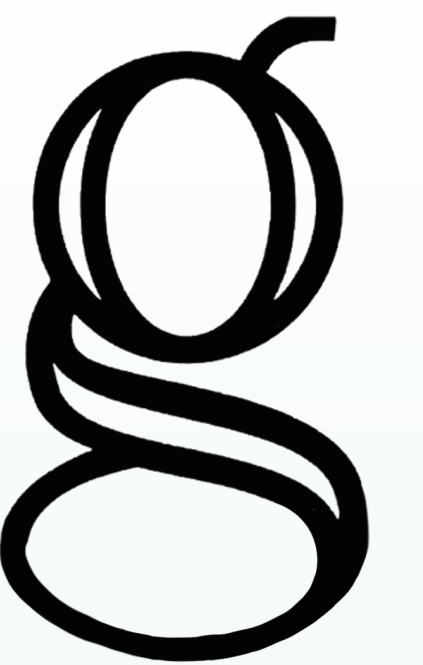




Modeling the Spread of Extremist Ideas on Reddit



Mitchell Fream, Nathan Hayes, Sahil Kochar, Paul Kolbeck, Charlie Schneider, Russell Schwartz, Olivia Sharon, Yuang Shen, Winslow Weiss, Robert Wolle
Mentor: Dr. Pierre-Emmanuel Jabin

BACKGROUND

Social media enables an unprecedented spread of ideas. Extremist ideas are no exception, and their spread online has had measurable effects in the real world (Alfano et al., 2018). We define extremist ideas as ideas that promote extremism, a radical ideological movement that embraces violence as a legitimate way to achieve its goals (Bötticher, 2017).

Developing a method for identifying extremist ideas and modeling their spread on social media could enable predictive and preventative abilities.

DATASET

- We sample text from Reddit communities with extreme ideas
- Subreddit organization is conducive to topic clustering
- Reddit comments are indexed through unique IDs and timestamps
- Interpersonal influence is quantified through positive and negative user votes on each comment

LEXICONS

- Lexicon-based approach
 - Avoids the obscurity of machine learning methods
 - Manual methods not fast enough
- Lexicons based off of features of extremist narratives (Furlow & Goodall, 2011)
 - Basing on narratives helps generalize to extreme ideas as a whole rather than targeting specific groups
- Extremist narratives separate the world into two diametrically opposed groups: "us" and "them" and then creates "a culture of fear capable of inspiring allegiance to 'us' and action against 'them.'" (Furlow & Goodall, 2011)
 - This led us to the lexicon categories:

Profanity	Problem with the World
Harassment	
Hate Speech	Us vs. Them
Violence	Insider Terminology

MODEL

- N_i - Engagement → Responsiveness of user to Reddit content
- O_i - Opinion → Frequency of extreme content generation
- V_j - Visibility → Likelihood of a comment being seen
- E_j - Extremism → Result of lexicon extremism score
- ϕ - Interaction Function → Willingness of an agent with opinion O_i to be influenced by a comment of extremism E_j .

$$\Delta O_i = N_i \frac{\sum \phi(O_i, E_j) v_j |O_i - E_j|}{\sum |\phi(O_i, E_j) v_j|}$$

FUTURE DIRECTIONS

- Clean up the lexicons, and add compounding effects to improve accuracy
- Fit model parameters to Reddit data to analyze subreddit change over time
- With the lexicons refined and the mathematical model fitted properly, analyze how extremist ideas propagate across Reddit

EXTREMISM SCORE

- Each lexicon corresponds to the categories as delineated above which is assigned an appropriate weight
- The number of hits in each category for a given comment are totaled and scaled by the corresponding weight
- The resulting sum is mapped to [0, 1) by a sigmoid function, giving the "percent" by which the comment is extreme
- Currently, we are considering how length of a comment impacts the reliability of this metric

REFERENCES AND ACKNOWLEDGEMENTS

Special thanks to Dr. Jabin, Dr. Lovell, and the Gemstone Team



References



Visualization of connections between users, measured as replies, in a single subreddit.