Spread of Threats of Political Violence on Twitter

2020 APSA, 14th September

Taegyoon Kim

Political Science | Social Data Analytics, Penn State University

Project Overview

The era of lethal/violent partisanship

- Partisans not just hate each other [Abramowitz and Saunders, 2008] but see violence as a viable means to deal with the political opposition [Kalmoe and Mason, 2018]

Figure 1. Majorie Green (left) and Violence during a BLM protest (right)

Project Overview (cont'd)

Violent Political Metaphors

- Description of politics in violent terms
- Politicians' use of violent political metaphor
  $\rightarrow$ Issue polarization [Kalmoe et al., 2018]
  $\rightarrow$ Tolerance for political violence [Kalmoe, 2014]

Project Overview (cont'd)

Little is known about threats of political violence

Figure 2. Trump's Tweet with a Threat of Political Violence

Project Overview (cont'd)

Social media platforms, including Twitter, are an essential part of political communication

Exploitation of those platforms by both politicians and citizens who write posts threatening, supporting, or inciting political violence as part of normalized expression of extreme partisan views [Zeitzoff, 2020]

Project Overview (cont'd)

Motivating questions

- What consequences does it have for inter-party relations, attitudes toward use of political violence, etc?
- Under what circumstances do politicians and citizens use violent threats on social media?

In this project

- Detection of threats of political violence
- Spread of threats of political violence

Conceptualization

Threats of political violence

“Political rhetoric bears an intention of physical violence against a political entity (threats, support, or incitement of physical violence against political opponents).”

Example Tweets

Figure 2. Example Political Tweets with a Threat of Violence

Detection of Threats of Political violence (cont'd)

Inter-coder agreement on training set

	Coder 1&2	Coder 2&3	Coder 1&3
Cohen's Kappa	0.462	0.573	0.672
Light's Kappa		0.569	
Fless's Kappa		0.573	
Krippendorff's Alpha		0.573	

Table 1. Agreement for Three-coder Manual Labelling

Inter-coder agreement similar works

0.54 [Theocharis et al., 2016], 0.45 [Wulczyn et al., 2017], 0.37 [Munger, 2017]

Detection of Threats of Political violence (cont'd)

Performance of ML Classifiers

	Precision	Recall	F-1
Logistic Regression	79.44	7.18	13.13
Random Forest	77.31	14.03	23.69
XGBoost	71.73	10.09	17.6
BERT	66.4	65.8	66.2

Table 2. Precision, Recall, and F-1 for Various Classifiers (N=9,500)

Spread of Threats of Political Violence

“How far do Tweets containing threats of political violence spread via indirect ties?”

“Who are the intermediaries (or broadcasters)”?

“Which side uses threats of political violence more frequently, Dem. or Rep.?”

“Which politicians are targeted? Gender asymmetry? Partisan asymmetry?”

Preliminary Analysis

Data Collection

- Boolean query (political + violent) on live Tweets (scraped in Feb 2020) $\rightarrow$ Manual annotation (no ML classifiers)
- Around 2,000 Tweets containing both violent and political keywords
- Around 5-10% of the Tweets contain violent political rhetoric

Preliminary Analysis (cont'd)

- Only a small subset of violent Tweets get spread

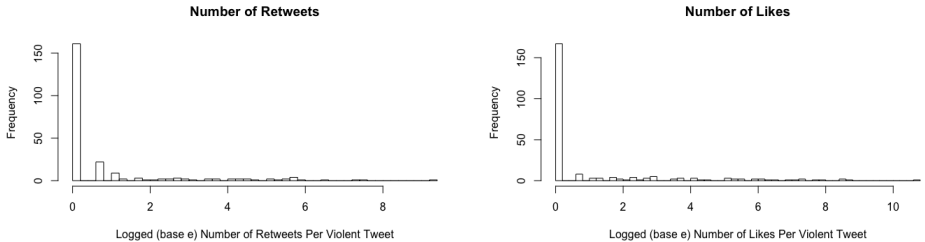

Figure 3. Distribution of Spread of Violent Tweets

Preliminary Analysis (cont'd)

- There is ideological homophily in retweeting of threats of political violence

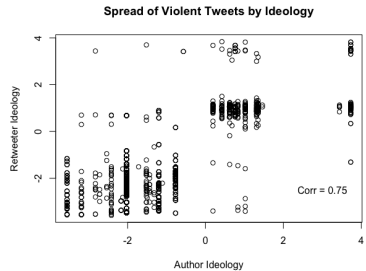

Figure 4. Ideological Homophily in the Spread of Violent Tweets

Preliminary Analysis (cont'd)

- There is *potentially* ideological asymmetry in the use of threats of political violence

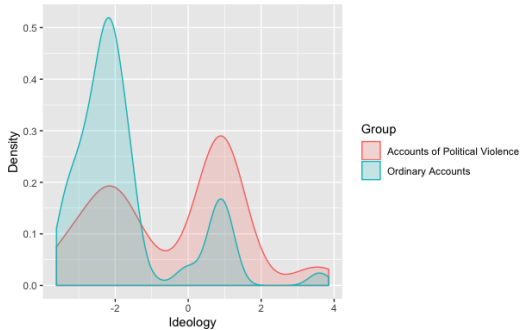

Figure 5. Ideological Asymmetry

Related Projects

How does morality and emotions affect use of threats of violence on Twitter?

- What are the consequences of exposure to online posts containing threats of political violence? How does it influence affective polarization? How does it affect attitudes toward political violence?
- How do offline political events revolving around moral issues affect (protests, legislation, or election campaign) affect use of violent threats on social media?

Q/A

Thank You!

Detection of Threats of Political Violence (back-up)

Boolean querying of Tweets through human/computer-generated “political” and “violent” keywords

1. **Boolean Querying for Political Tweets** A list of human-generated political keywords are used to filter/download Tweets via the Streaming API.
2. **Boolean Querying for Violent Tweets** A list of violent keywords are computer-generated and used to filter the downloaded “political Tweets”. A logistic regression fit on an external corpus that contains threatening Wikipedia newspaper comments and top-200 uni- or bi-gram features with the largest coefficients (e.g. shoot, kick , execute, etc) are extracted.

Detection of Threats of Political Violence (back-up)

Problems of the Boolean querying: low recall (missing out violent expressions) and low precision (violent keywords are in fact not used for a threat of violence - e.g., “killing time”).

3. **Keyword-expansion for High Recall** As in [King et al., 2017], keywords that most effectively separate violent Tweets and non-violent Tweets are extracted.

4. **Active-learning Classification for High Precision** Tweets whose class an ML classifier is uncertain about (e.g., the predicted probability around 0.5) is iteratively added/labelled.

Spread of Threats of Political Violence (back-up)

Two types of communication networks

- **Reconstruction of Following Network** A tie represents whether a user follows another user. It represents potential for information flow (exposure to violent political rhetoric)
- **Reconstruction of Retweet Network** A tie represents whether a user shared another users' Tweet. It reflects flow of information (actual adoption/spread of violent political rhetoric)

Spread of Threats of Political Violence (back-up)

Two different (but related) analyses

- **Basic characteristics of the Networks** The number of nodes, the number of ties, degree distribution, clustering, average path length, and network centralization measures to see if they are in line with other online networks.
- **Spread of Violent Tweets along Multiple Hops of Communication Ties** Users can experience unwanted exposure from users with no direct tie (the impact of a small number of violent Tweets can be amplified by intermediary users who spread them to an audience of huge size).
    - Shortest path estimation for all the pairs of Retweeter/Retweetee on the following network to reveal how far Tweets of violent political rhetoric travel.
    - Use of a betweenness-centrality measure on the following and Retweet networks to examine the distribution of influence on the spread of violent political rhetoric.

References

- Abramowitz, A. I. and Saunders, K. L. (2008).
  Is polarization a myth?
  The Journal of Politics, 70(2):542–555.
- Kalmoe, N. P. (2014).
  Fueling the fire: Violent metaphors, trait aggression, and support for political violence.
  Political Communication, 31(4):545–563.
- Kalmoe, N. P., Gubler, J. R., and Wood, D. A. (2018).
  Toward conflict or compromise? how violent metaphors polarize partisan issue attitudes.
  Political Communication, 35(3):333–352.
- Kalmoe, N. P. and Mason, L. (2018).