

CounterTerrorism

@ *Archives Unleashed 2.0*

Emily Kalah Gade, University of Washington

Dan Kerchner, George Washington University

Data Set

383,527 Tweets from:

- 1,153 accounts of suspected American and/or U.S. based ISIS supporters; curated by Program on Extremism analysts. Approx. 300 people associated with the accounts.
- 2013-2016
- Mix of English & Arabic

ALSO: NPS Ideology Classifier/Scoring Algorithm looking for key terms, based on expert analyst coding

Goals

- Apply Ideology Scoring methodology to accounts (English tweet text only)
- Use User Mentions to establish and graph account relationships
- Explore: Does similarity of ideology correlate with likelihood of interaction (expressed as Mentions)?

Techniques

Line-oriented JSON (Tweets) → Python script → 2 CSV files:

1. Per tweet: screen_name and text
2. Per mention: screen_name, mentioned screen name, date

→ R scripts →

- matrices of relationships (by time period) → network graphs
- ideology score (by screen_name)

Results

- Patterns appear to validate Program on Extremism anecdotal analysis that there are "generators" and "amplifiers"
- Generated ideology scores on small sample

Next Steps

- Visualize relationship weights
- Distinguish between retweets vs. mentions
- Time-evolve relationship graphs
- Correlate with ideology scores
- Reconsider stack: Too computationally intensive for R on Mac → leverage cloud instance w/ map-reduce cluster (e.g. use `pig`)

January 2015
Mention-relationships

