

Platform for Analyzing Social Media Data to Understand Extremist Activities

Trivalent and Saffron

Parantapa Goswami
Nadia Derbas, Frédérique Segond

May 22, 2018



VISEO

Spread of Extremism

Traditionally

Through physical interaction in social environments: places of worship, prisons, meeting venues ...

Age of Internet

Through virtual environment of the Internet, social media sites: Twitter, Facebook, Tumblr, Instagram, YouTube ...

- promoting extremist ideology and propaganda
- recruitment of individuals (foreign fighters)
- raising funds
- operational planning using encrypted communications

Using online medium ISIS successfully recruited about 25,000 foreign fighters from Europe, North America, Syria and Iraq.

— 2015 U.S government report

Countering Extremism

- Disrupting and disconnecting extremist channels: continuously search and disable social media accounts
 - Government organizations
 - Social media platforms

In response to the Paris attacks in November 2015, the hacker community Anonymous took down more than 20,000 Twitter accounts that were allegedly linked to ISIS.

- Methods deployed to categorize such accounts is mostly imperfect

Anonymous mistakenly blocked the social media accounts of the U.S president Barack Obama, the White House, the BBC, the New York Times, and many other anti-ISIS accounts.

—BBC News, 26 Nov 2015

Models of Radicalization

Multiple models have emerged from psychology and social sciences

Factors that drive people to get radicalized and indicators to detect that

- identity crisis, poverty, discrimination

Roots of radicalization

- Micro-level - individual level
- Meso-level - group or community level
- Macro-level - global level, the influence of government and society at home and abroad

Stages of how the radicalization process happens and evolves

- stages: pre-radicalization, self-identification, indoctrination, jihadisation



1 Factors and Indicators

2 Roots

3 Identification

VISEO

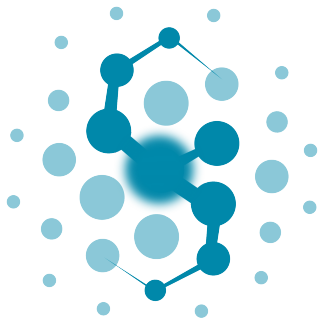


1 Factors and Indicators

2 Roots

3 Identification

WISEO



SAFFRON

- Co-funded by the Internal Security Fund of the European Union
- 2 year project (Feb 2016 - Jan 2018)
- Partners: 3 companies, 1 university, 1 law-enforcement agency
- Goal: Using ICT technologies, mainly semantic analysis, and field studies to understand recruitment communication strategies on social media, as well as the needs, cultural and social contexts that generate radicalization.

Saffron: Building Indicators

Radicalizations drivers capture the reasons why a person gets radicalized

- identity, otherness, discrimination, media stereotyping, religious issues etc.

Radicalization indicators generalize text content that may indicate a process of radicalization or online recruitment.

Drivers and indicators served as a basis to define information which needs to be extracted and structured from the social networks.

Example Indicators (created by law-enforcement experts)

- Interest manifested for activities conducted by terrorist organizations
- Increased level of violence in one's speech

Saffron: Analyzing Social Media

- Keywords
 - Semantic analysis tools developed by Eloquent, Grenoble
 - The tool extracts keywords from social media accounts
- Concepts
 - Concepts are defined by the law-enforcement and counter-terrorism experts by exploring Twitter messages
 - Indicators are linked to one or more “concepts” by experts

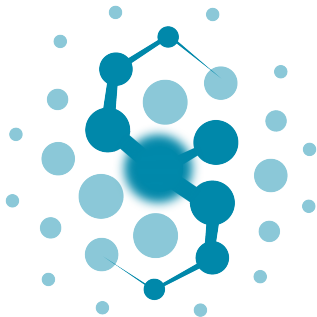
Indicator: Increased level of violence in ones speech

Related concepts : Kill, Hate, Violence ...

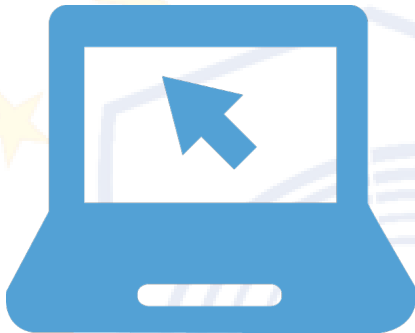
- Keywords are clustered under the defined concepts: using an topic modeling approach

Concept “Kill”: death, murder, war ...

Concept “conversion”: repent, revert umma, honor, brothers ...



SAFFRON



WISEO



1 Factors and Indicators

2 Roots

3 Identification

VISEO



- Funded by *European Union's Horizon 2020 research and innovation programme*
- Duration: 3 years (Jun 2017 - May 2020)
- Partners: : 4 industries, 5 universities, 12 law enforcement agencies (LEA), 1 research lab
- Multidisciplinary: informatics, social sciences, European legal systems, security research

Trivalent: Roots of Radicalization

Understanding the Roots of Radicalisation on Twitter.

M. Fernandez, M. Asif, H. Alani. WebSci 2018.

Participation of an User in Social Media

- creating and posting new content
- sharing content posted by someone within the network

Assumptions

- **Micro (individual)** roots: captured by posts **created** by the user.
- **Meso (social)** roots: captured by posts **shared** by the user.

Trivalent: Roots of Radicalization

Modeling Roots of Radicalization

- For an user u
 - Micro roots: set of *original* created posts: P_u^o
 - Meso roots: set of shared posts: P_u^s
- All posts in P_u^o, P_u^s are broken into n-grams (unigrams, bigrams and trigrams)
- Each set P_u^o and P_u^s is represented using the n-gram-frequency vector of all the posts it contains

Radicalization Lexicon

- Indicator concepts, keywords extracted in Saffron project
- Lexicon created by the Trivalent partner **International Institute for Counter Terrorism, Israel**
- Total 556 terms
- This lexicon is also converted to a n-gram-frequency vector similar to P_u^o and P_u^s

Trivalent: Modeling Roots of Radicalization

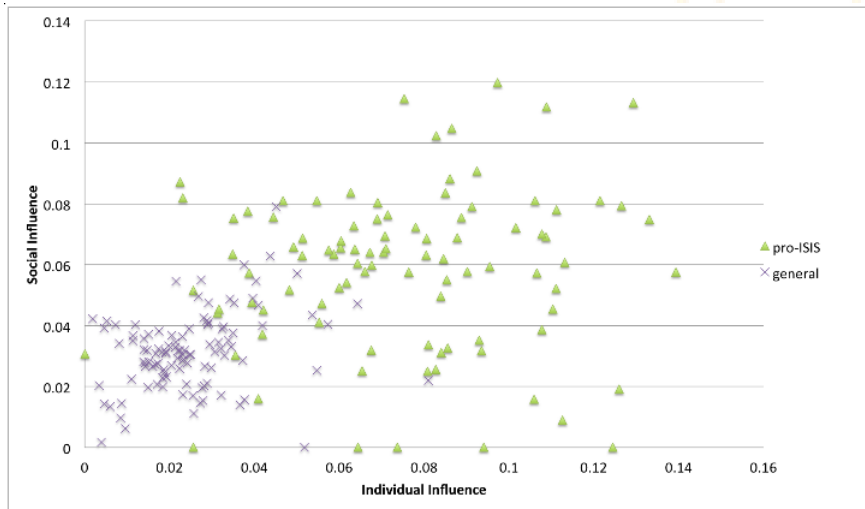
Radicalization Influence

- Micro Influence or **individual influence** of user u : cosine similarity between the representation vectors of P_u^o and radicalization lexicon
- Meso Influence or **social influence** of user u : cosine similarity between the representation vectors of P_u^s and radicalization lexicon

Dataset

- 112 pro-ISIS Twitter accounts (with 17,350 tweets) from Kaggle
- 112 not-pro-ISIS (anti-ISIS or neutral) Twitter accounts (with 197,743 tweets) randomly crawled and verified by annotators

Trivalent: Modeling Roots of Radicalization





1 Factors and Indicators

2 Roots

3 Identification

WISEO

Trivalent: Identification of extremism on social media

The meaning of “radical”

According to the UKs CONTEST strategy, radicalization refers to the process by which people come to support violent extremism and, in some cases, then to join extremist groups.

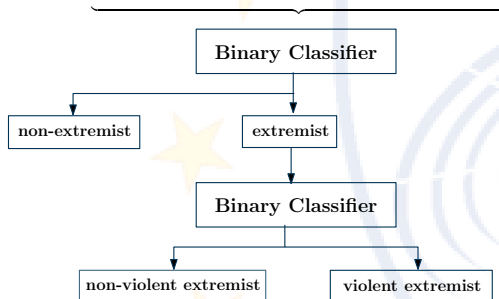
[...] to be a radical is to reject the status quo, but not necessarily in a violent or even problematic manner. [...] We can distinguish two phenomena:

- Radicalization that leads to violence, in which individuals come to undertake terrorist activity, or directly aid or abet terrorism. They are referred to as “terrorists” .
- Radicalisation that does not lead to violence, in which individuals come to hold radical views in relation to the status quo but dont undertake terrorist activity. They are referred to as “radicals” .

J. Bartlett, J. Birdwell, and M. King (2010). The edge of violence

Trivalent: Identification of extremism on social media

An hierarchical classification approach



Trivalent: Identification of extremism on social media

Twitter Dataset

- 566 users exhibiting pro-ISIS stances (Saif et al., 2017)
- 566 users exhibiting anti-ISIS stances while discussing about Middle East issues: bloggers, journalists, etc. (Saif et al., 2017)
- 64 “recognized” radical users (Rowe & Hassan, 2016)

Dataset Cleaning

- Overlapping datasets are merged
- Accounts with no English tweets are removed

Final Labeled Dataset

- 418 pro-ISIS users
- 551 non-radical users
- **Additional** 460 verified Twitter accounts (celebrities, politicians, musicians, etc)

Trivalent: Identification of extremism on social media

Feature Extraction

- Tool used: COGITO, Expert Systems semantic intelligence platform for text understanding
- For each Twitter account, COGITO extracted a total of 174,524 features

Categorize features (159): such as accounts related to

- Arrests, police custodies and detentions
- Bankruptcy offences and corporate crime
- Institutions working against terrorism

Information features (174,365): accounts mentioning

- Places: Syria, Gaza, Egypt etc.
- Organizations: United Nations, Hamas etc.
- People: Obama, Trump, Netanyahu etc.
- Other concepts: brotherhood, music, war etc.

Trivalent: Identification of extremism on social media

Feature Selection using Chi-squared Test (Forman 2003)

- test the dependency between each feature and the labels (extremist/non-extremist)
- 2,000 features retained with the highest χ^2 value (higher χ^2 value means higher “discriminating power”)
 - Categorize: 157
 - Information: 1,843 (places - 200, organizations - 137, people - 87, other - 1,419)

Classifier

- Random forest
- 10-fold cross validation
- Results (average of 10-folds)
 - Precision = 96%
 - Recall = 84%
 - F-score = 89%

Classifier Testing

- Tested on 55 held-out radical users
 - Recall 77%
- Tested with previously unseen non-radical and general users
 - Recall 100%

- Influencer detection
 - Active extremists are social media influencers
 - Influencer detection methodologies can be applied to detect extremist influencers
- Representation learning for complex social media data
 - Deep learning techniques

Thank you!
Questions, Comments...



WISEO