

Understanding adversarial interactions against politicians on social media

Yiqing Hua, Cornell Tech

Dataset

Twitter activities from Sept, 2018 to Nov of **786** candidates running for U.S. House of Representatives (87%): **431** (D), **355** (R). **231** female, **555** male. In total, we have **1.7M** candidate posted tweets, replies and mentions, **6.5M** retweets of candidates from **1M** users.

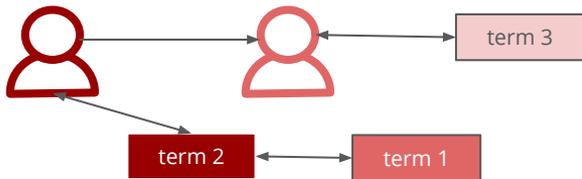
Methodology

Characterizing **candidates** who receive adversarial interactions:

- Identify adversarial interactions using Toxicity scoring.
- Quantify **directed** adversarial interactions per candidate **via user party preference**.
- Correlate candidate characteristics with amount of directed adversarial interactions received **using logistic regression**

Discovering subtle adversarial interactions inspired by SENTPROP:

- Construct User-Term graph per candidate
- Assign weights to adversarial users
- Propagate weights using random walk



Results:

Candidates who receive adversarial Interactions:

Overall attention is the main predictor of adversarial interactions, while candidate **party or gender** are **not** significant factors.

Adversarial interactions may be **subtle** and **tailored** to their targets, often **missed** by general language detection tools.

Published in **[ICWSM 2020]**.

Users who engage in Adversarial Interactions:

10% of the users created over **35%** of the adversarial interactions.

Adversarial users exhibit **different** behavioral patterns than normal user, showing a tendency to **seek out conflicts**.

They involve in **fewer supportive** interactions and pay more attention to **opponent** candidates.

Published in **[CHI 2020]**.

Find my papers on yiqing-hua.com

Contact: yiqing@cs.cornell.edu