

The abuse of social media: misinformation

Saptarshi Ghosh
and Mainack Mondal

CS 60017
Autumn 2019



Roadmap

Previous lecture

Misinformation in social media: Specifically we will focus on the spread of misinformation in whatsapp today

This lecture:

How different are **fake news content** from **real news**

What happened? The Spread of Fake News Publisher Content During the 2016 U.S. Presidential Election

(Ceren Budak, WWW'19)

Findings

- i. traditional news producers outperformed fake news producers in aggregate
- ii. the prevalence of content produced by fake news publishers increased over the course of the campaign
- iii. changes in such prevalence were closely following changes in net Clinton favorability

Dataset

Fake news publishers

- Combined list from 5 well known sources

- If two or more sources identify as fake, then fake

- Any news from fake news publishers is fake

Real news publishers

- Websites whom Alexa categorized as “news”

- Removed the fake news publishers

Each day 5000 random tweets with terms “Clinton” and “Trump”

- Removed tweet without external urls

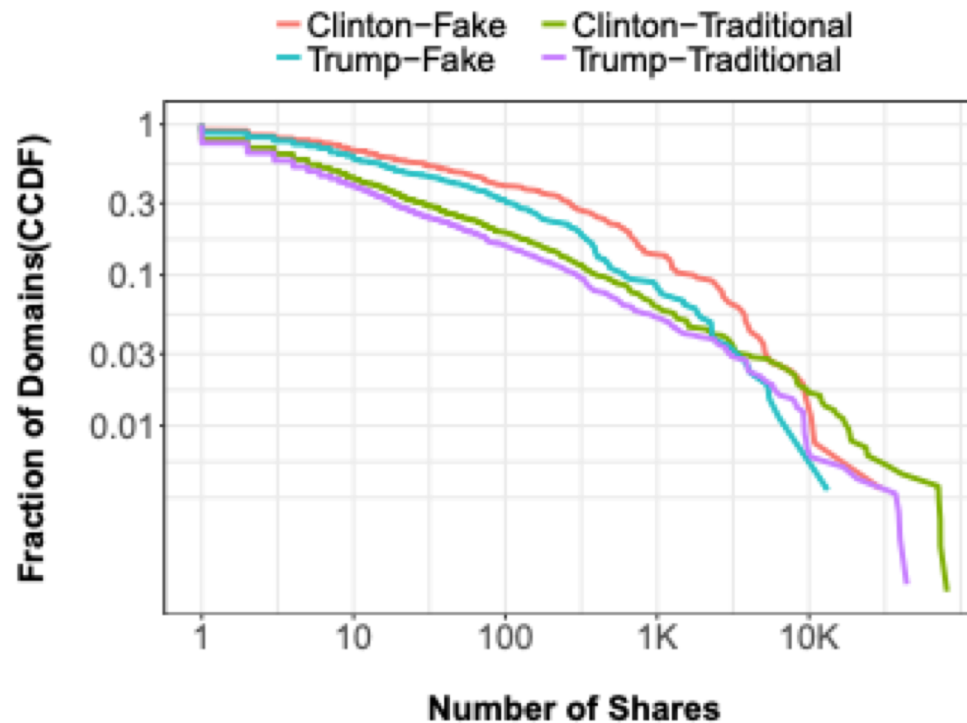
- Twitter API to create a tweet object

Gall up survey during election months

- What did you heard about Trump | Clinton that affected you?

How popular are fake and real news?

Popularity of Fake and traditional news domains
#shares of the same url

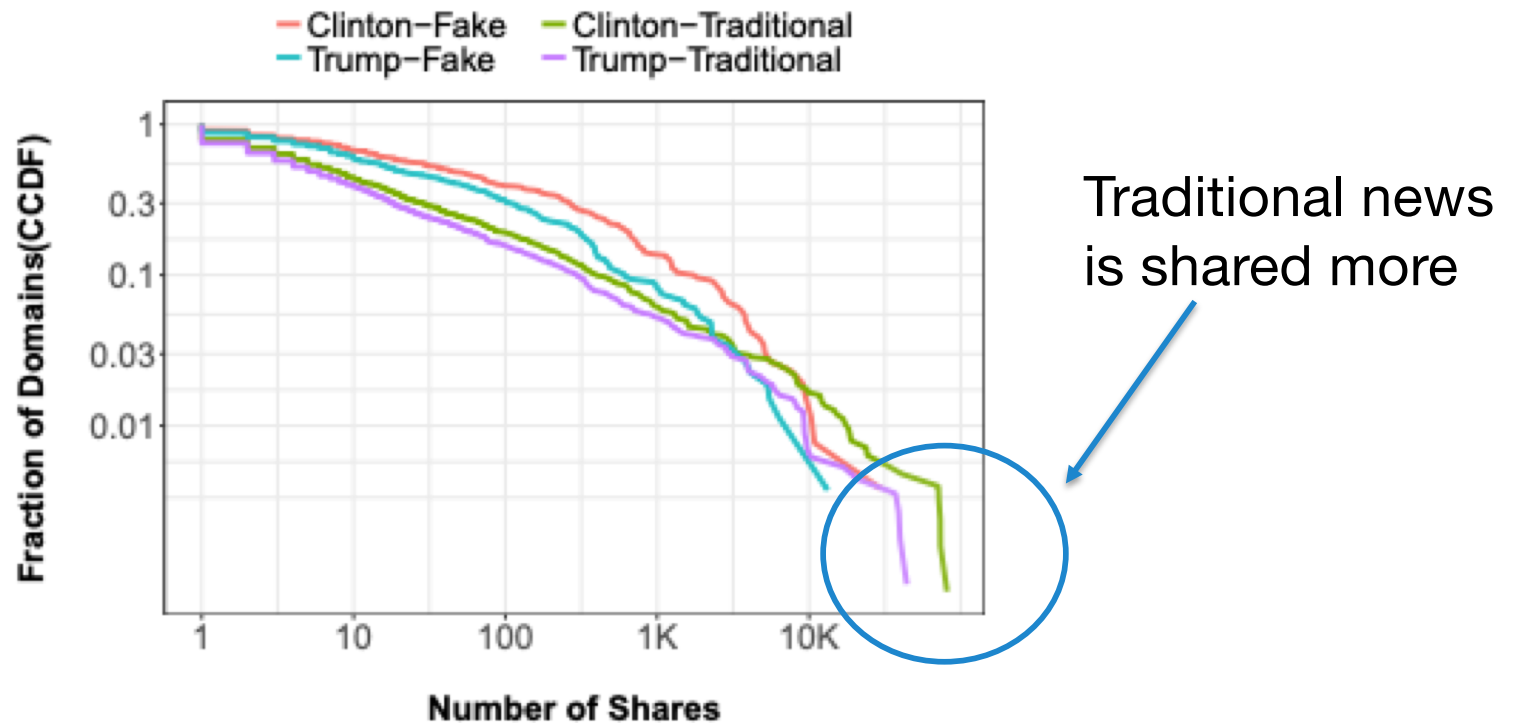


Very similar distributions

Fake and traditional news shared around similar #times

How popular are fake and real news?

Popularity of Fake and traditional news domains
#shares of the same url



Very similar distributions

Fake and traditional news shared around similar #times

Findings

- i. traditional news producers outperformed fake news producers in aggregate
- ii. the prevalence of content produced by fake news publishers increased over the course of the campaign
- iii. changes in such prevalence were closely following changes in net Clinton favorability

Do fake news affect public opinion?

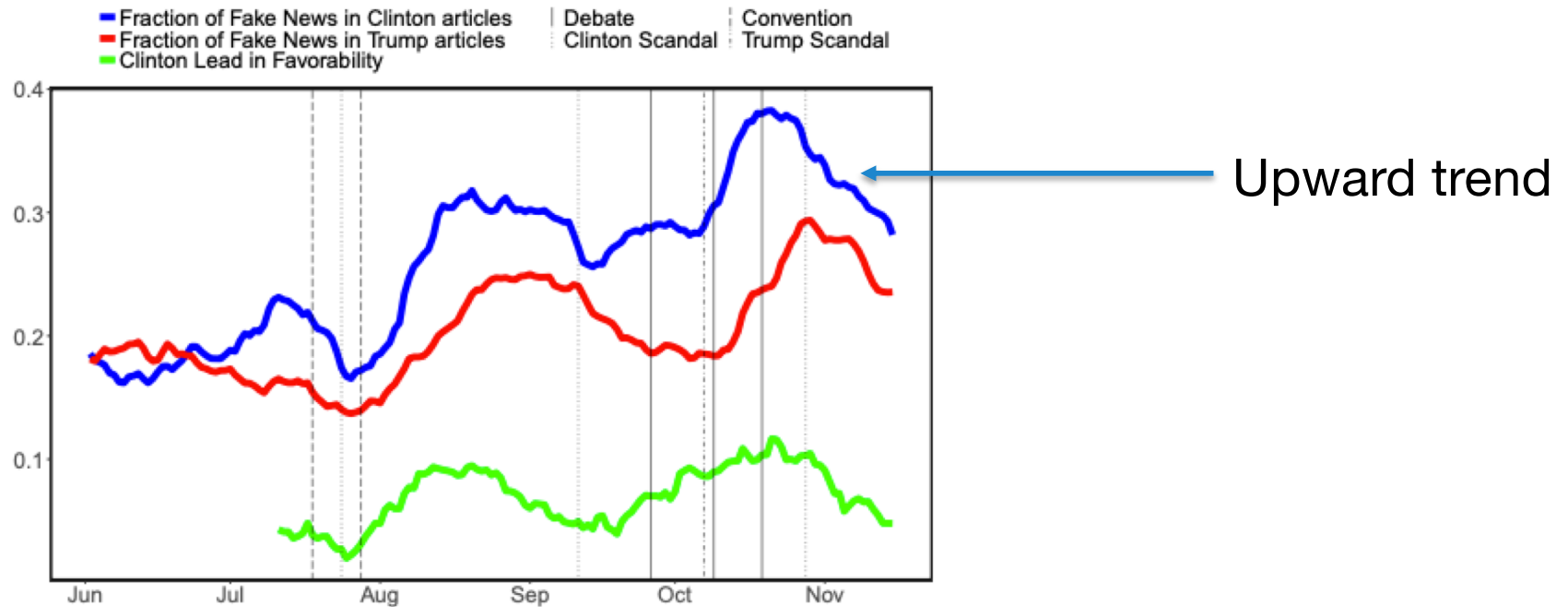
$$\text{Metric for fakeness} = \frac{\text{\#fake tweets about Clinton or Trump each day}}{\text{Total \#fake news that day}}$$

$$\text{Metric for favorability} = \frac{\text{\#Survey takes favoring Clinton or Trump}}{\text{Total \#survey takers}}$$

Do fake news affect public opinion?

Metric for fakeness = $\frac{\#fake\ tweets\ about\ Clinton\ or\ Trump\ each\ day}{Total\ \#fake\ news\ that\ day}$

Metric for favorability = $\frac{\#Survey\ takes\ favoring\ Clinton\ or\ Trump}{Total\ \#survey\ takers}$



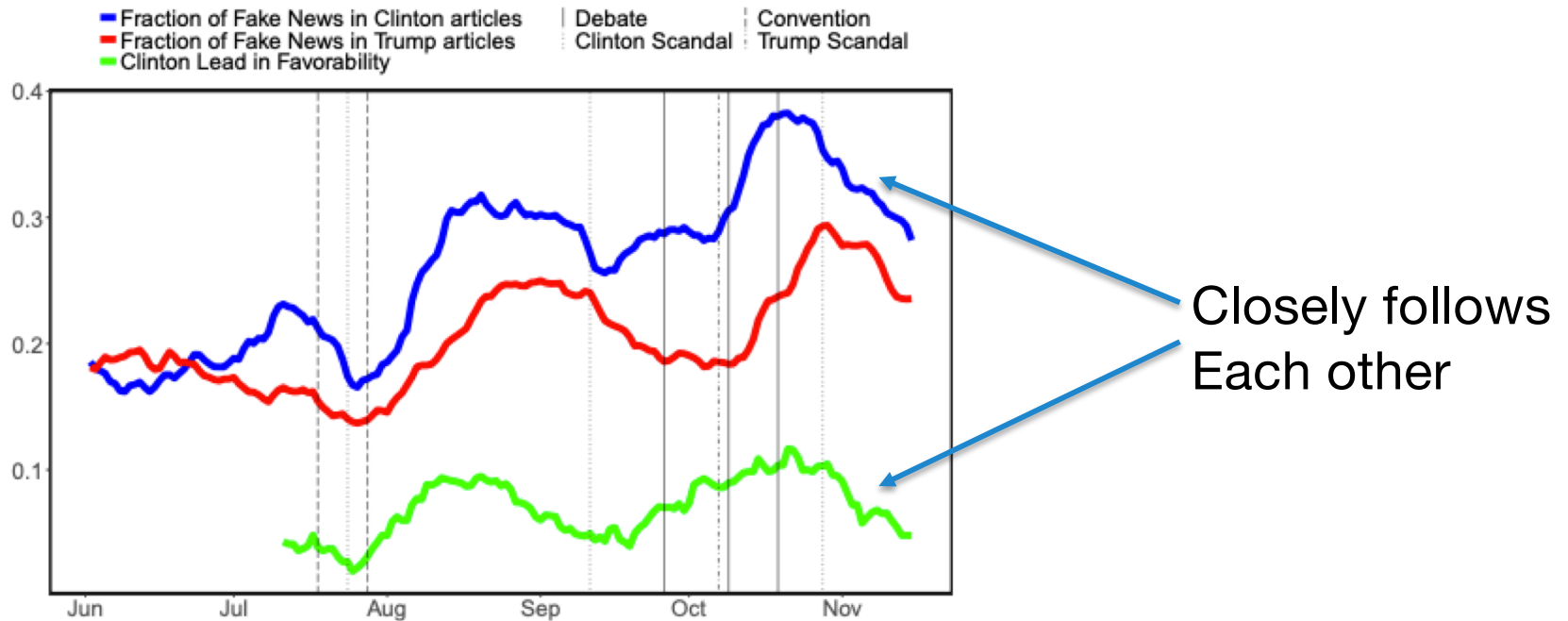
Findings

- i. traditional news producers outperformed fake news producers in aggregate
- ii. the prevalence of content produced by fake news publishers increased over the course of the campaign
- iii. changes in such prevalence were closely following changes in net Clinton favorability

Do fake news affect public opinion?

Metric for fakeness = $\frac{\#fake\ tweets\ about\ Clinton\ or\ Trump\ each\ day}{Total\ \#fake\ news\ that\ day}$

Metric for favorability = $\frac{\#Survey\ takes\ favoring\ Clinton\ or\ Trump}{Total\ \#survey\ takers}$



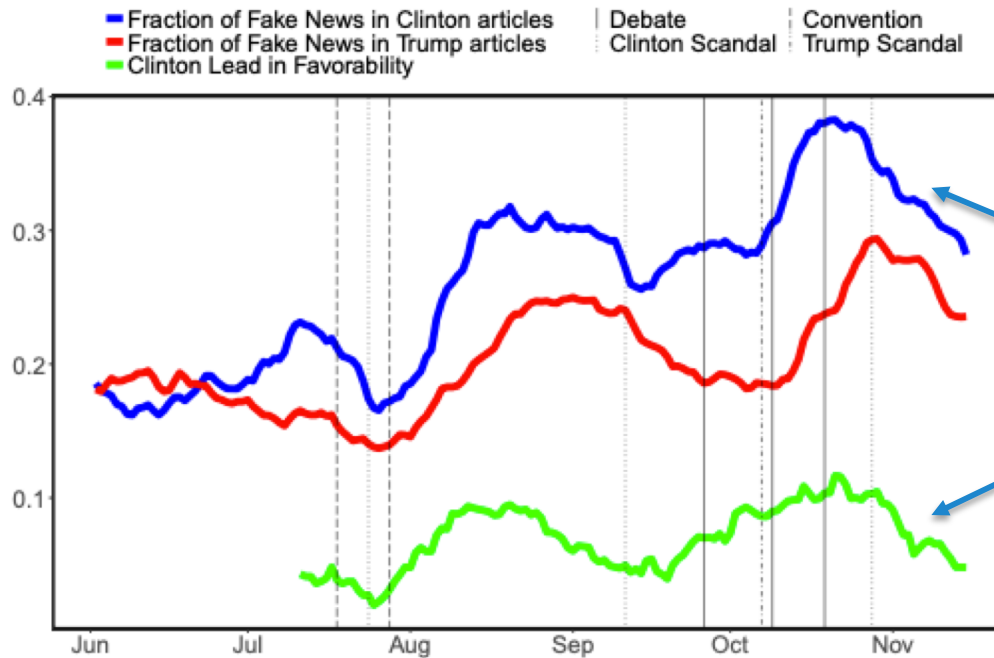
Findings

- i. traditional news producers outperformed fake news producers in aggregate
- ii. the prevalence of content produced by fake news publishers increased over the course of the campaign
- iii. changes in such prevalence were closely following changes in net Clinton favorability

Do fake news affect public opinion?

Metric for fakeness = $\frac{\#fake\ tweets\ about\ Clinton\ or\ Trump\ each\ day}{Total\ \#fake\ news\ that\ day}$

Metric for favorability = $\frac{\#Survey\ takes\ favoring\ Clinton\ or\ Trump}{Total\ \#survey\ takers}$



But trend in favorability is around 4 days earlier than fake news

Findings

- i. traditional news producers outperformed fake news producers in aggregate
- ii. the prevalence of content produced by fake news publishers increased over the course of the campaign
- iii. changes in such prevalence were closely following changes in net Clinton favorability
- iv. Fake news follows offline user opinion

Few words about end sem

Read the papers in course webpage (dataset, methodology, findings)

There will be questions on those

25% of the exam will be from topics covered in mid sem

Understand and remember privacy definitions/
k-anonymity/differential privacy

How to design metrics in a given social context

How to interpret graphs/data to find answers