Identifying political hate speech and Analysing its Spread

Malvika Mohan
Julius Coleman
Meghana Parab

Information School
UNIVERSITY *of* WASHINGTON

W

What is the problem?

"**The Trump Effect**" is a term coined after an increase in bullying in school post the 2016 elections but has expanded to include religious and racial hate speech/bullying by adults during his presidential term

Donald J. Trump
@realDonaldTrump

I always treated the Chinese Virus very seriously, and have done a very good job from the beginning, including my very early decision to close the "borders" from China - against the wishes of almost all. Many lives were saved. The Fake News new narrative is disgraceful & false!

4:46 AM · Mar 18, 2020 · Twitter for iPhone

56.5K Retweets 25.1K Quote Tweets 291.7K Likes

NEWS   POLITICS   COVID-19   U.S. NEWS   OPINION   BUSINESS   WORLD

AMERICA'S DIVERSE VOTE

After Trump's Covid-19 diagnosis, anti-Asian tweets and conspiracies rose 85%: report
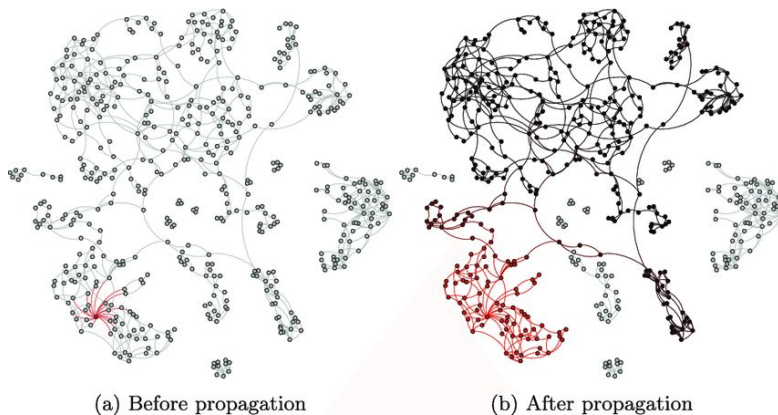
Information School
UNIVERSITY of WASHINGTON

Why is it hard?

- Classification of hate speech is limited to hateful keyword detection

- Challenges in detecting early onset and triggers of hate speech

(a) Before propagation   (b) After propagation

Information School
UNIVERSITY *of* WASHINGTON

What are we trying to do

- Identify political hate speech on Twitter during Trump's presidential term and determine how its spread has changed prior to his tenure
- Identify potential triggers/causes of spread in hate speech during Trump's presidential term

Current Landscape – Literature Review

How is it done today :

- Hate speech detection limited to the 2016 election [1]
- Crowdsource labelling [2]

Limitations :

- Lack of analysis of hate speech over a span of time
- Reliability of crowdsourced data

What's new in your approach and why do you think it will be successful?

- **Full 4 year presidential term**

    - Gives us a expansive set of data to analyze

    - Interested in the President's influence while in office rather just a election cycle

    - Compare hate speech spread before/after the 2016-2020 presidential term

Who cares?

- **The Public**

    - Informs them on the direct effects of such online behavior from a political figure

- **Social Media Platforms**

    - Truly grasp the dangers of allowing such behavior on their platforms without

        regulation

If you are successful what difference will it make?

- **The Public**

    - Allows them to make more informed decisions when it comes to choosing a

        candidate.

- **Social Media Platforms**

    - Can be used to help create better policies when it comes to political figures.

Expected Innovation and Creativity

Our project will analyze the spread of hate speech throughout the 4-year presidential term from 2016-2020

Most research done in this area has focused on hate speech only for the few months preceding the elections

Information School
UNIVERSITY *of* WASHINGTON

W

Data Sources

- http://trumptwitterarchive.com/ - archive of all of trump's tweets (around 56,571 tweets)
- Data from twitter API based on search queries (20-25k for each of the four years of his presidency).
- (https://hatebase.org/) - lexicon of hate speech vocabulary terms to classify tweets.

Information School
UNIVERSITY *of* WASHINGTON

List of Activities

Timeline	Deliverable
Week 1	**Gather data, do exploratory analysis, and data cleaning**
Week 2	**Build machine learning model (focusing on text analysis,NLP) to classify tweets as hate/normal**
Week 3	**Evaluate model performance and fine-tune the ML model. Build visualizations to analyse the spread of hate speech.**
Week 4	**Create final project document and presentation**

Information School
UNIVERSITY *of* WASHINGTON

W

Cloud recording URL

https://washington.zoom.us/rec/share/gfoRWrGzNFZdEnwutBtLeYDH9e0Fy9kc8t5J_kzl-UZOa4w3TFI8UJqs2KDud9s.ONgeCtwfvVBnuX5o

References

Edwards, G. S. (2018, January 14). The Effect of President Trump's Election on Hate Crimes by Griffin Sims

      Edwards, Stephen Rushin :: SSRN. Retrieved from

      https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3102652

Davidson, T., Warmsley, D.,Macy, M. & Weber,I. (2017, March 11). Automated Hate Speech Detection and the

      Problem of Offensive Language. Retrieved from https://arxiv.org/pdf/1703.04009.pdf

Information School
UNIVERSITY *of* WASHINGTON

W

Thank You!

Information School
UNIVERSITY *of* WASHINGTON