

Tracking the spread of misinformation on Twitter.  
Evidence from the 2016 U.S. election.

Pablo Barberá

**Networked Democracy Lab**

USC

[netdem.org](http://netdem.org)

David García

**Chair of Systems Design**

ETH Zürich

[dgarcia.eu](http://dgarcia.eu)



## SCOTT FOVAL, NATIONAL FIELD DIRECTOR AT AMERICANS UNITED FOR CHANGE

You remember the Iowa state fair thing where Scott Walker grabbed the sign out of the dude's hand and then the dude gets kind of roughed up right in front of the stage right there on camera?



12:05 / 16:26

VISIT PROJECTVERITASACTION CC N. HD M D

## Rigging the Election - Video I: Clinton Campaign and DNC Incite Violence at Trump Rallies

Project Veritas Action  
The logo for Project Veritas Action Fund, featuring the word "Veritas" in a stylized font with "PROJECT" above it and "ACTION FUND" below it.



122,118

6,420,507 views



RADIO SHOW NEWS VIDEOS STORE TOP STORIES BREAKING NEWS CONTACT



## "SPIRIT COOKING": CLINTON CAMPAIGN CHAIRMAN PRACTICES BIZARRE OCCULT RITUAL

Menstrual blood, semen and breast milk: Most bizarre Wikileaks revelation yet

Paul Joseph Watson | Infowars.com - NOVEMBER 4, 2016

4067 Comments



## TODAY ON THE ALEX JONES SHOW

On this **Thursday, Feb. 9** edition of the Alex Jones Show, we discuss Trump's 'PHENOMENAL' tax cuts coming soon and how the White House needs to floor the accelerator to...

READ MORE



Clinton    Corruption

## Doctor Who Removed Clinton's Blood Clot Found Dead

October 13, 2016    Christopher Kermett    28 Comments



## BREAKING: FBI Now 99% Certain at Least 5 Foreign Agencies Hacked Into Hillary's Server

Jim Hoft Nov 2nd, 2016 6:31 pm – 432 Comments

FBI Now 99% Certain at Least 5 Foreign Agencies Hacked Into Hillary's Server



What do these news stories have in common?

What do these news stories have in common?

They were widely shared  
on Twitter

# Misinformation was widely shared on Twitter

## Rigging the Election - Video I: Clinton Campaign and DNC Incite Violence at Trump Rallies

55K shares

**"SPIRIT COOKING": CLINTON CAMPAIGN CHAIRMAN PRACTICES BIZARRE OCCULT RITUAL**

Menstrual blood, semen and breast milk: Most bizarre WikiLeaks revelation yet

[Paul Joseph Watson](#) | Infowars.com - NOVEMBER 4, 2016

4067 Comments



22K shares

Clinton Corruption

## Doctor Who Removed Clinton's Blood Clot Found Dead

17K shares

## BREAKING: FBI Now 99% Certain at Least 5 Foreign Agencies Hacked Into Hillary's Server

[Jim Hoft](#) Nov 2nd, 2016 6:31 pm — 432 Comments

15K shares

# Rigging the Election - Video I: Clinton Campaign and DNC Incite Violence at Trump Rallies

55K shares

**"SPIRIT COOKING": CLINTON CAMPAIGN CHAIRMAN PRACTICES BIZARRE OCCULT RITUAL**

Menstrual blood, semen and breast milk: Most bizarre WikiLeaks revelation yet

[Paul Joseph Watson](#) | Infowars.com - NOVEMBER 4, 2016

4067 Comments



Clinton Corruption

## Doctor Who Removed Clinton's Blood Clot Found Dead

October 13, 2016 Christopher Kemmett 28 Comments

22K shares

## BREAKING: FBI Now 99% Certain at Least 5 Foreign Agencies Hacked Into Hillary's Server

[Jim Hoft](#) Nov 2nd, 2016 6:31 pm — 432 Comments

17K shares

The New York Times

## *Emails in Anthony Weiner Inquiry Jolt Hillary Clinton's Campaign*

15K shares

 **politics**

Election Results Nation W

Live election results and coverage

10K shares

15K shares

# Misinformation has real consequences

MEDIA

The New York Times

## *In Washington Pizzeria Attack, Fake News Brought Real Guns*

---

By CECILIA KANG and ADAM GOLDMAN DEC. 5, 2016

---

# Misinformation has real consequences

MEDIA

The New York Times

## *In Washington Pizzeria Attack, Fake News Brought Real Guns*

---

By CECILIA KANG and ADAM GOLDMAN DEC. 5, 2016

---

ASIA PACIFIC

The New York Times

## *Reading Fake News, Pakistani Minister Directs Nuclear Threat at Israel*

---

By RUSSELL GOLDMAN DEC. 24, 2016

## What we know so far

“Top fake election news stories generated **more total engagement on Facebook** than top election stories from 19 major news outlets combined”

Craig Silverman, *Buzzfeed*

## What we know so far

“Top fake election news stories generated **more total engagement on Facebook** than top election stories from 19 major news outlets combined”

Craig Silverman, *Buzzfeed*

“The average American saw and remembered 0.92 pro-Trump fake news stories and 0.23 pro-Clinton fake news stories, with just **over half of those who recalled seeing fake news stories believing them**”

Allcott and Gentzkow (2017)

*Social Media and Fake News in the 2016 Election*

## What we don't know

- ▶ What proportion of social media posts during the campaign were spreading misinformation?

## What we don't know

- ▶ What proportion of social media posts during the campaign were spreading misinformation?
- ▶ Who was sharing misinformation? What % were bots?

## What we don't know

- ▶ What proportion of social media posts during the campaign were spreading misinformation?
- ▶ Who was sharing misinformation? What % were bots?
- ▶ Did ideological polarization contribute to the spread of misinformation?

## What we don't know

- ▶ What proportion of social media posts during the campaign were spreading misinformation?
- ▶ Who was sharing misinformation? What % were bots?
- ▶ Did ideological polarization contribute to the spread of misinformation?
- ▶ What was the reach of misinformation? Who was exposed to it?

## What we don't know

- ▶ What proportion of social media posts during the campaign were spreading misinformation?
- ▶ Who was sharing misinformation? What % were bots?
- ▶ Did ideological polarization contribute to the spread of misinformation?
- ▶ What was the reach of misinformation? Who was exposed to it?
- ▶ Why was misinformation so prevalent?

## What we don't know

- ▶ What proportion of social media posts during the campaign were spreading misinformation?
- ▶ Who was sharing misinformation? What % were bots?
- ▶ Did ideological polarization contribute to the spread of misinformation?
- ▶ What was the reach of misinformation? Who was exposed to it?
- ▶ Why was misinformation so prevalent?
  - ▶ Emotional valence of misinformation: anger, disgust

## What we don't know

- ▶ What proportion of social media posts during the campaign were spreading misinformation?
- ▶ Who was sharing misinformation? What % were bots?
- ▶ Did ideological polarization contribute to the spread of misinformation?
- ▶ What was the reach of misinformation? Who was exposed to it?
- ▶ Why was misinformation so prevalent?
  - ▶ Emotional valence of misinformation: anger, disgust

# Our two datasets

## **Keyword-based dataset**

- ▶ 83.8m tweets mentioning “hillary”, “clinton”, “hillaryclinton”, “#i’mwithher” Oct. 10 to Nov. 9, 2016.

# Our two datasets

## **Keyword-based dataset**

- ▶ 83.8m tweets mentioning “hillary”, “clinton”, “hillaryclinton”, “#i’mwithher” Oct. 10 to Nov. 9, 2016.
- ▶ 24.1m tweets with URLs from 84,429 unique domains

# Our two datasets

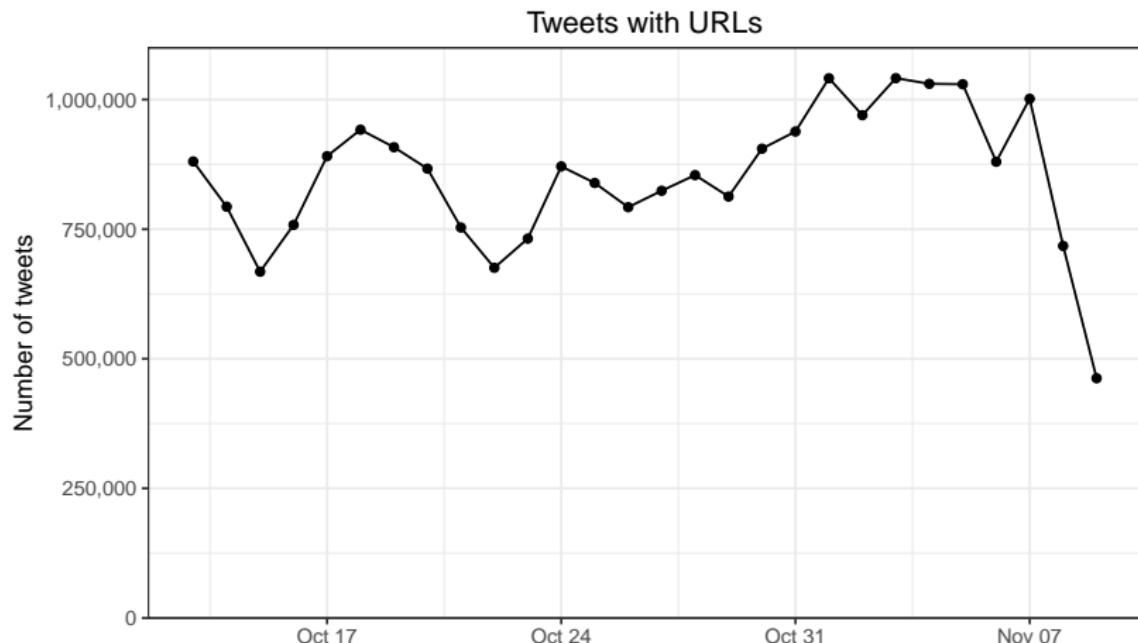
## **Keyword-based dataset**

- ▶ 83.8m tweets mentioning “hillary”, “clinton”, “hillaryclinton”, “#i’mwithher” Oct. 10 to Nov. 9, 2016.
- ▶ 24.1m tweets with URLs from 84,429 unique domains
- ▶ After cleaning/expanding URLs, 3.8 million unique URLs

# Our two datasets

## Keyword-based dataset

- ▶ 83.8m tweets mentioning “hillary”, “clinton”, “hillaryclinton”, “#i’mwithher” Oct. 10 to Nov. 9, 2016.
- ▶ 24.1m tweets with URLs from 84,429 unique domains
- ▶ After cleaning/expanding URLs, 3.8 million unique URLs



# Our two datasets

## **Keyword-based dataset**

- ▶ 83.8m tweets mentioning “hillary”, “clinton”, “hillaryclinton”, “#i’mwithher” Oct. 10 to Nov. 9, 2016.
- ▶ 24.1m tweets with URLs from 84,429 unique domains
- ▶ After cleaning/expanding URLs, 3.8 million unique URLs

## **User-based dataset**

# Our two datasets

## **Keyword-based dataset**

- ▶ 83.8m tweets mentioning “hillary”, “clinton”, “hillaryclinton”, “#i’mwithher” Oct. 10 to Nov. 9, 2016.
- ▶ 24.1m tweets with URLs from 84,429 unique domains
- ▶ After cleaning/expanding URLs, 3.8 million unique URLs

## **User-based dataset**

- ▶ 205K Twitter accounts matched with voter records

# Our two datasets

## **Keyword-based dataset**

- ▶ 83.8m tweets mentioning “hillary”, “clinton”, “hillaryclinton”, “#i’mwithher” Oct. 10 to Nov. 9, 2016.
- ▶ 24.1m tweets with URLs from 84,429 unique domains
- ▶ After cleaning/expanding URLs, 3.8 million unique URLs

## **User-based dataset**

- ▶ 205K Twitter accounts matched with voter records
- ▶ 4.6m tweets with URLs Oct. 3 to Nov. 9, 2016.

# Our two datasets

## Keyword-based dataset

- ▶ 83.8m tweets mentioning “hillary”, “clinton”, “hillaryclinton”, “#i’mwithher” Oct. 10 to Nov. 9, 2016.
- ▶ 24.1m tweets with URLs from 84,429 unique domains
- ▶ After cleaning/expanding URLs, 3.8 million unique URLs

## User-based dataset

- ▶ 205K Twitter accounts matched with voter records
- ▶ 4.6m tweets with URLs Oct. 3 to Nov. 9, 2016.
- ▶ 688K tweets with URLs from “hard news” websites  
(Bakshy et al 2015)

# Our two datasets

## Keyword-based dataset

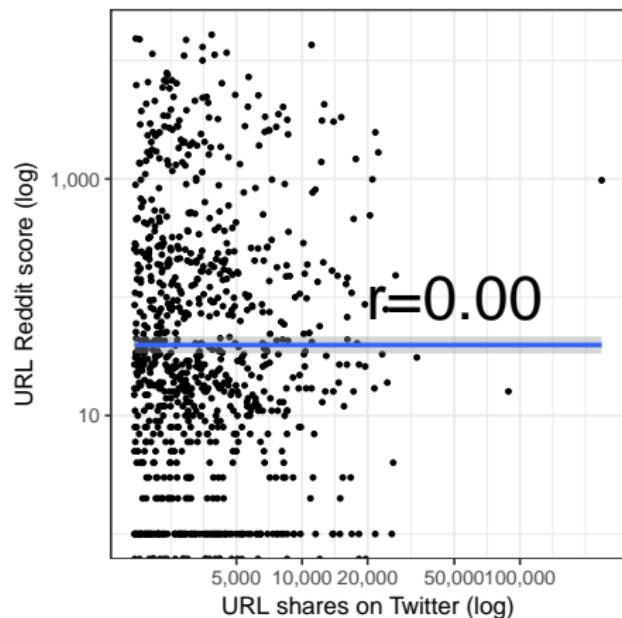
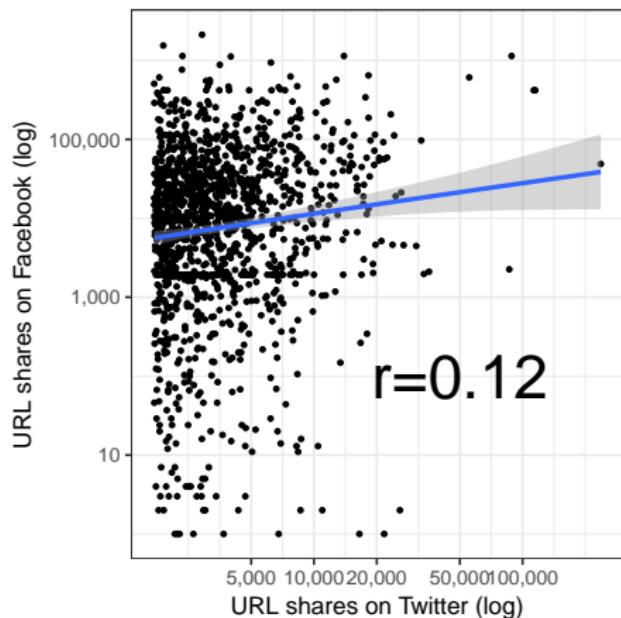
- ▶ 83.8m tweets mentioning “hillary”, “clinton”, “hillaryclinton”, “#i’mwithher” Oct. 10 to Nov. 9, 2016.
- ▶ 24.1m tweets with URLs from 84,429 unique domains
- ▶ After cleaning/expanding URLs, 3.8 million unique URLs

## User-based dataset

- ▶ 205K Twitter accounts matched with voter records
- ▶ 4.6m tweets with URLs Oct. 3 to Nov. 9, 2016.
- ▶ 688K tweets with URLs from “hard news” websites  
(Bakshy et al 2015)
- ▶ 141K tweets with URLs and same keywords as first dataset.

# One quick caveat about Twitter data

Social media metrics of top 1,500 most shared URLs on Twitter



**Data:** Facebook Graph API, Reddit API

# Defining misinformation

How to define [misinformation](#)?

# Defining misinformation

How to define **misinformation**?

Berinsky, 2015; Bullock et al, 2015; Gaines et al, 2007; Kuklinski et al, 2000;  
Nyhan et al, 2013, ... many others

# Defining misinformation

How to define **misinformation**?

Berinsky, 2015; Bullock et al, 2015; Gaines et al, 2007; Kuklinski et al, 2000;  
Nyhan et al, 2013, ... many others

**Our definition:**

1. News stories about politics (not entertainment)

# Defining misinformation

How to define **misinformation**?

Berinsky, 2015; Bullock et al, 2015; Gaines et al, 2007; Kuklinski et al, 2000;  
Nyhan et al, 2013, ... many others

**Our definition:**

1. News stories about politics (not entertainment)
2. Pretend to present facts (not opinion)

# Defining misinformation

How to define **misinformation**?

Berinsky, 2015; Bullock et al, 2015; Gaines et al, 2007; Kuklinski et al, 2000;  
Nyhan et al, 2013, ... many others

**Our definition:**

1. News stories about politics (not entertainment)
2. Pretend to present facts (not opinion)
3. Demonstrably false or misleading (excluding satire)

# Defining misinformation

## How to define **misinformation**?

Berinsky, 2015; Bullock et al, 2015; Gaines et al, 2007; Kuklinski et al, 2000;  
Nyhan et al, 2013, ... many others

### Our definition:

1. News stories about politics (not entertainment)
2. Pretend to present facts (not opinion)
3. Demonstrably false or misleading (excluding satire)

**Data:** 146 domains that produce mostly ‘fake news’:

breitbart.com, dailycaller.com, therealstrategy.com, truthfeed.com,  
thegatewaypundit.com, infowars.com, lifezette.com, zero hedge.com,  
newsninja2012.com, feed24hnews.com, freebeacon.com, ...

# Defining misinformation

## How to define **misinformation**?

Berinsky, 2015; Bullock et al, 2015; Gaines et al, 2007; Kuklinski et al, 2000;  
Nyhan et al, 2013, ... many others

### Our definition:

1. News stories about politics (not entertainment)
2. Pretend to present facts (not opinion)
3. Demonstrably false or misleading (excluding satire)

**Data:** 146 domains that produce mostly ‘fake news’:

breitbart.com, dailycaller.com, therealstrategy.com, truthfeed.com,  
thegatewaypundit.com, infowars.com, lifezette.com, zero hedge.com,  
newsninja2012.com, feed24hnews.com, freebeacon.com, ...

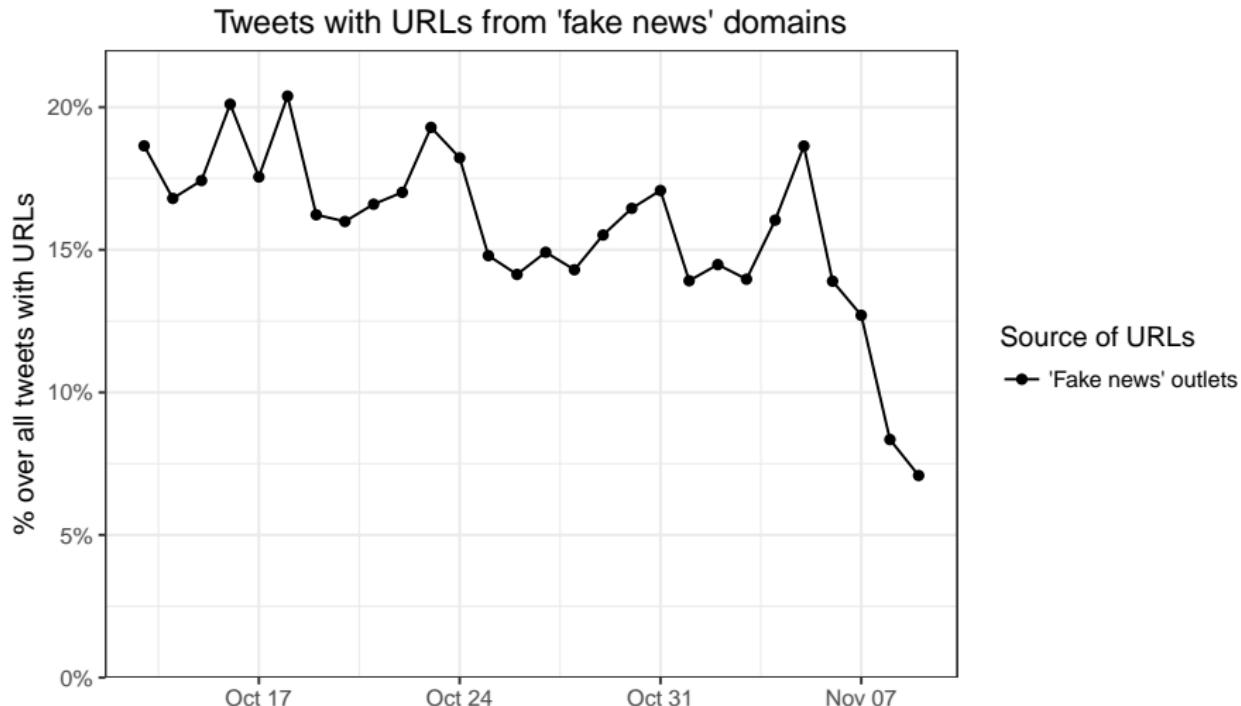
**Source:** based on list compiled by Melissa Zimdars and most  
popular domains in our dataset (manually checked)

What proportion of social media posts during the campaign were spreading misinformation?

What proportion of social media posts during the campaign were spreading misinformation?

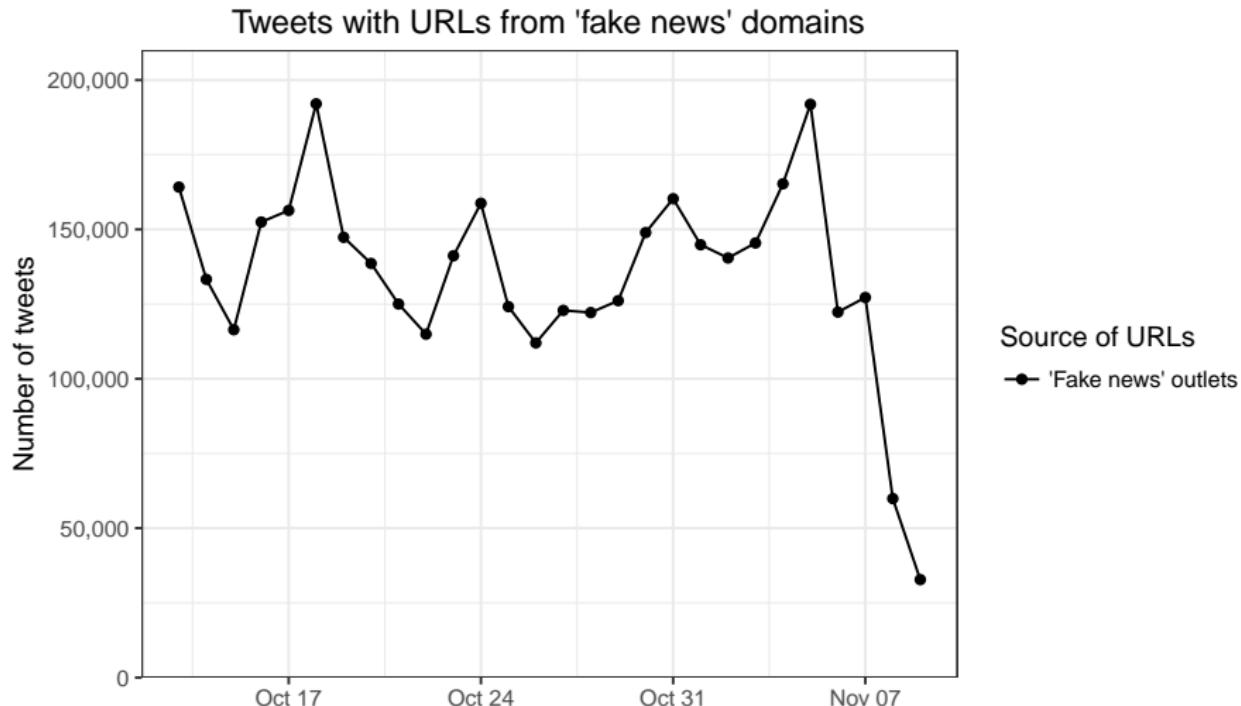
At least 16%

# Prevalence of misinformation



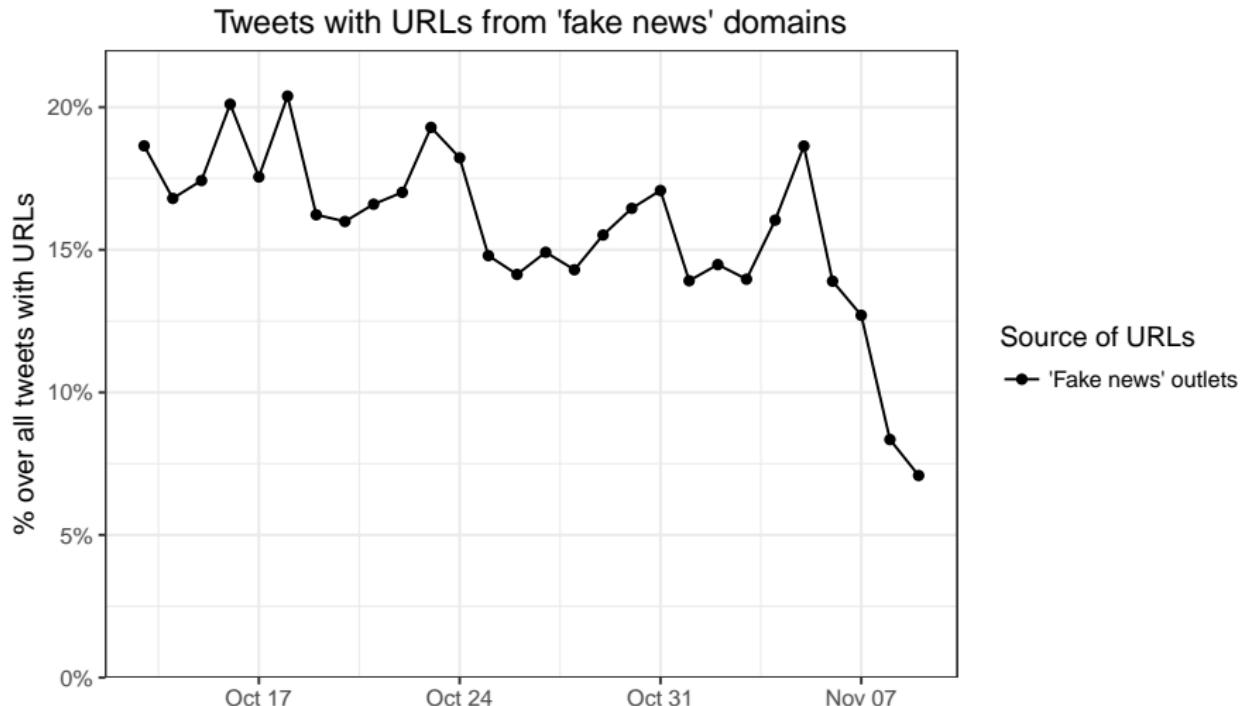
**Data:** 24.1 million tweets with URLs, keyword-based dataset.  
'Fake news' URLs from 146 domains, manually labeled.

# Prevalence of misinformation



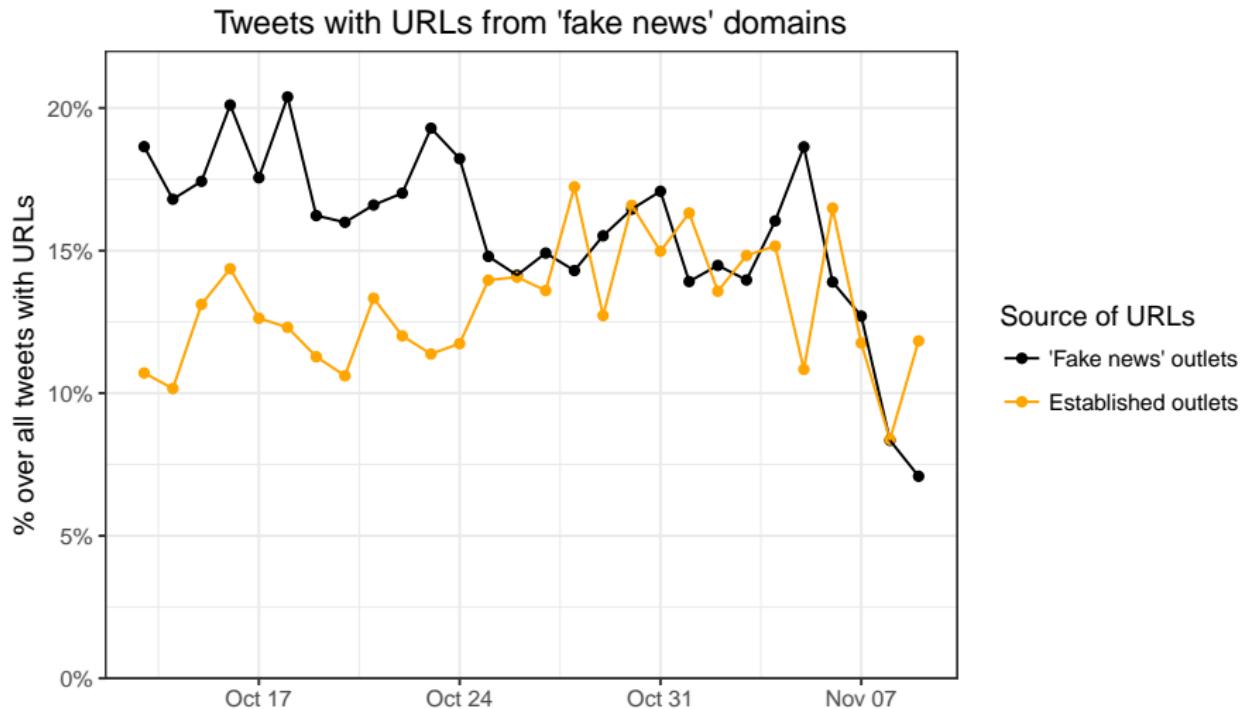
**Data:** 24.1 million tweets with URLs, keyword-based dataset.  
'Fake news' URLs from 146 domains, manually labeled.

# Prevalence of misinformation



**Data:** 24.1 million tweets with URLs, keyword-based dataset.  
'Fake news' URLs from 146 domains, manually labeled.

# Prevalence of misinformation



16 established media outlets: NYT, Fox, WSJ, WaPo, ABC News, Guardian, NY Post, NBC News, ... (13% of URL shares)

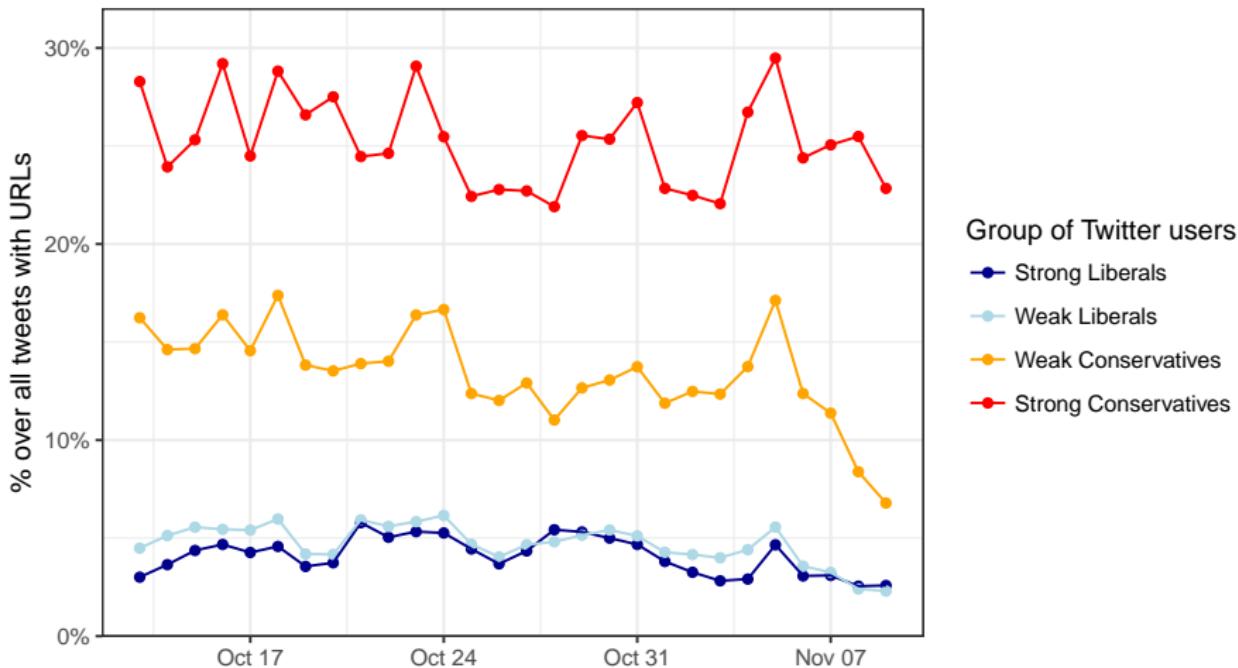
Who was sharing misinformation?

Who was sharing misinformation?

Right-leaning Twitter users,  
women, Republicans, 40+ years  
old, white, from red states

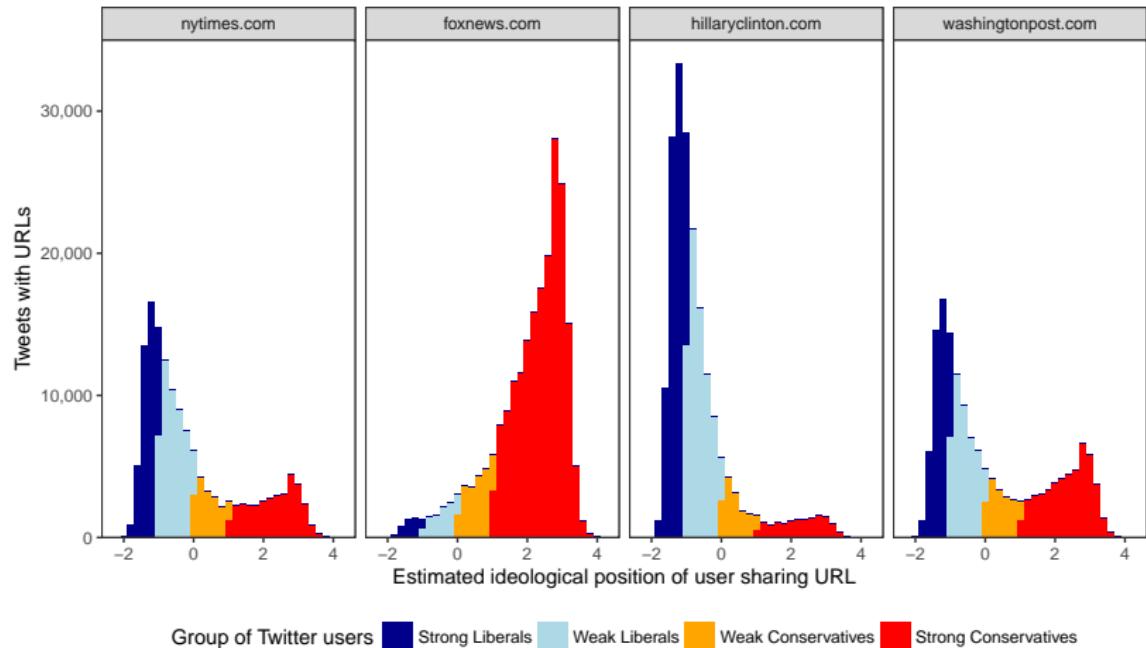
# Who was sharing misinformation?

Tweets with URLs from 'fake news' domains



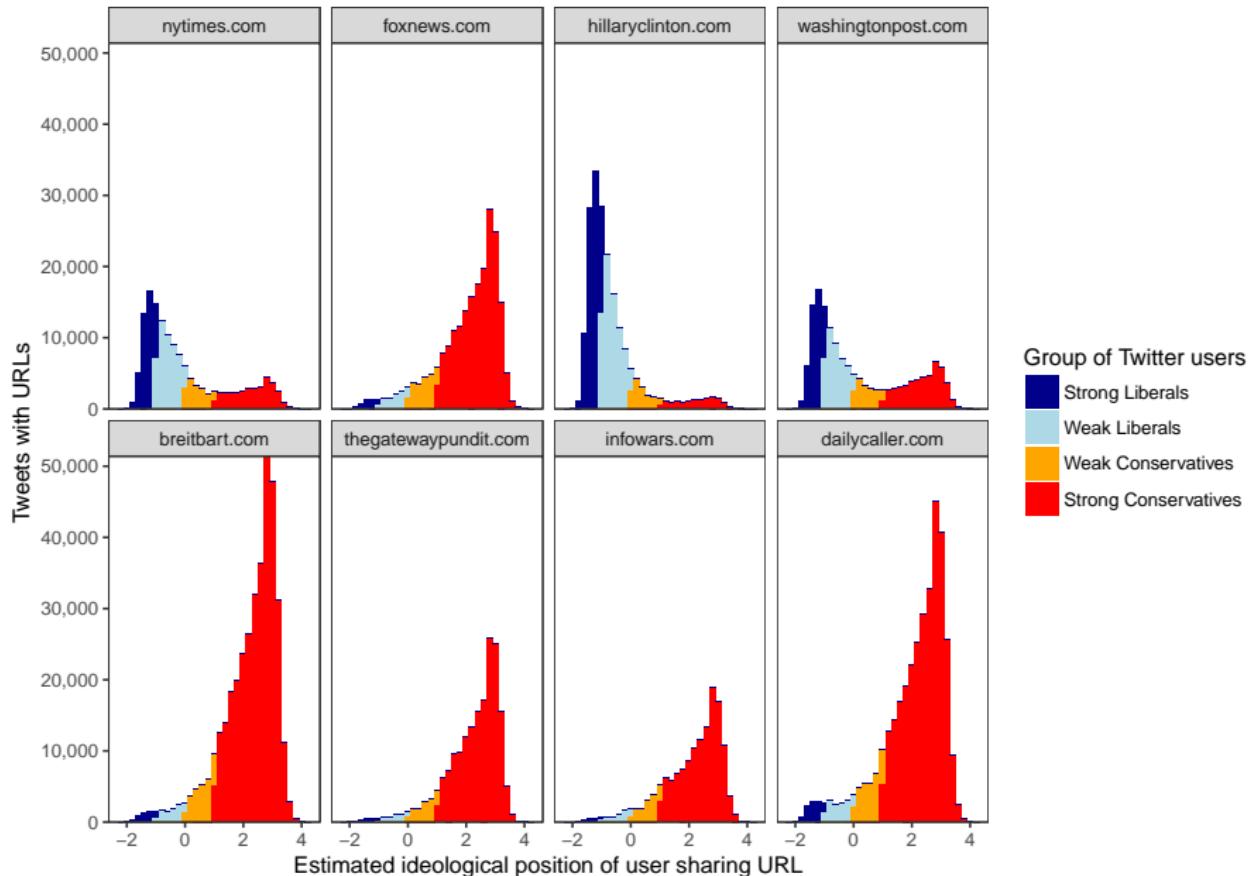
Ideology estimates based on Barberá et al (2015) for 12.6 million users following 3+ political accounts.

# Who was sharing misinformation?

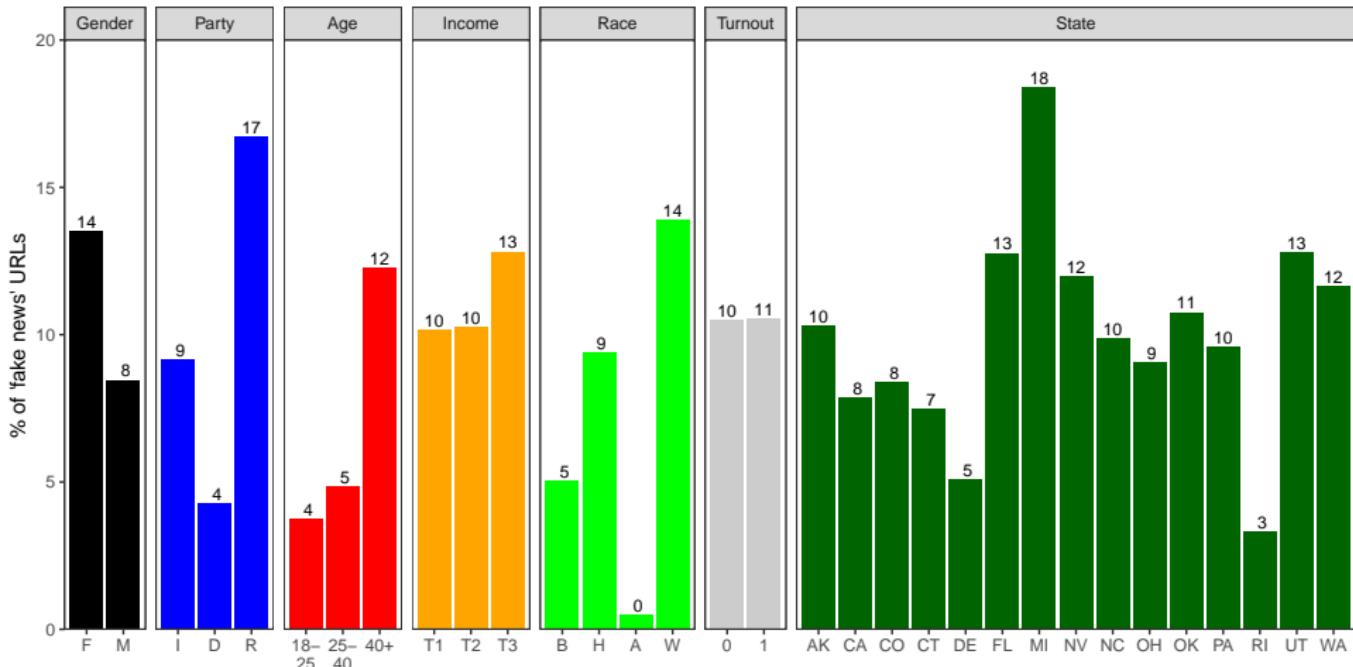


Ideology estimates based on Barberá et al (2015) for 12.6 million users following 3+ political accounts.

# Who was sharing misinformation?

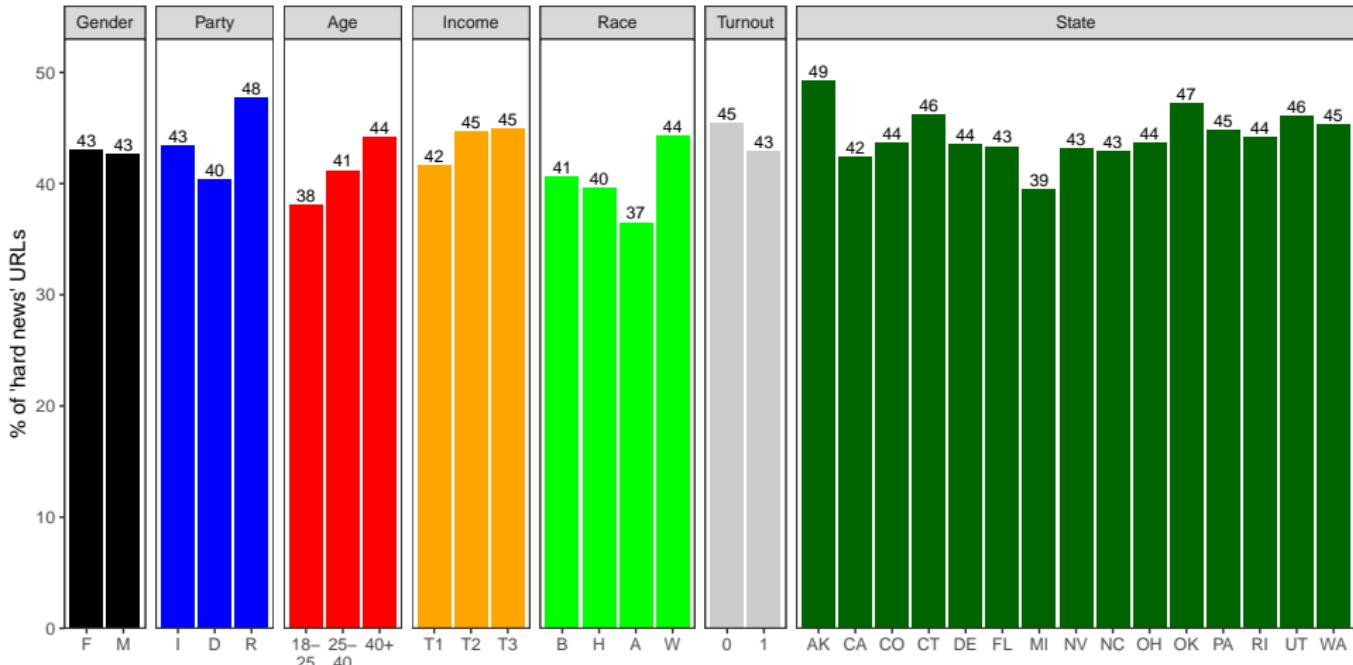


# Who was sharing misinformation?



**Data:** 205K Twitter accounts matched with voter files in 15 states.

# Who was sharing hard news URLs?



**Data:** 205K Twitter accounts matched with voter files in 15 states. List of 'hard news' URLs from Bakshy et al (2015)

What proportion of accounts spreading misinformation were bots?

What proportion of accounts spreading misinformation were bots?

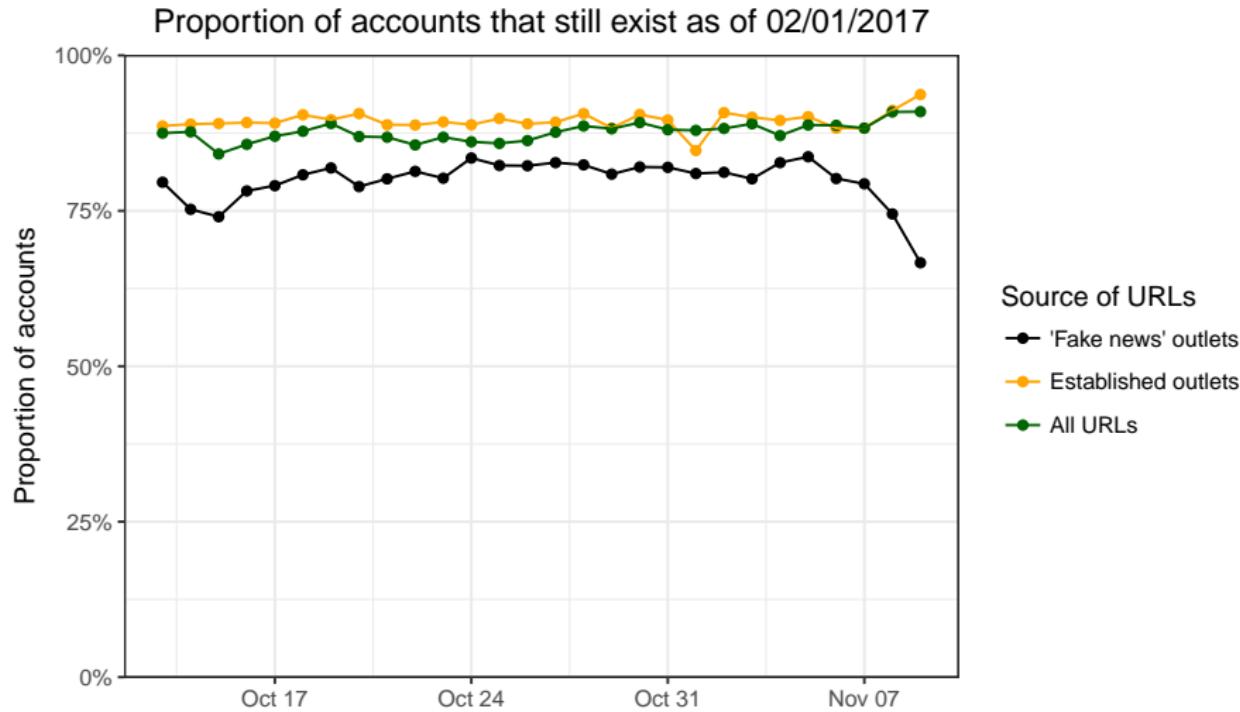
Our best guess: 15-30%

## What % of accounts were bots?



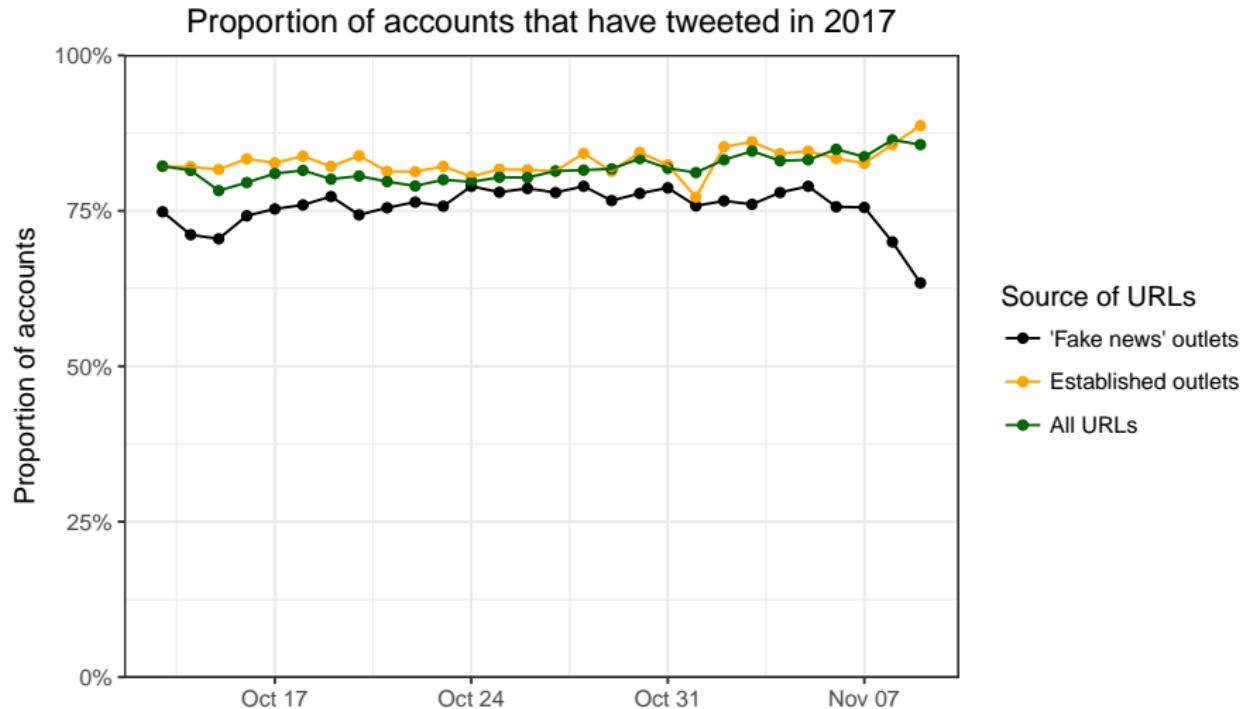
71% of accounts tweeting ‘fake news’ URLs still exist as of February 2017

# What % of accounts were bots?



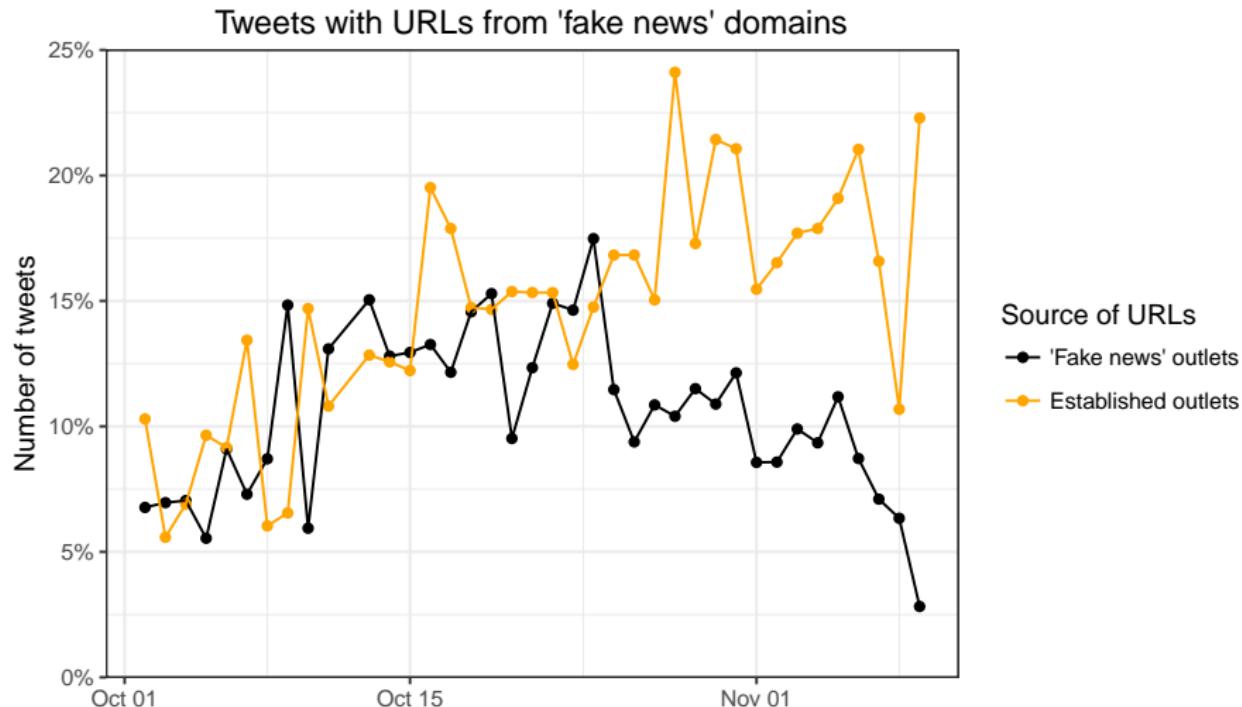
71% of accounts tweeting ‘fake news’ URLs still exist as of February 2017; in contrast to 85% of accounts tweeting URLs from 16 established outlets.

# What % of accounts were bots?



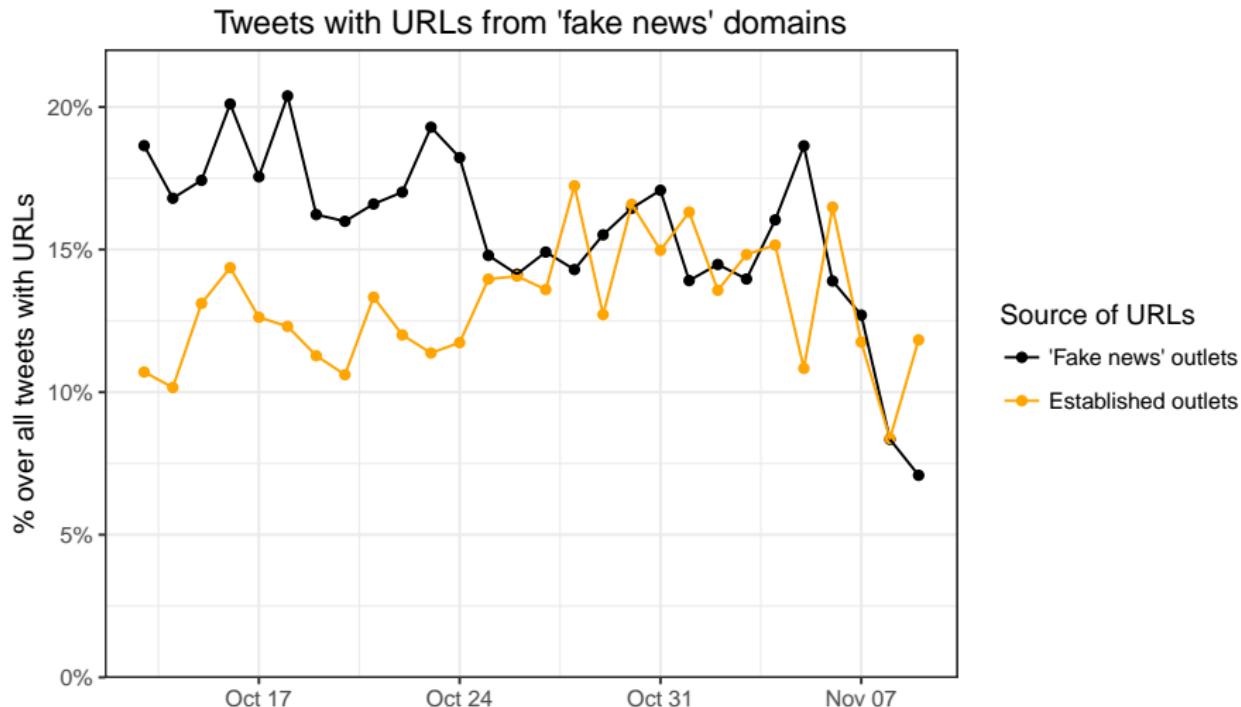
67% of accounts tweeting ‘fake news’ URLs have tweeted in 2017; in contrast to 79% of accounts tweeting URLs from 16 established outlets.

# What % of accounts were bots?



**Data:** 205K Twitter accounts matched with voter files in 15 states. (10% fake news URLs, 15% established outlets)

# What % of accounts were bots?



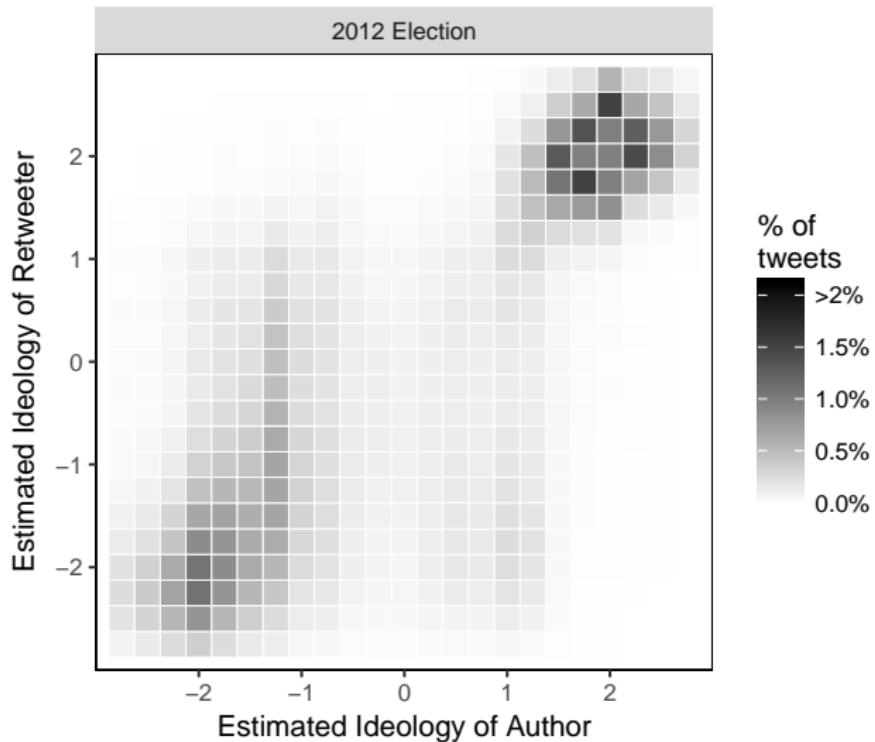
**Data:** 24.1m tweets with URLs, keyword-based dataset. (16% fake news URLs, 13% established outlets)

Did online polarization contribute to the spread  
of misinformation?

Did online polarization contribute to the spread  
of misinformation?

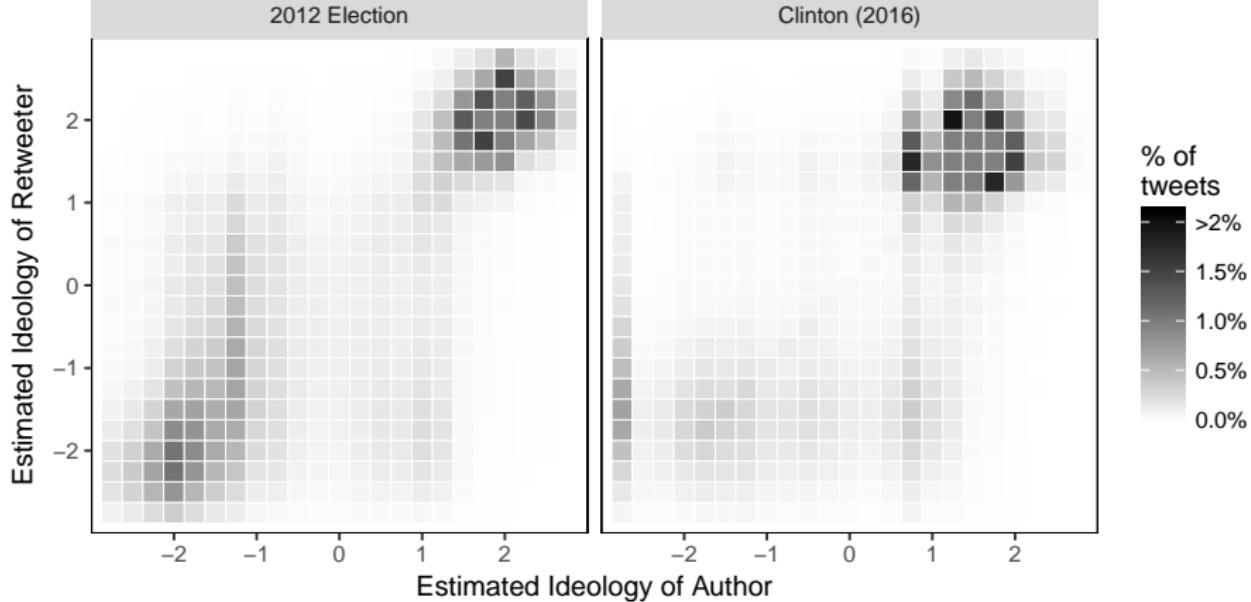
Probably not

# Polarization in information diffusion on Twitter

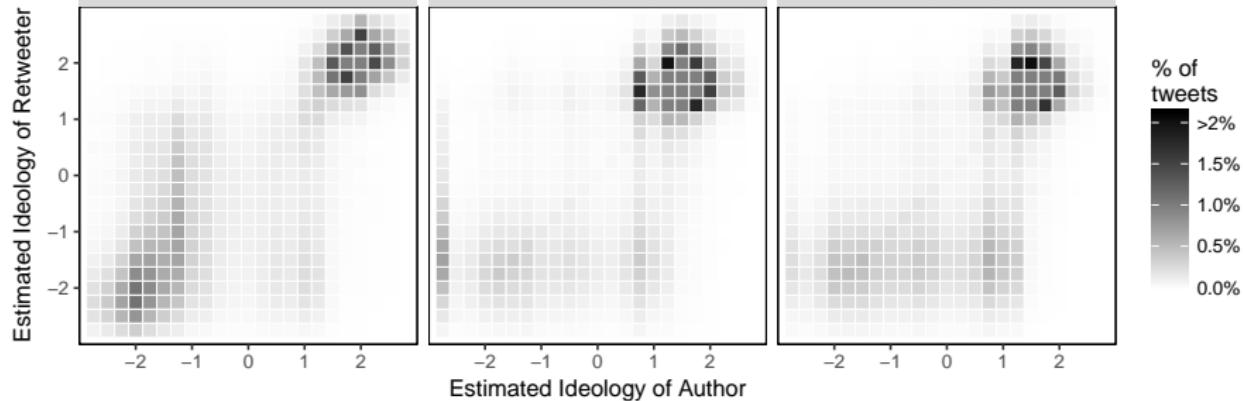


**Source:** Barberá et al (2015) *Psychological Science*

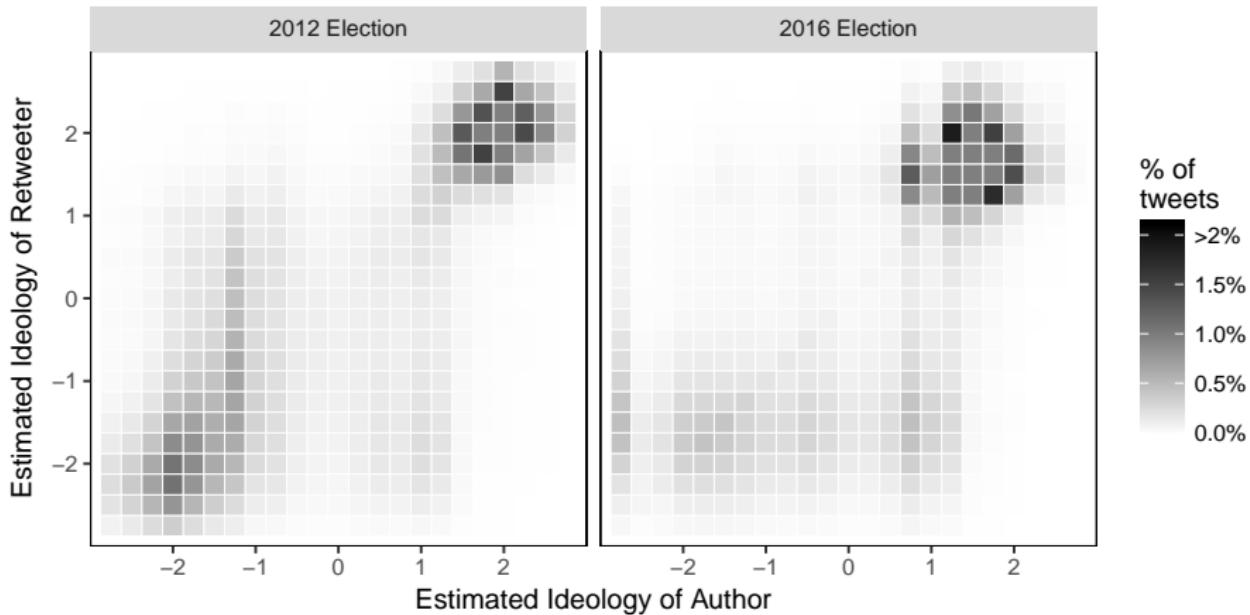
# Polarization in information diffusion on Twitter



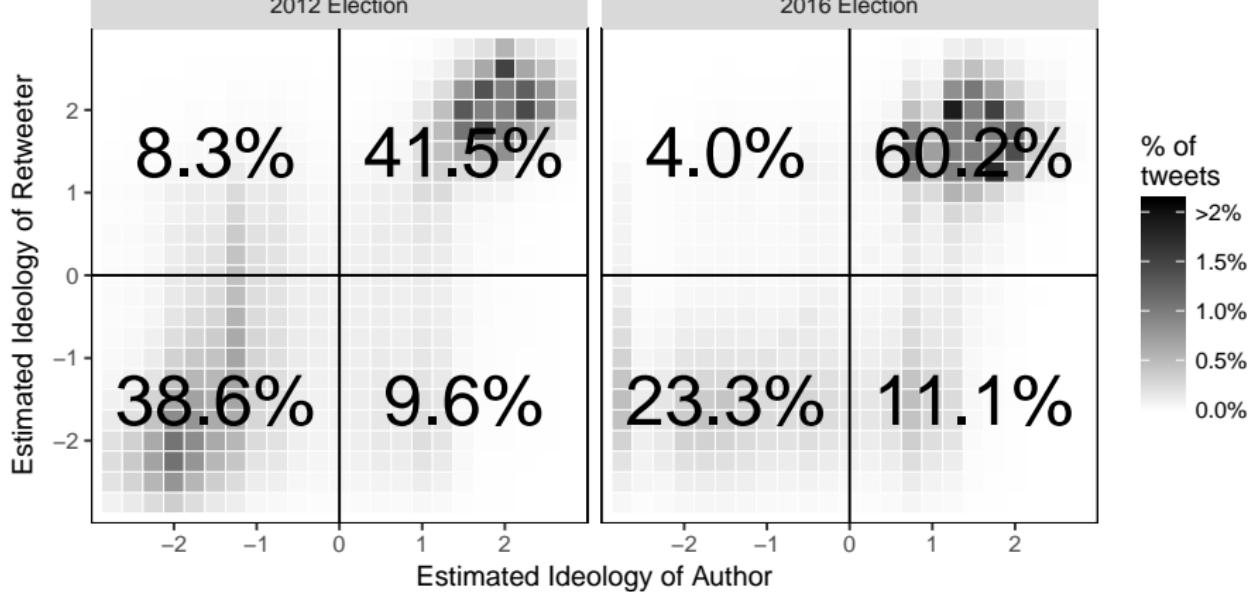
# Polarization in information diffusion on Twitter



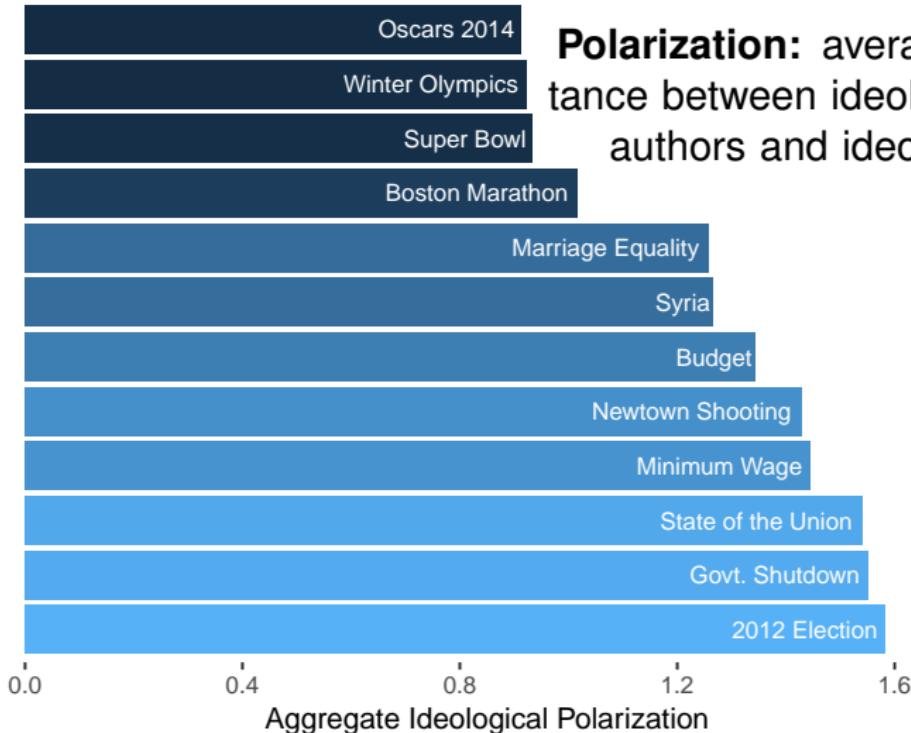
# Polarization in information diffusion on Twitter



# Polarization in information diffusion on Twitter

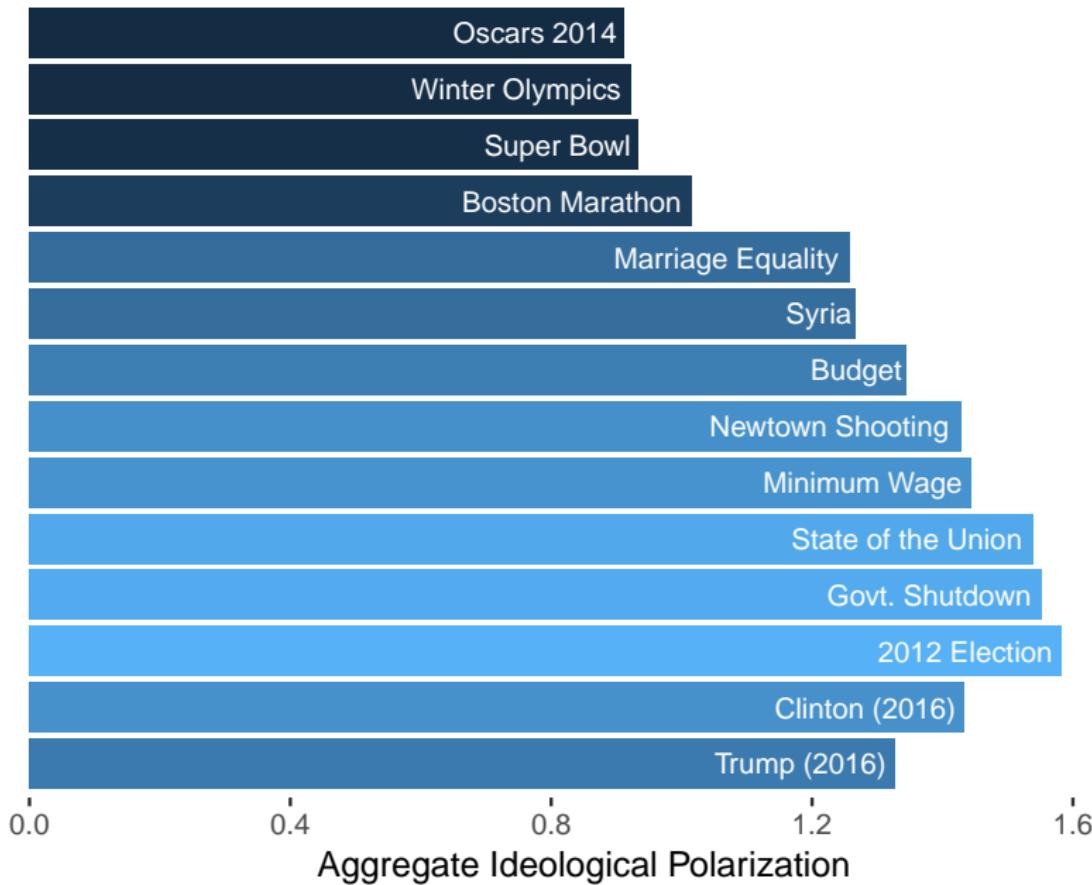


# Polarization in information diffusion on Twitter



**Polarization:** average absolute distance between ideology of retweeted authors and ideological center

**Source:** Barberá et al (2015) *Psychological Science*



## What (we think) we know

1. At least **16%** of all URLs shared on Twitter during the campaign contained some degree of misinformation

## What (we think) we know

1. At least **16%** of all URLs shared on Twitter during the campaign contained some degree of misinformation
2. **Right-leaning Twitter users, women, Republicans, 40+ years old, white, from red states** were more likely to share misinformation

## What (we think) we know

1. At least **16%** of all URLs shared on Twitter during the campaign contained some degree of misinformation
2. **Right-leaning Twitter users, women, Republicans, 40+ years old, white, from red states** were more likely to share misinformation
3. Probably **15-30%** of all tweets with 'fake news' URLs were sent by bots

## What (we think) we know

1. At least **16%** of all URLs shared on Twitter during the campaign contained some degree of misinformation
2. **Right-leaning Twitter users, women, Republicans, 40+ years old, white, from red states** were more likely to share misinformation
3. Probably **15-30%** of all tweets with 'fake news' URLs were sent by bots
4. High levels of exposure to diversity (in comparison to offline exposure) probably **contributed to the spread of misinformation**

**Thanks for your attention!**

Pablo Barberá

**Networked Democracy Lab    Chair of Systems Design**

USC

[netdem.org](http://netdem.org)

David García

ETH Zürich

[dgarcia.eu](http://dgarcia.eu)