

Towards Measuring Adversarial Twitter Interactions against Candidates in the US Midterm Elections

Yiqing Hua

Thomas Ristenpart

Mor Naaman

Cornell Tech, Cornell University

Presentation for ICWSM 2020

Find the slides and paper on yiqing-hua.com



- Connect with constituents
- Express opinions
- Campaign for the race

Thank you and
agreed!

That's bad idea.



Medicare for all
NOW!

Thank you and agreed!

That's bad idea.



Most annoying woman ever...

SHUT UP!

Medicare for all NOW!

Alexandria
Occasionally-Coherent!

#!\$@#!% ^!#%\$%

!@#\$



- Quantify **directed** adversarial interactions per candidate
- Discover **candidate-specific** adversarial interactions

#!\$@#!%^!#%\$%

!@#\$



- Quantify **directed** adversarial interactions per candidate
- Discover **candidate-specific** 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.

Previous Work

Quantify **directed** adversarial interactions per candidate

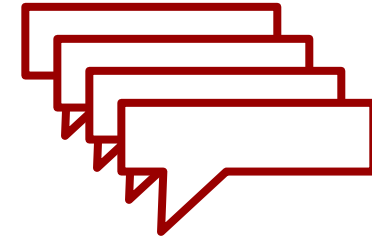
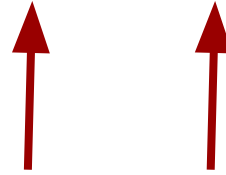
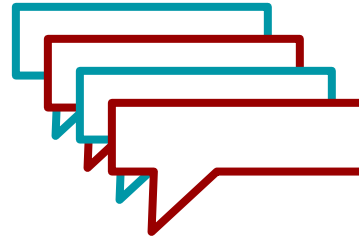
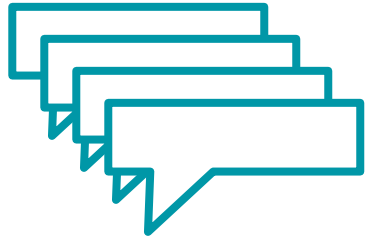
Measure **general** adversarial interactions

Mondal et al. (2017); Finkelstein et al. (2018); Chatzakou et al. (2017); Ribeiro et al. (2018); Gorrell et al. (2018)

Discover **candidate-specific** adversarial interactions

Categorize **general** adversarial interactions

Founta et al. (2018); Matias et al. (2015); ElSherief et al. (2018)



Data Collection

Identify Adversarial Interactions

Quantify **Directed** Adversarial Interactions per Candidate

U.S. Midterm Election Twitter Dataset 2018

1.7M Twitter interactions with **validated** candidate accounts from **786** candidates running for U.S. House of Representatives (87%) between **September 17th, 2018** to **November 6th**.

Dataset published on [Figshare](#)
Find the link at yiqing-hua.com



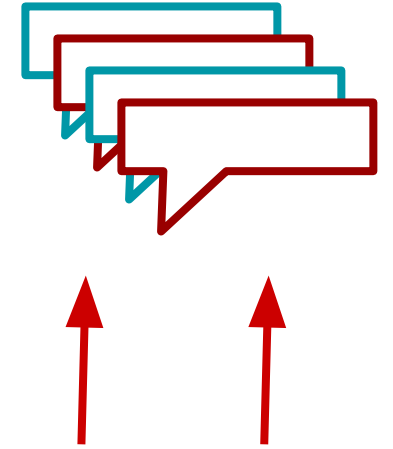
yiqing-hua.com

Adversarial Interactions



Behaviors on social media that intended to **hurt, embarrass, or humiliate** a targeted individual.

Identify Adversarial Interactions



Use **Toxicity** scoring from Perspective API to identify adversarial interactions

"a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion."

We took several steps to validate this approach, please refer to details in the paper.

Adversarial, but directed at whom?



Omar Navarro ✓ @RealOmarNavarro · Oct 2

When is Maxine Waters going to be held accountable for her actions? She's asking our government to look into things and obviously they go through with it. Like the FBI being dished at me for sharing a letter. It's time they investigate her takeout threat.

287 2.6K 4.2K

Replying to @RealOmarNavarro @

And sheeeee ugly!

6:06 AM - 3 Oct 2018

Interaction icons: reply, retweet, like, direct message

Follow

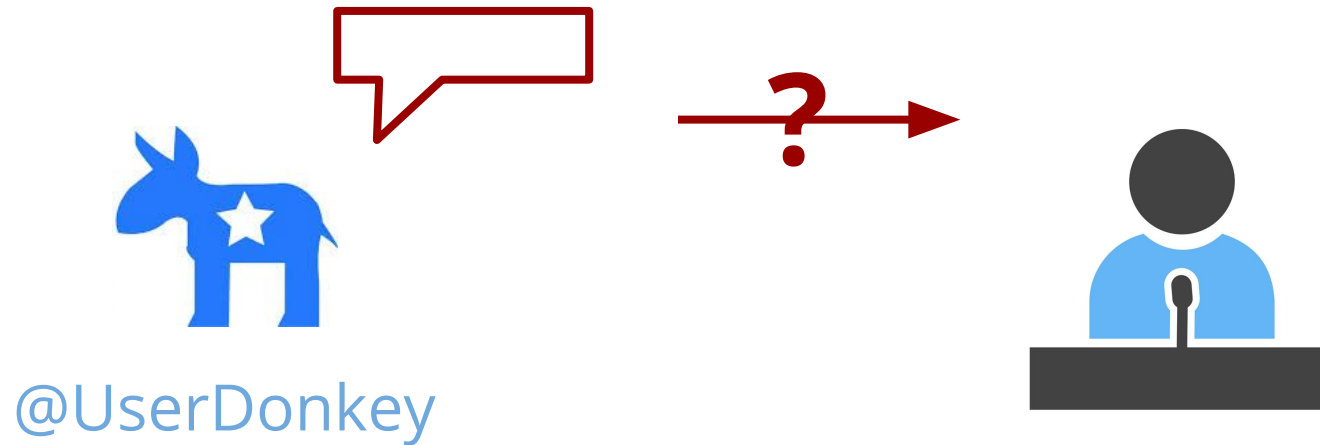


Maxine Waters ✓
@RepMaxineWaters



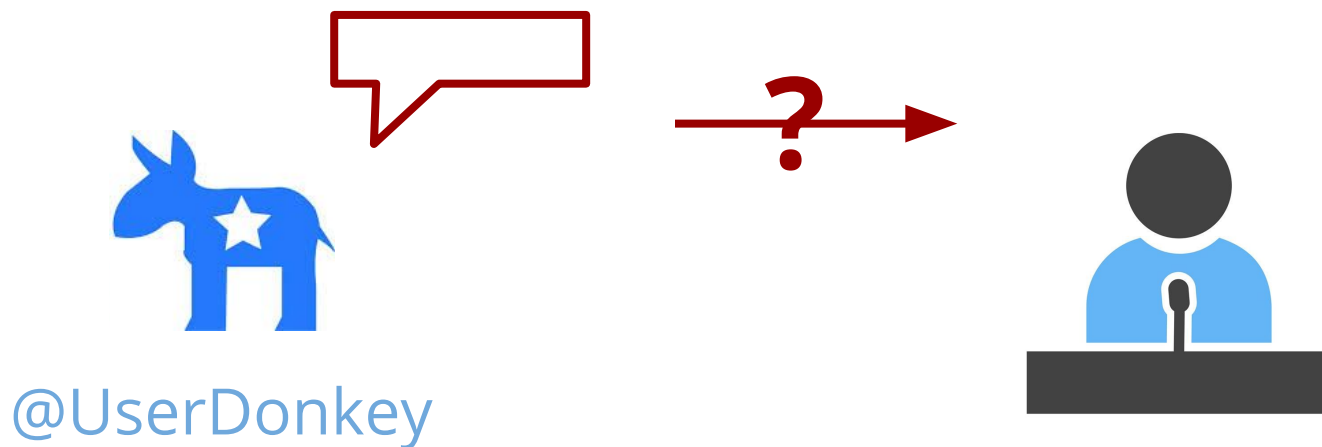
Omar Navarro ✓
@RealOmarNavarro

Directionality with Party Preference



Heuristic: when the **user party preference** is the **same** as candidate's affiliated party, the user is **not** adversarial directed at the candidate.

Directionality with Party Preference



Heuristic: when the **user party preference** is the **same** as candidate's affiliated party, the user is **not** adversarial directed at the candidate.

- Hashtags in user profile [Conover (2011)]
- Retweet Pattern [Conover (2011)]
- Following Pattern [Barberá (2015)]

 **User party preference**

What *Candidate Characteristics* attract more *Directed Adversarial Interactions*?

Control

Attention received from opponent users
Number of followers



Candidate Attributes

Gender
Affiliated party

What *Candidate Characteristics* attract more *Directed Adversarial Interactions*?

Control

- + Attention received from opponent users
- Number of followers



$$R^2 = 0.388$$

Candidate Attributes

Gender
Affiliated party

Limitations of General Language Detection Tools

@IlhanMN did you and your **brother** have fun on your honeymoon?

Fact Checks › Politics

Did U.S. Rep. Ilhan Omar Marry Her Brother?

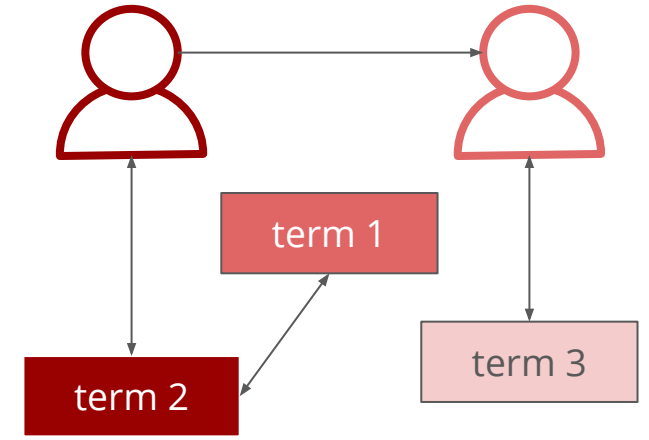
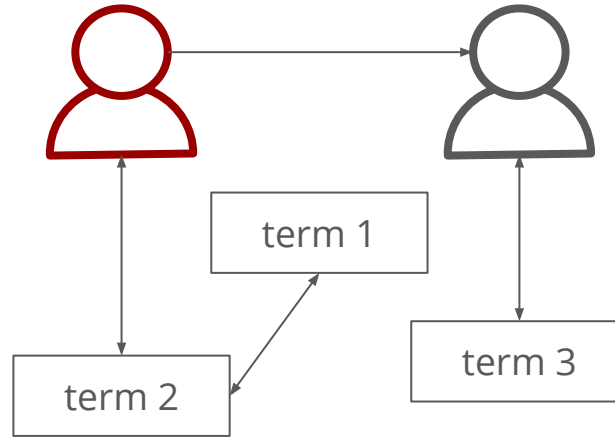
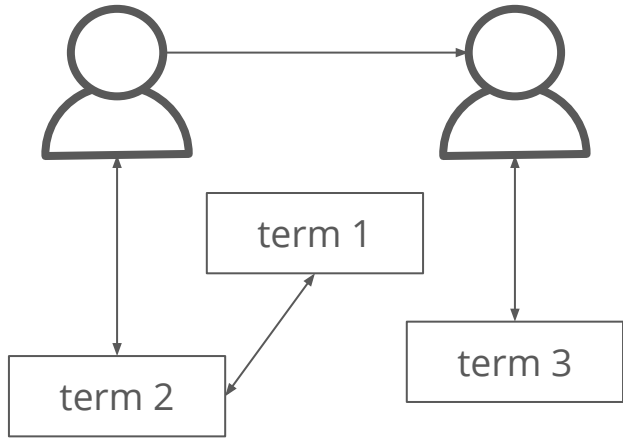
Rep. Ilhan Omar has been the subject of recurrent controversies, including ones regarding her marriages and family immigration history.

BETHANIA PALMA

PUBLISHED 15 FEBRUARY 2019



Image via [Wikimedia Commons](#)



Construct
User-Term Graph
per Candidate

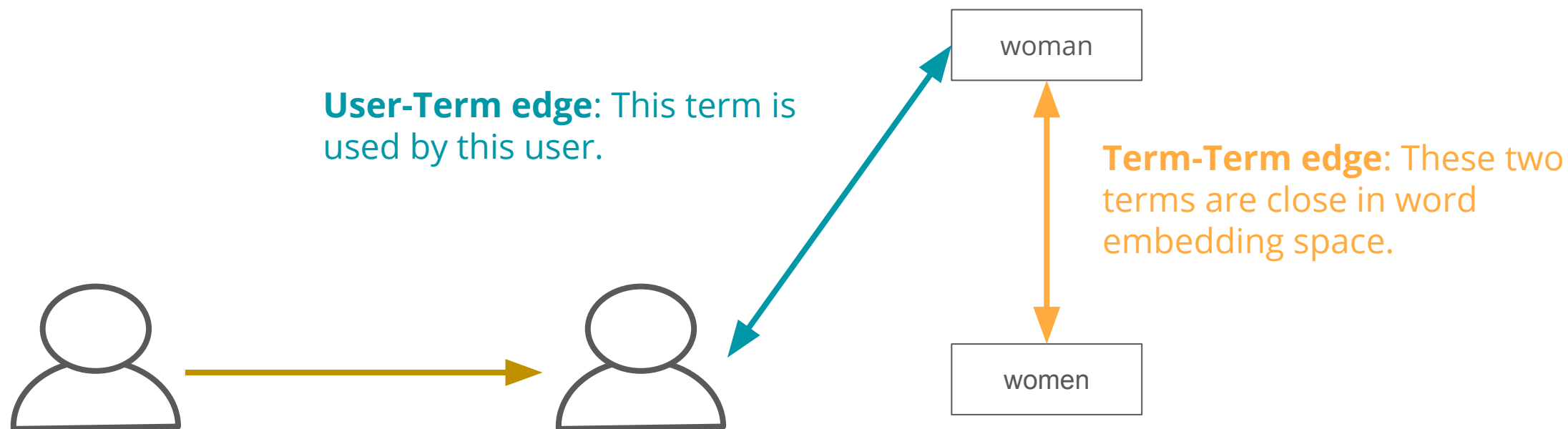
Assign weights to
Adversarial Users

Propagate Weights using
random walk

Based on SENTPROP (Hamilton et al.
2016)

User-term graph to discover candidate-specific adversarial terms

Based on SENTPROP (Hamilton et al. 2016)



User-User edge: This user follows the other user.

User-Term edge: This term is used by this user.

Term-Term edge: These two terms are close in word embedding space.

#!\$@#!% ^!#%\$%

!@#\$



- Quantify **directed** adversarial interactions per candidate
- Discover **candidate-specific** 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.

Adversarial interactions with political candidates

[Characterizing Twitter Users Who Engage in Adversarial Interactions against Political Candidates.](#) [CHI2020]

[Towards Measuring Adversarial Twitter Interactions against Candidates in the US Midterm Elections.](#) [ICWSM2020]

yiqing-hua.com

yiqing@cs.cornell.edu