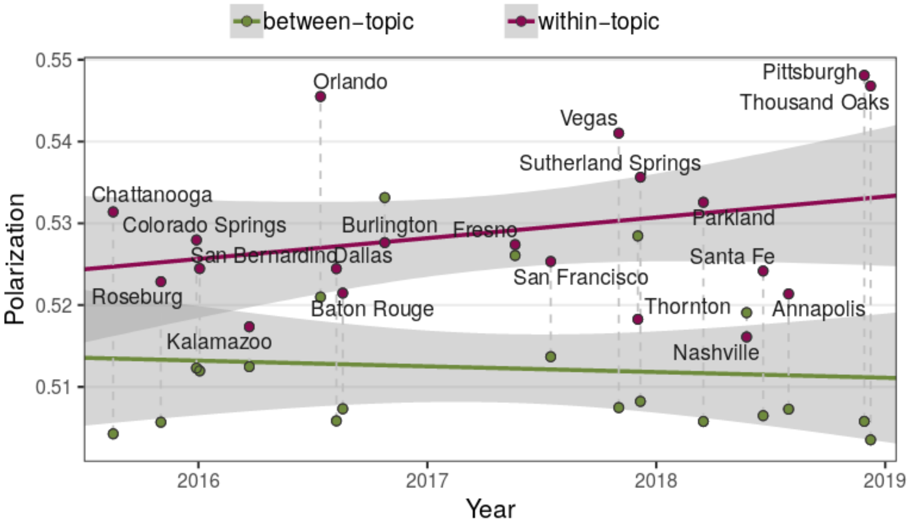
Topic Content

Topic	10 Nearest Stems
news (19%)	break, custodi, #breakingnew, #updat, confirm, fatal, multipl, updat, unconfirm, sever
investigation (9%)	suspect, arrest, alleg, apprehend, custodi, charg, accus, prosecutor, #break, ap
shooter's identity & ideology (11%)	extremist, radic, racist, ideolog, label, rhetor, wing, blm, islamist, christian
victims & location (4%)	bar, thousand, california, calif, among, los, southern, veteran, angel, via
laws & policy (14%)	sensibl, regul, requir, access, abid, #gunreformnow, legisl, argument, allow, #guncontolnow
solidarity (13%)	affect, senseless, ach, heart, heartbroken, sadden, faculti, pray, #prayer, deepest
remembrance (6%)	honor, memori, tuesday, candlelight, flown, vigil, gather, observ, honour, capitol
other (23%)	dude, yeah, eat, huh, gonna, ain, shit, ass, damn, guess

- The embedding method resulted in more coherent topics (better MTurk validation for words and tweets) than a topic model. $k = 8$ got best coherence.
  - Appendix reports samples of tweets for each topic (but does not say how samples were selected).

Between-topic vs within-topic polarization

- Within-topic polarization: compute π separately by the tweet clusters.
- Between-topic polarization: Compute π using cluster counts, rather than token counts.



Trends in within-topic polarization

- Most polarized topics: shooter's identity & ideology (.55), laws & policy (.54)

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.

- "measuring polarization of topics for other events over time is noisy".

Partisanship of Topics, by Race of Shooter

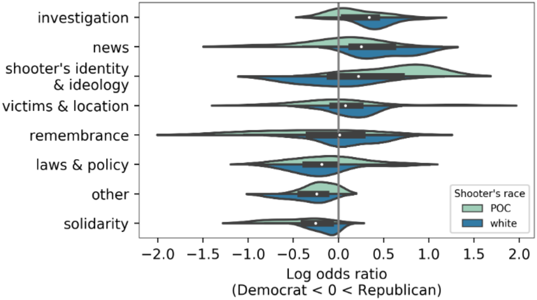


Figure 7: The plot shows the kernel density of the partisan log odds ratios of each topic (one observation per event). The white points show the median and the black rectangles the interquartile range across events.

Partisan Framing Devices: Words

- Partisanship of phrases from GST model:

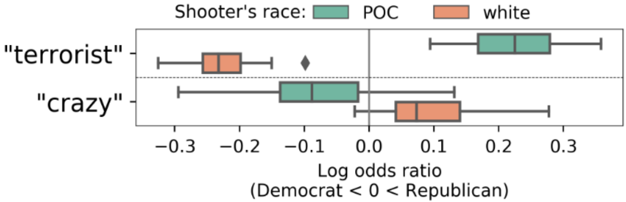
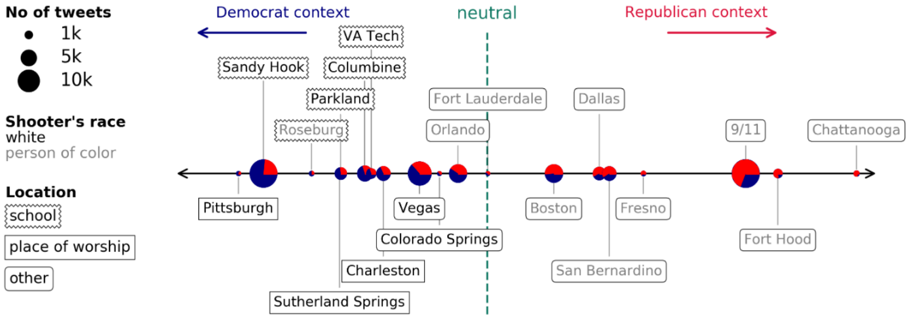


Figure 8: The log odds ratios of "terrorist" and "crazy" across events, grouped by the shooter's race. The boxes show the interquartile range and the diamond an outlier.

- Partisan valence of "terrorist" and "crazy" flip depending on race of shooter (these words have the largest racial difference in the joint vocabulary).

Partisan Framing Devices: Events

- Partisanship of keywords for previous events from GST model:



- Democrats invoke white shooters, Republicans invoke POC shooters.

Affect

- Starting point: Emotion lexicon from Mohammad and Turney (2013), available at `saifmohammad.com`.
  - 14,182 words assigned to sentiment (positive/negative) and emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust).
- Domain propagation:
  - pick 5-11 representative words per emotion category (Appendix E)
  - for each word in vocabulary, compute average distance to each member of each category. take 30 closest words as lexicon.

sadness senseless, loss, tragedi, lost, devast, sad, love, griev, horrif, terribl, pain, violenc, condol, broken, hurt, feel, victim, mourn, horrifi, will, grief, 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

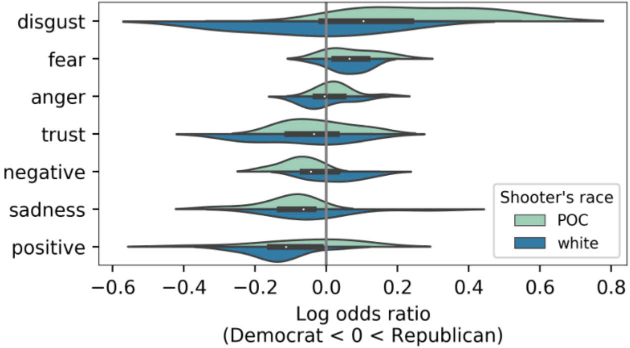
anger gun, will, murder, kill, violenc, wrong, shoot, bad, death, attack, feel, shot, action, arm, idiot, crazi, crimin, terrorist, mad, hell, crime, blame, fight, ridicul, insan, shit, die, threat, terror, hate

fear danger, threat, fear, arm, gun, still, shooter, attack, feel, fight, hide, murder, shot, shoot, bad, kill, chang, serious, 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

Partisanship of Affect Categories

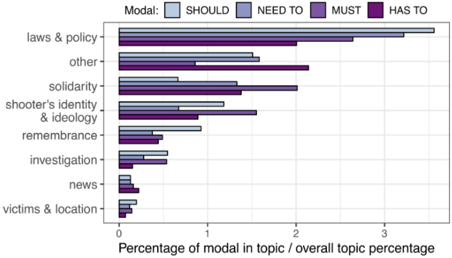
- Compute GST partisanship scores using affect-category counts:



- Disgust affect flips along partisan lines depending on race of shooter.

Modality

This roller coaster debate **MUST** STOP! Sensible gun ownership is one thing but assault weapons massacre innocent lives. The savagery of gore at #Parkland was beyond belief & **must** be the last.
In times of tragedy **shouldn't** we all come together?! Prayers for those harmed in the #PlannedParenthood shooting.
Communities **need to** step up and address white on white crime like the Las Vegas massacre. White men are out of control.
he BLM protest shooting, planned parenthood, now cali... domestic terrorism will crumble this country, SANE PPL **HAVE TO** FIGHT BACK
Shooting cops is horrible, cannot be condoned. But **must be** understood these incidents are outgrowth of decades of police abuses. #BatonRouge
1. Islamic terrorists are at war with us 2. Gun free zones = kill zones 3. Americans **should be** allowed to defend themselves #Chattanooga
Las Vegas shooting Walmart shooting and now 25 people killed in Texas over 90 people killed Mexico **should** build that wall to keep the US out
CNN reporting 20 dead, 42 injured in Orlando night club shooting. Just awful. The US **must** act to control guns or this carnage will continue.



- Count the four most frequent necessity modals in the data: should, must, have to, need to.
  - in this context, they are used as calls to action.
- Democrats use modals more than Republicans; Republicans are more fatalistic.

Comments

- This is an impressive array of NLP tools aimed at the same research question.
  - could be moving toward a standard for analyzing interpretable dimension in language.
- For all outcomes, would help to compare to other types of events, and to show pre-trends.
  - there is no baseline for polarization for comparison.
  - they do not distinguish whether outcomes are driven by different people selecting into tweeting, vs within-user changes.