

Mining Social Media



Lab 9

Mining Social Media



Lab 9

Joe Rogan

- Host of “most-listened-to podcast globally” acc. to Spotify

Ben Shapiro

- Lawyer, columnist, author, and conservative political pundit

Ratatouille

- 2007 Pixar film about a rat that cooks

This conversation never happened



SOCIAL SCIENCE

The spread of true and false news online

Soroush Vosoughi,¹ Deb Roy,¹ Sinan Aral^{2*}

We investigated the differential diffusion of all of the verified true and false news stories distributed on Twitter from 2006 to 2017. The data comprise ~126,000 stories tweeted by ~3 million people more than 4.5 million times. We classified news as true or false using information from six independent fact-checking organizations that exhibited 95 to 98% agreement on the classifications. Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information. We found that false news was more novel than true news, which suggests that people were more likely to share novel information. Whereas false stories inspired fear, disgust, and surprise in replies, true stories inspired anticipation, sadness, joy, and trust. Contrary to conventional wisdom, robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it.

support their position whereas sources that were not labeled reliable or had no connection to the institution presented, remained in academic classification explicitly avoided this paper and in verifiable terms “the terms fake news imply a willful disinformation make any claims about the source of the information focus our attention have been verified.”

We also purpose of the term news constitutes news as a source of the asserted asserted claim may defend this decision section on “reliability” define news as an institution in it and a man-

Recent changes to Twitter



2006 – Twitter founded by Jack Dorsey, Noah Glass, Evan Williams, & Biz Stone with Dorsey as CEO

2014 – Twitter daily users peaks at 661,000,000

Feb-Oct 2020 – Multiple groups attempt to “oust” Dorsey

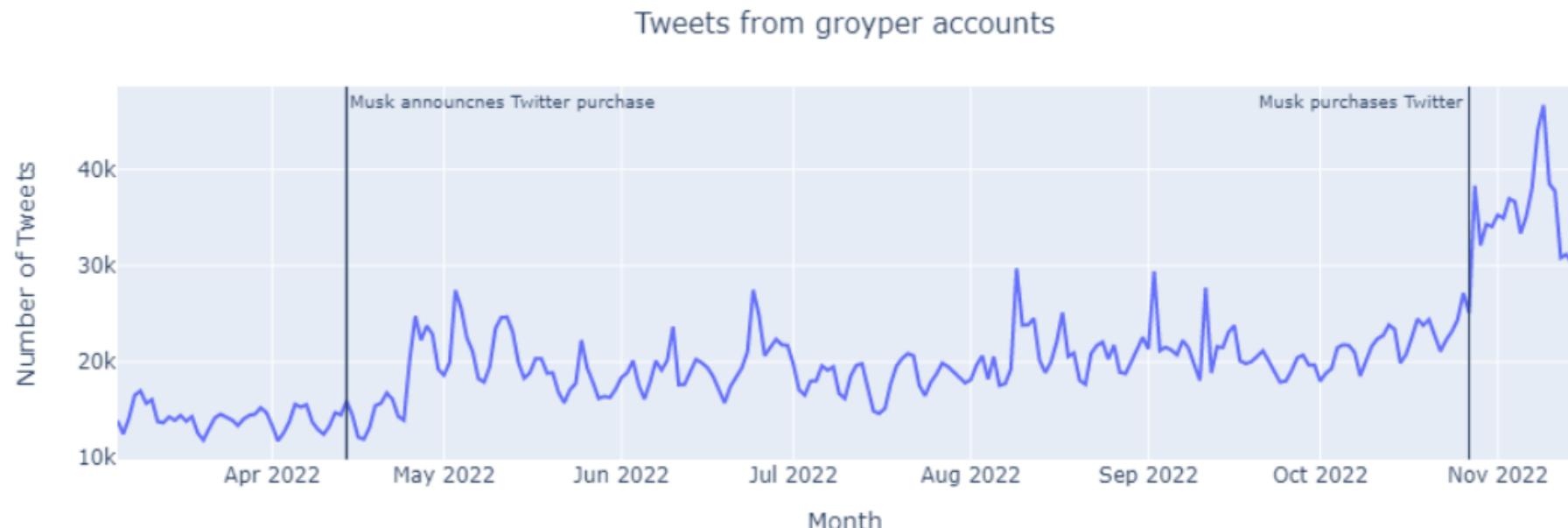
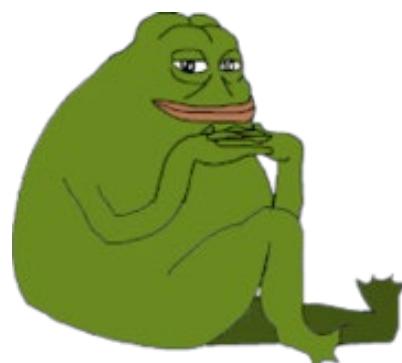
May 2022 – Dorsey resigns and Musk offers to buy

Oct 2022 – Musk takes over; lays off 5,500 employees

Oct 2022 – Dorsey starts “[Bluesky](#)” as a Twitter alternative

Apr 2023 – Changes to Twitter API, the point is to get rid of bots and promote free speech

Groypers, sometimes called the **Groyper Army**, are a group of [white nationalist](#) and [far-right](#) activists, provocateurs and [internet trolls](#) who are notable for their attempts to introduce far-right politics into mainstream [conservatism in the United States](#), their participation in the [2021 United States Capitol attack](#) and the protests leading up to it, and their [extremist](#) views. They are known for targeting other conservative groups and individuals whose agendas they view as too moderate and insufficiently [nationalist](#).^{[3][4]} The Groyper movement has been described as white nationalist, [homophobic](#), [nativist](#), [fascist](#), [sexist](#), [antisemitic](#), and an attempt to rebrand the declining [alt-right](#) movement.^{[2][5][6][7]}



Agenda

- Scraping Reddit with R
- Turning words into numbers

“If you can quantify it, you can do science with it”

- Dr. Sarah Woolley (Columbia Psychology)

Unsupervised learning

Supervised

Ground truth available

label = 1



label = 9



label = 1



label = 4



Unsupervised

Ground truth unavailable



Unsupervised learning

- Clustering
- Abstraction
- Generative models

Machine learning (ML) algorithms are increasingly integrated into every aspect of our lives. Whether it is recommending movies, diagnosing diseases, or defending against cyberattacks, ML has emerged as one of the dominant technologies of our age, powering countless advances in diverse fields such as robotics, particle physics, and personal finance. Practitioners typically distinguish between three main branches of ML: supervised, unsupervised, and reinforcement learning.¹ The first pertains to prediction problems, e.g., labeling images. The third pertains to policy optimization, e.g., learning boardgames. The subject of this article is the second branch, for which pithy descriptions are not nearly so straightforward.

[Watson \(2023\) On the Philosophy of Unsupervised Learning](#)



Search Reddit

Get app

Log In

...



Hot Everywhere

r/MadeMeSmile • 6 hr. ago

Join

A girl tapped me on the shoulder and gave me her number at the gym today :)



POPULAR COMMUNITIES

r/explainlikeimfive
22,783,027 members

r/IAmA
22,572,889 members

r/classicwow
596,500 members

r/Instagram
856,627 members

r/NintendoSwitch
5,911,189 members

See more

subreddit

 r/news • 7 hr. ago

Join

...

Video shows California police fatally shooting teenager who was reported kidnapped



thread

[https://www.theguardian.com/us-news/2024/apr/01/california-police-video-shooting-15-...](https://www.theguardian.com/us-news/2024/apr/01/california-police-video-shooting-15-year-old-boy)

 18K 

 1.3K

 Share

up/down-votes

 r/AskReddit • 10 hr. ago

Join

...

What's the worst drug ever?

 5K 

 6.4K

 Share

comments

 r/facepalm • 11 hr. ago

Join

...

And this is how a new person in the neighborhood announces themselves, pretty aggressive. I'm not taking the tray of muffins over.

