

# Social Media Research Methods

Dr. Patricia Rossini

Department of Communications & Media



# Me, in a nutshell

Tenure-track Derby Fellow [@](#) University of Liverpool, UK

Postdoctoral researcher [@](#) *School of Information Studies, Syracuse U, USA (2017-2019)*

Ph.D. Communications [@](#) *Federal University of Minas Gerais, Brazil (2017)*

## Research interests:

Political communication + technologies

Informal political talk online

Incivility and intolerance

Computational social sciences

Political campaigns online

Deliberation

Misinformation

# Me, in a nutshell

## *Funded projects*

- “Visual Misinformation in Global Perspective: Platforms, Devices, and Users.” Facebook.
  - Principal Investigator: Cristian Vaccari. Co-Principal Investigators: Patricia Rossini, Michael Chan, Shira Dvir-Gvirsman, Iginio Gagliardone, Raquel Recuero, Nicole Stremlau.
- "WhatsApp Misinformation Award: WhatsApp as a source of political participation and (mis)information in Brazil". WhatsApp.
  - Principal Investigator: Patricia Rossini. Co- Principal Investigators: V. Veiga, E. Baptista J. Stromer-Galley.
- "Devising Metrics for Assessing Echo Chambers, Incivility, and Intolerance". Twitter.
  - Principal Investigator: Rebekah Tromble. Co- Principal Investigators: Patricia Rossini, N. Tintarev, D. Hovy, M. Meffert, J. Stromer-Galley.

# Workshop Schedule

## Day 1:

- Social Media Research: Challenges & Opportunities
- Social Media Data & APIs
- Ethics of Social Media Research
- Data collection on SM

## *Hands on:*

- Collecting data using the Twitter API
- Understanding Twitter data
- Group work: define topic & RQs

# Workshop Schedule

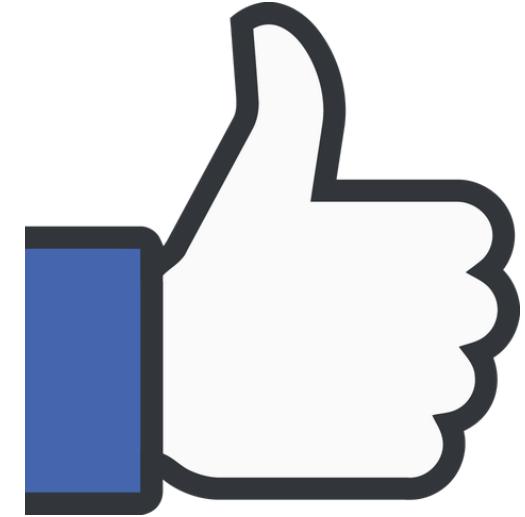
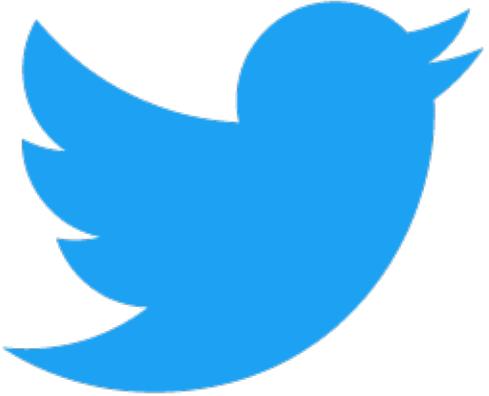
Day 2:

- Data Analysis Techniques: an overview

*Hands on:*

- Time series, textual, and sentiment analysis
- Group work: collect, analyze and present your work

*(if time allows: Facepager & collecting data from Facebook/YouTube)*



Why care so much about social media?

---



# THE AVERAGE BRIT CHECKS THEIR PHONE 10,000 TIMES A YEAR, STUDY FINDS

We live in a digital world, but has our dependency on our phones gone overboard?



RIVELY  
LTH

Click to follow  
Indy/Life

Very  
approa

# The social media (r)evolution

*% of U.S. adults who use at least one social media site*

80

60

40

20

0

Feb 7, 2019  
U.S. adults: 72%

2006

2008

2010

2012

2014

2016

2018

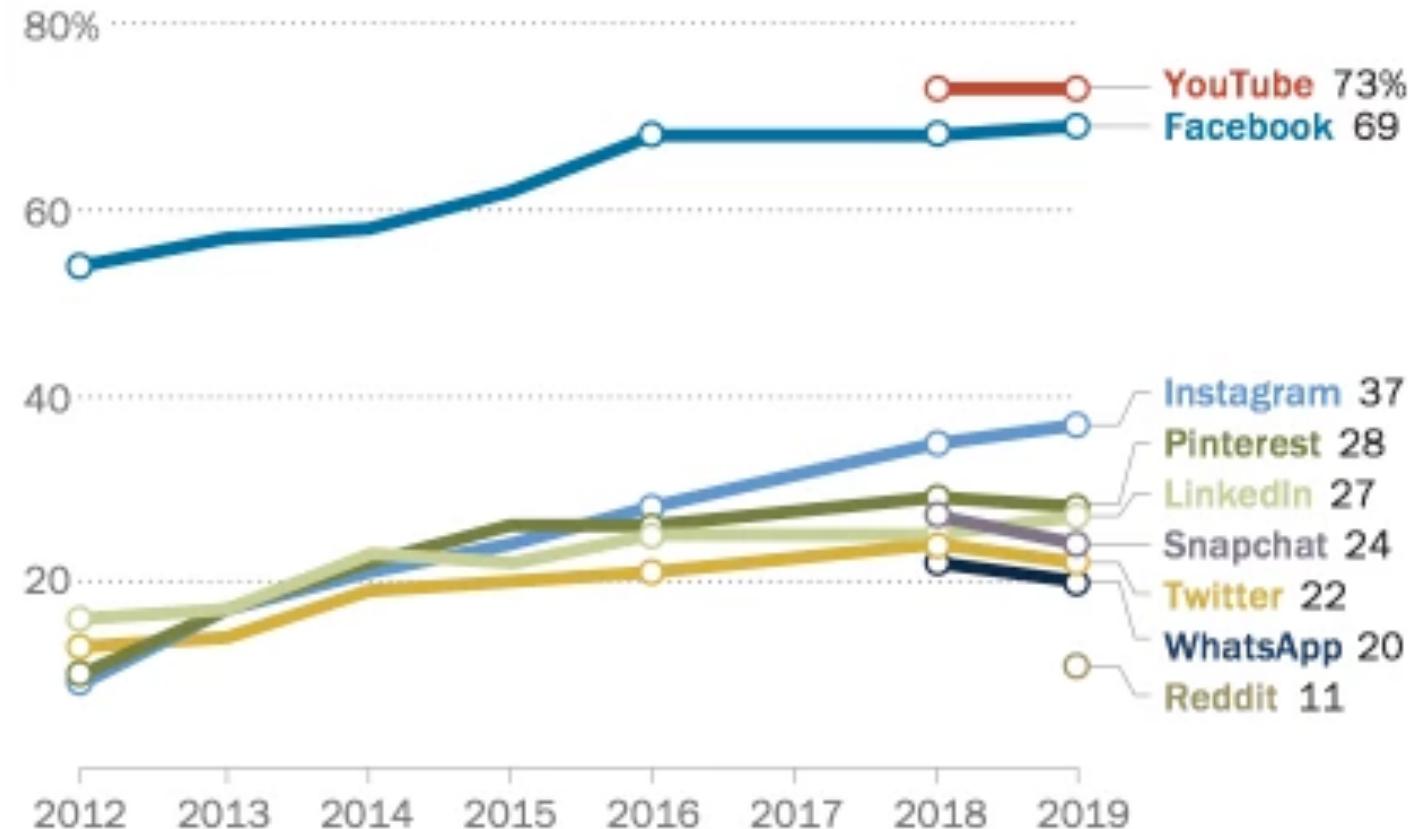
Source: Surveys conducted 2005-2019.

PEW RESEARCH CENTER

# The social media (r)evolution

## Facebook, YouTube continue to be the most widely used online platforms among U.S. adults

% of U.S. adults who say they ever use the following online platforms or messaging apps online or on their cellphone

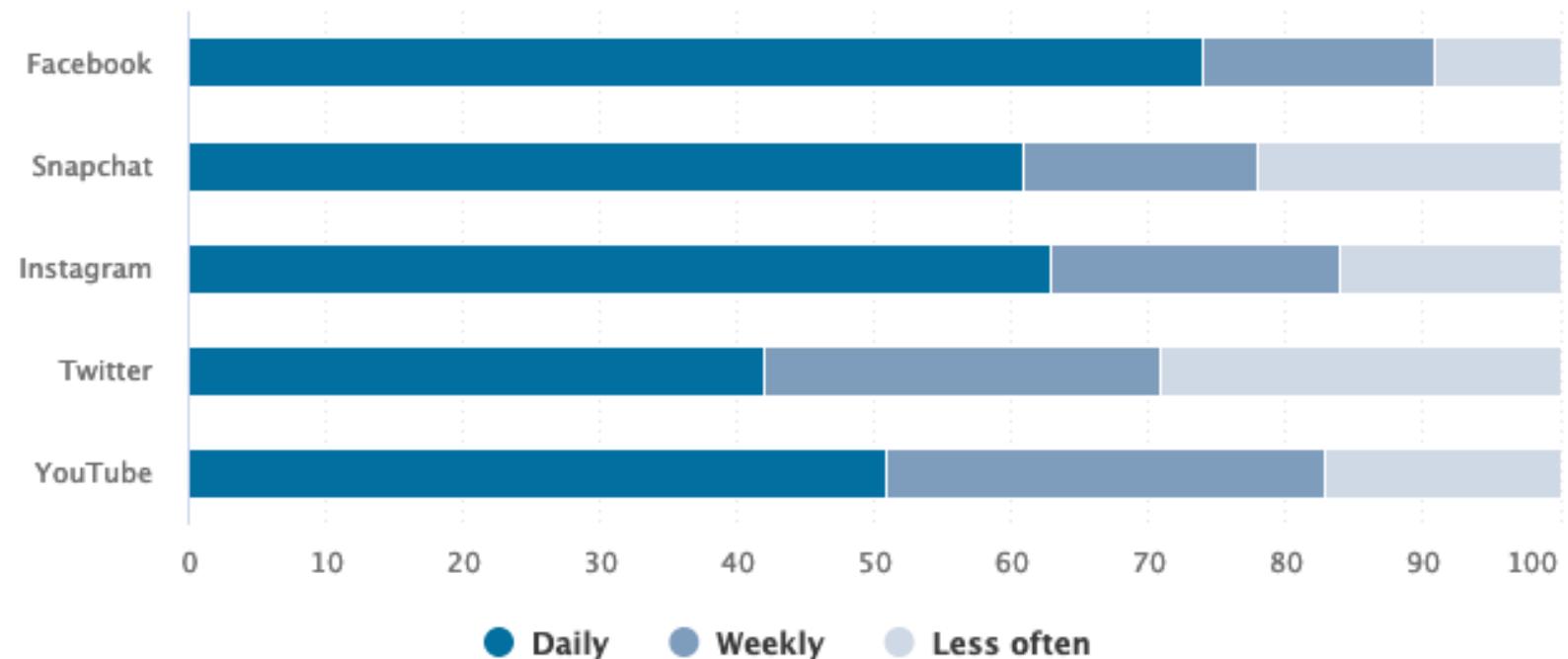


Note: Pre-2018 telephone poll data is not available for YouTube, Snapchat and WhatsApp. Comparable trend data is not available for Reddit.

Source: Survey conducted Jan. 8-Feb. 7, 2019.

# The social media (r)evolution

*Among the users of each social media site, the % who use that site with the following frequencies*

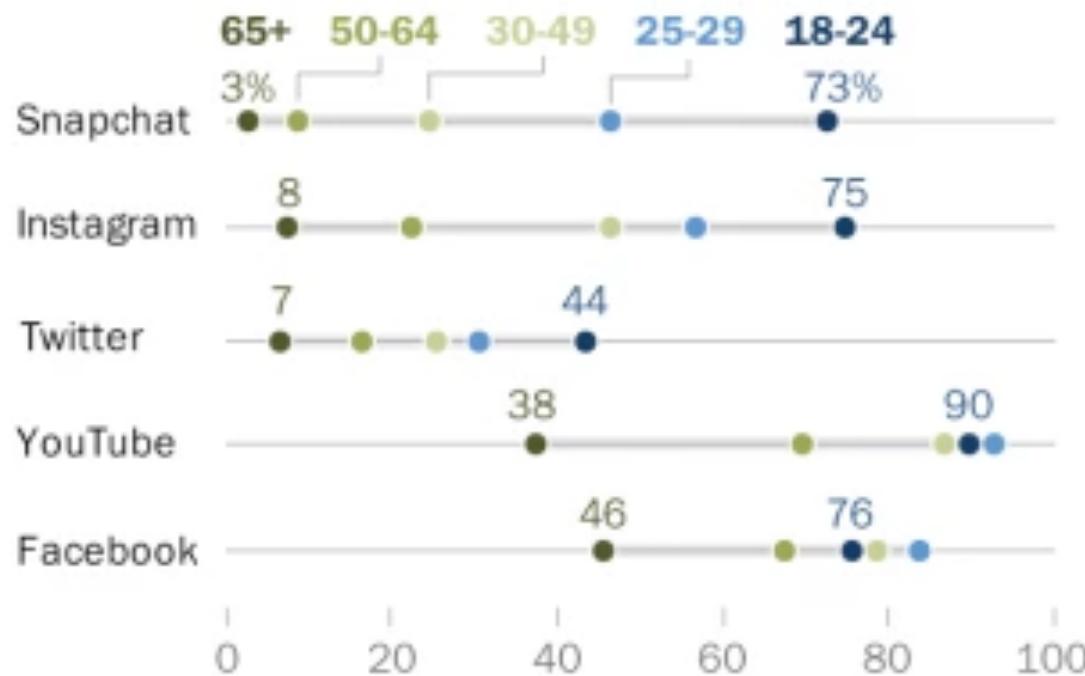


Note: Numbers may not add to 100 due to rounding.

Source: Survey conducted Jan. 8 to Feb. 7, 2019.

## Snapchat and Instagram are especially popular among 18- to 24-year-olds

% of U.S. adults in each age group who say they ever use ...



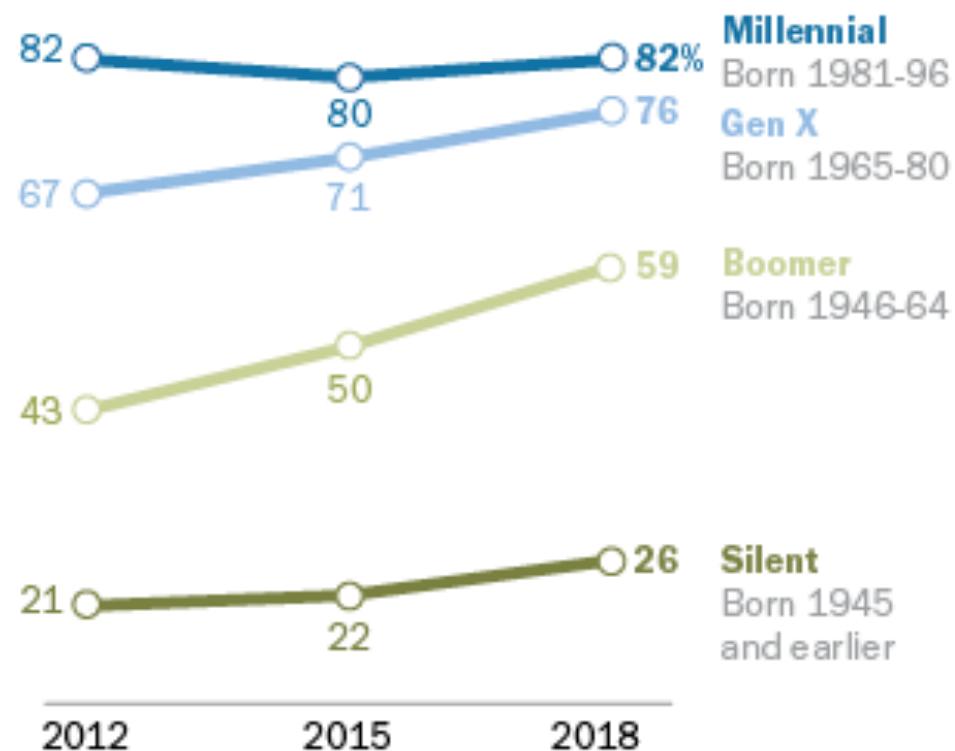
Note: Respondents who did not give an answer are not shown.

Source: Survey conducted Jan. 8-Feb. 7, 2019.

PEW RESEARCH CENTER

## Similar shares of Gen Xers and Millennials now use Facebook; Boomers growing fast

% of all adults in each generation who say they use Facebook



Source: Survey conducted Jan. 3-10, 2018.

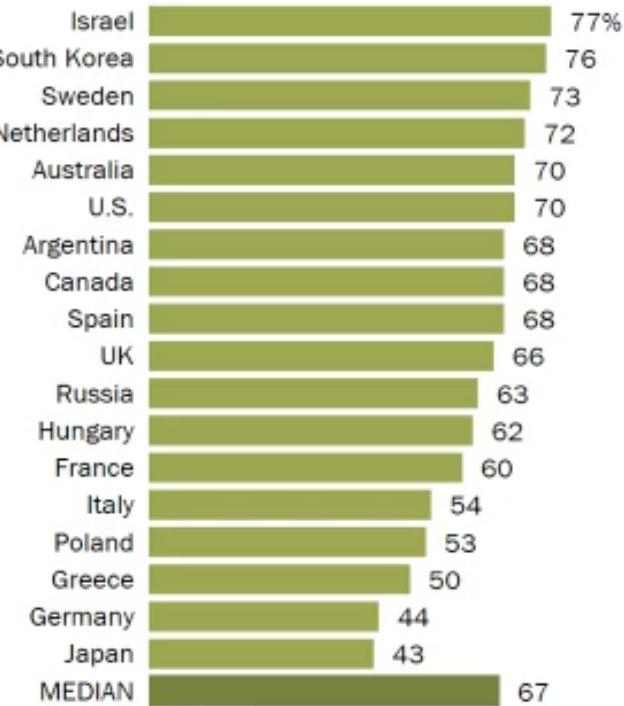
PEW RESEARCH CENTER

# The social media (r)evolution

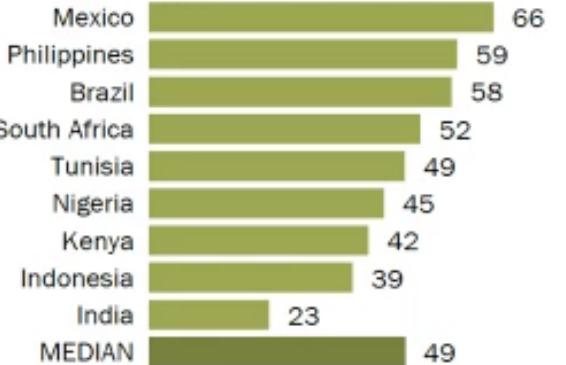
## Many in both advanced and emerging economies use social media

% of adults who use social media

### Advanced economies



### Emerging economies



Source: Spring 2018 Global Attitudes Survey, Q47.

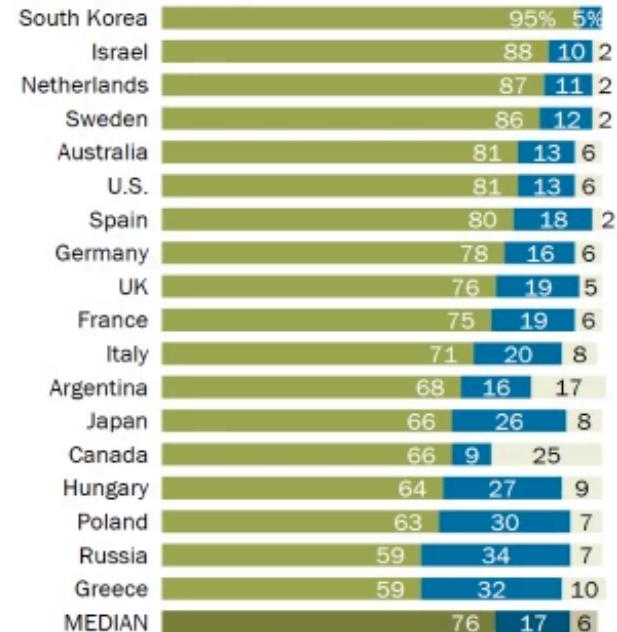
PEW RESEARCH CENTER

## Smartphone ownership in advanced economies higher than in emerging

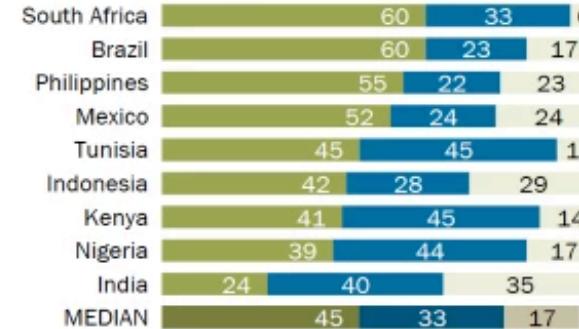
% of adults who report owning ...

- A smartphone
- A mobile phone that is not a smartphone
- No mobile phone

### Advanced economies



### Emerging economies



Source: Spring 2018 Global Attitudes Survey, Q45 & Q46.

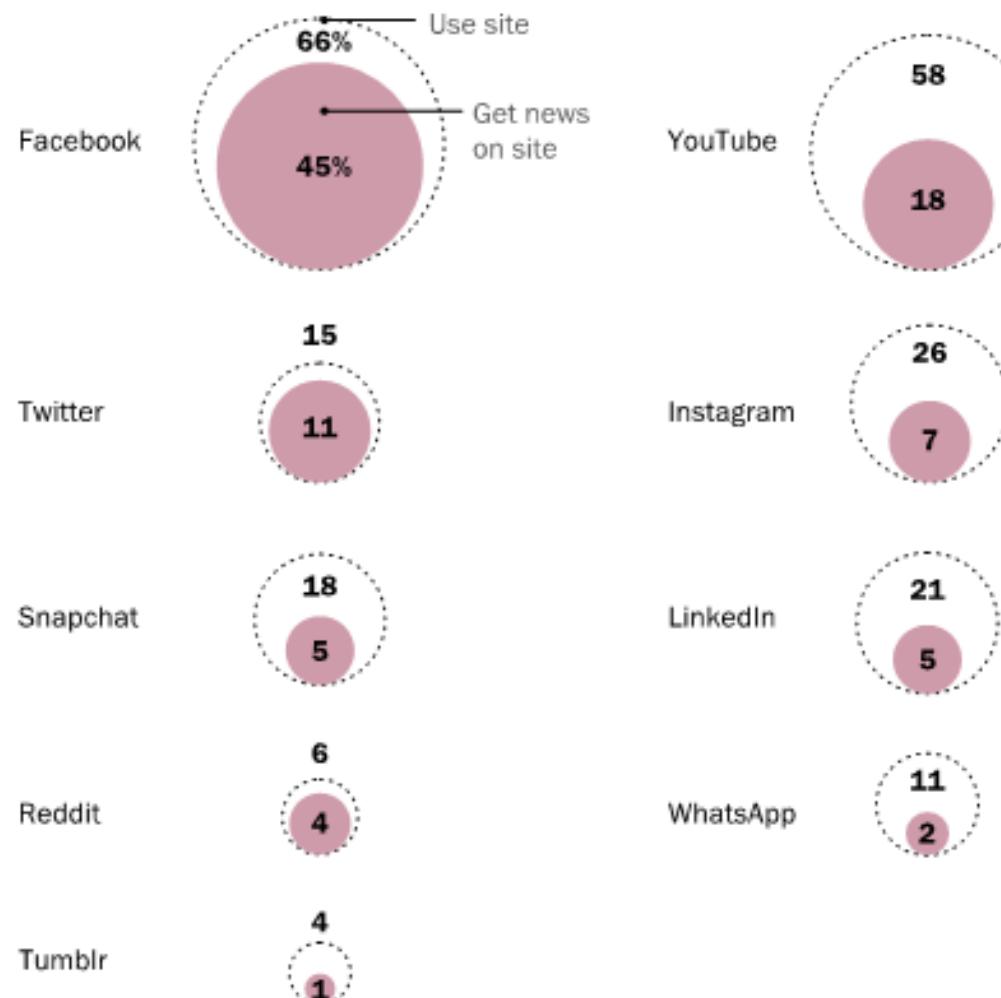
PEW RESEARCH CENTER

# The social media (r)evolution

## Social media sites as pathways to news

@patyrossini

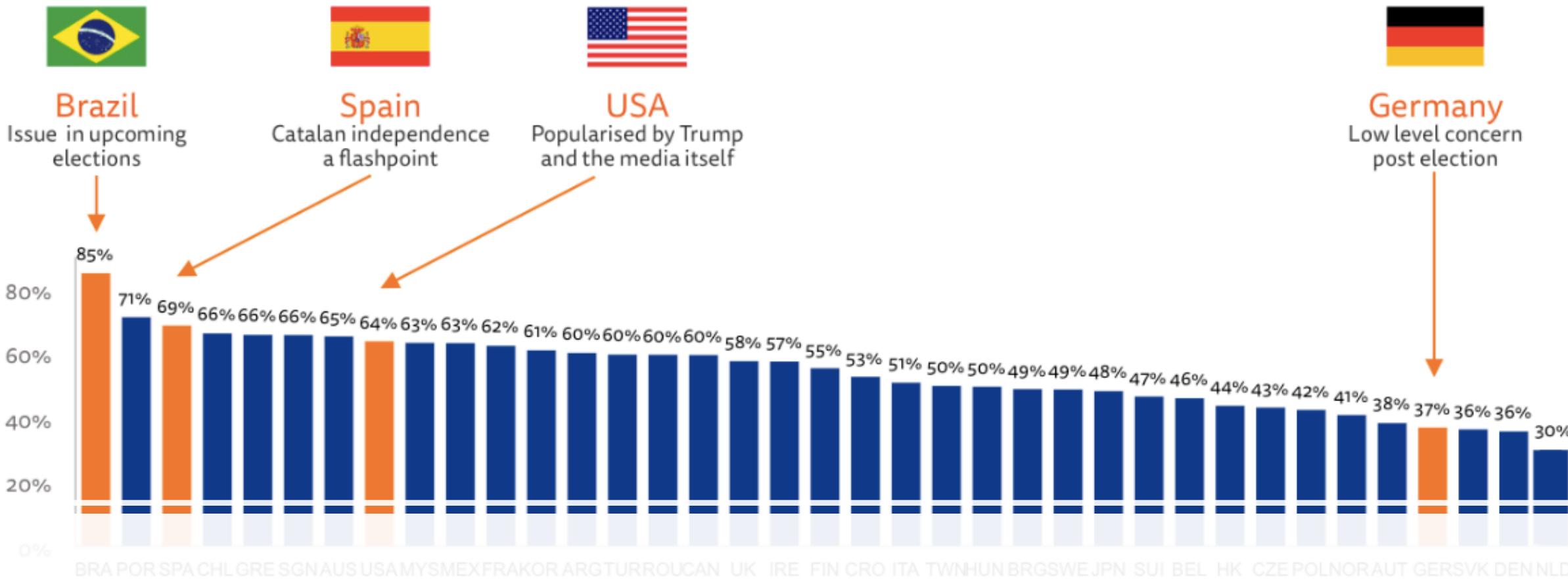
% of U.S. adults who use each social media site and % of U.S. adults who get news from each site



Source: Survey conducted Aug. 8-21, 2017.  
"News Use Across Social Media Platforms 2017"

PEW RESEARCH CENTER

# Varied concern about whether online news is real or fake



Q. FAKE NEWS 1. Please indicate your level of agreement with the following statement. Thinking about online news, I am concerned about what is real and what is fake on the internet.  
Base: Total sample in each market

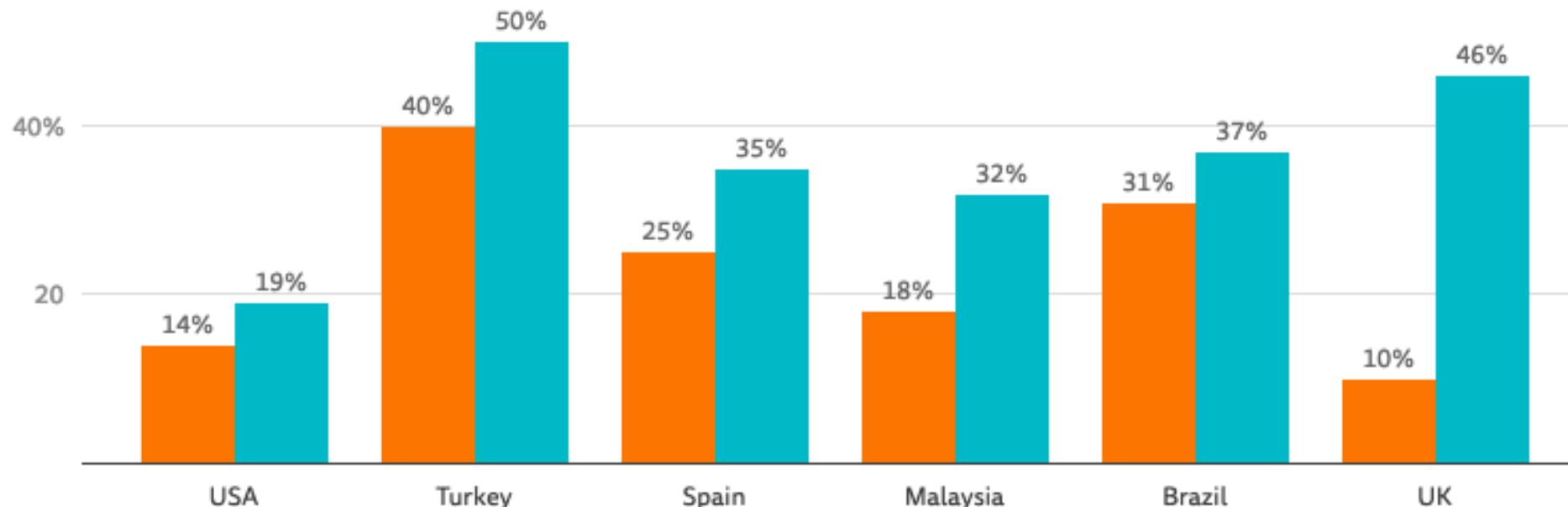


Fake news is an international problem, most notably in Brazil. Nic Newman/Reuters

# PROPORTION THAT TRUSTS MOST NEWS FROM SOCIAL MEDIA MOST OF THE TIME

*Selected countries*

■ National average ■ Members of news/politics groups on FB/WA



Q12\_2019\_FB/WA. Facebook/WhatsApp allows you to set up, join, and participate in groups, where you can discuss news or related topics with like-minded people. Which, if any, of the following have you used in the past month on Facebook/WhatsApp itself?

Q6\_2018\_2. Please indicate your level of agreement with the following statements: I think I can trust news in social media most of the time.

*Base: Total sample/member of news/politics groups on Facebook or WhatsApp: UK = 2023/125, USA = 2012/197, Spain = 2005/243, Brazil = 2013/477, Turkey = 2074/549, Malaysia = 2101/380.*



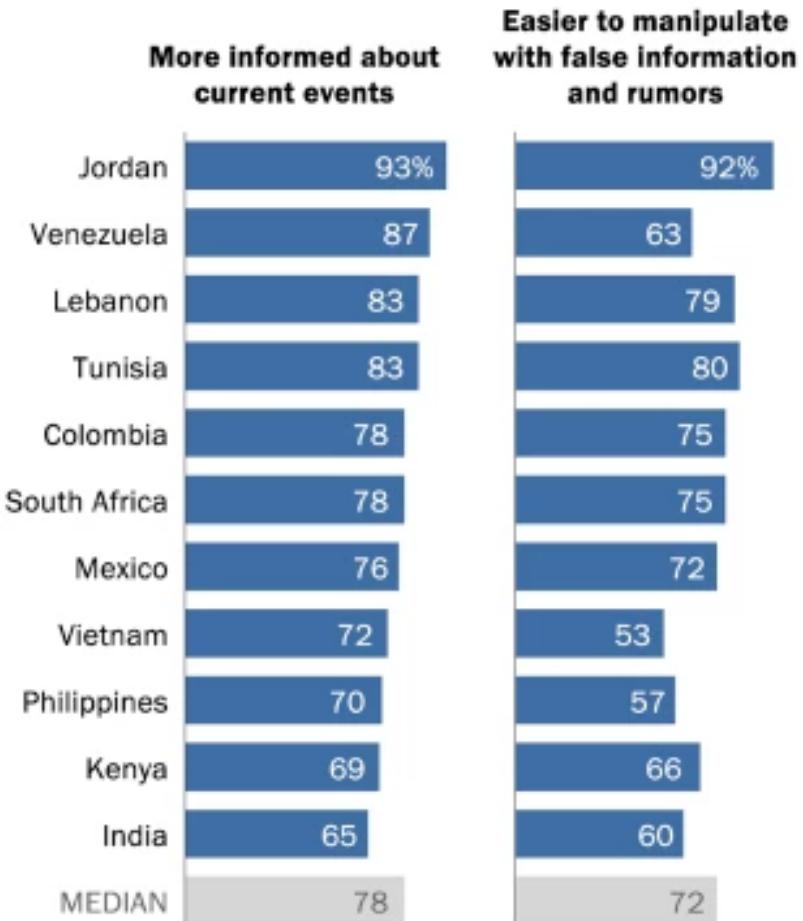
[Get the data](#) • [Embed](#)



# The social media (r)evolution

## Majorities say access to technology has made people more informed, but also easier to manipulate

% of adults who say access to mobile phones, the internet and social media have made people ...



Source: Mobile Technology and Its Social Impact Survey 2018. Q15a & Q15d.

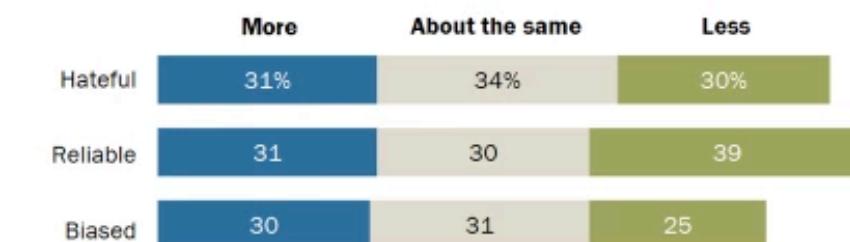
"Publics in Emerging Economies Worry Social Media Sow Division, Even as They Offer New Chances for Political Engagement"

## Social media users in 11 emerging economies regularly encounter negative content on these platforms

% of social media platform and messaging app users who frequently/occasionally see articles and other content when they use social media that ...



% of social media platform and messaging app users who say that the news and information they get on social media is more/about the same/less compared with other sources



Note: Percentages are 11-country medians. Social media and messaging app users include those who said they use one or more of the seven specific online platforms measured in this survey.

Source: Mobile Technology and Its Social Impact Survey 2018. Q32c, Q32d, Q32f, Q33a & Q33b.

"Publics in Emerging Economies Worry Social Media Sow Division, Even as They Offer New Chances for Political Engagement"

PEW RESEARCH CENTER

# Why care about social media?

- It's pervasive in our society and central to many everyday activities
- It mediates relationships, access to news and information, community building and so on
- Scale: everyone is on, and anything can spread quickly
- It has the potential to influence users' behaviors
  - When misused, it can pose real-life threats to democratic values (e.g. misinformation) and to society as a whole



What is social media  
research?

# Social Media Research

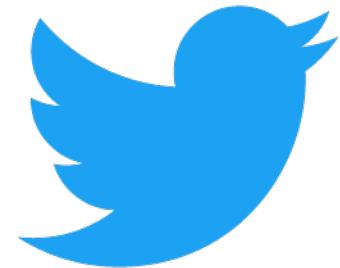
Research that focuses on analyzing social media..

- **Behaviors**
- **Data**
- **Infrastructures**
- **Affordances**
- **Practices**

Social Media research is not  
limited to studies that uses data  
from social media.

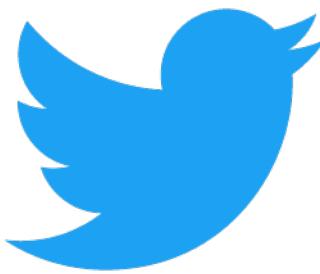
# Social Media Research Approaches

- Social media as **a source of data**
  - User behaviors (e.g. clicks, likes, shares)
  - Interpersonal networks (homogeneity/heterogeneity)
  - Elite behavior (e.g. celebrities, politicians)
  - Field experiments
- Social media **effects on behavior**
  - Political campaigns
  - Activism and social movements
  - Interpersonal communication & social capital
  - Political/cultural attitudes and behavior



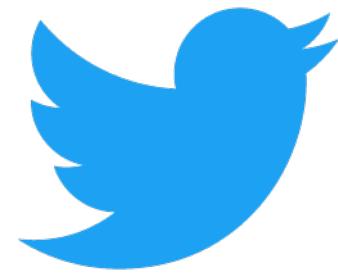
# Social Media Research

- Focus on **data from social media**:
  - Content
  - Engagement
  - Information Diffusion / Flow
  - Relationships
  - Can focus on institutional or public discourse
- Focus on **users' perceptions or behaviors**
  - Experiments
  - Surveys
  - Qualitative work (e.g. interviews, participant observation, (n)ethnography)



# Social Media Research

- Focus on **social media design & affordances**:
  - Design vs. behavior
  - Algorithms
  - Privacy
  - Network structures
- Focus on **social media companies' practices**
  - Policies and terms of use
  - Practices (e.g. moderation, Advertising, targeting)



# Collecting Data: The API

## API: Application Programming Interface

- A “contract” between two software programs for how they will interact, share, or transmit data
  - It relies on some sort of authentication (user level, application level)
- Allows developers to create applications that use data from social networking and Web platforms
- Allows researchers/analysts to analyze user behavior from data from social networking and Web platforms

# The API

- **Facebook Graph API:**
  - the core of Facebook Platform
  - enables developers to read from and write data into Facebook
  - enables the use of Facebook to login to other apps
  - presents a simple, consistent view of the Facebook social graph, uniformly representing objects in the graph (e.g., people, photos, events, and pages)
  - Presents the connections between them (e.g., friend relationships, shared content, and photo tags).
- Since early 2018 (post Cambridge Analytica scandal), the API is virtually closed for research and all ‘apps’ have to be reviewed.

# The API

- Twitter has **some different Public APIs:**
  - Streaming API
    - Push-based
    - Constant flow of data – used for tracking developing events and hashtags
    - Good for planned events, can't get historical data
    - Has limitations – only the firehose (enterprise) delivers 100% of what's posted.
    - Most queries will capture 1% of all tweets. For small events, that might include all tweets.
  - REST API
    - Social graphs – followers, friends
    - Engagement metrics (favorites, replies, retweets)
    - Search (for the past 7 days)

**Twitter has different levels of access: Standard (free), Premium & Enterprise (paid)**

# The API

- On Instagram:
  - Instagram Graph API (similar to Facebook)
  - API enables downloading data about users, relationships, media, comments, likes, tags, location.
  - Get lists of followers and followed-by
  - Data from *public profiles* is available for collecting. For private profiles, only the owner can collect the data.

# API policies

- APIs are very strict in terms of what you can collect and do.
- Eg:
  - you can't **share data with you collected (Twitter/Facebook)**
  - you can't **use data collected through the API for commercial purposes**
  - **There are strict limits to the amount of data you can collect (particularly on Twitter)**
- But... people don't always follow the rules (Cambridge Analytica!)
- <http://www.trumptwitterarchive.com/>

# Mobile Instant Messaging (MIMs)

Research has largely focused on social media platforms (public/semi-public data).

Yet, while SM use is decreasing globally, the use of messaging apps is rising (Reuters Digital News Report 2018 & 2019)

Messaging apps are a lot more challenging to study.  
Can you guess why?



# Mobile Instant Messaging (MIMs)

- Private communication – and often encrypted
- Data is not available through APIs
- Can't get 'big'/representative data
- Requires 'traditional' or 'creative' approaches
  - Interviews
  - Focus groups
  - Participant observation
  - Content analysis
  - Surveys





What are the  
challenges of social  
media research?

# Challenges

- Access to platform-level data is conditional on ToS and API use
  - Facebook cracked down on research using its API in the spring of 2018, introducing an app review that is nearly impossible to clear
  - Twitter is implementing new app review measures and cracking down on API calls and limits
  - Techniques such as *scrapping* are seen as alternatives to APIs, but they violate ToS

# Challenges

- Big data
- Population bias
- Sampling & Case selection
- Spam/bots/non-genuine behavior
- Privacy
- Ethics

# Big data biases & sampling

- Population bias
  - Sociodemographic characteristics are correlated with social media use:
    - E.g. wealthier, more educated also more likely to use SM
  - Representativeness
    - Observed behavior (e.g., talking about politics or a TV show) tends to refer to a particularly active subset of the population
- Algorithmic bias
  - Proprietary API does not return all publicly available information (e.g. Twitter) and is not transparent about what/how much is left out

# Ethical Research on Social Media

Presented by Dr. Emily Johnson  
University of California, Berkeley

July 15, 2023  
10:00 AM - 11:30 AM  
Virtual Event

Abstract:  
This presentation will explore the ethical challenges and best practices for conducting research on social media platforms. We will discuss issues such as data privacy, informed consent, and the potential for bias and manipulation in research designs. Additionally, we will examine the role of social media in shaping public discourse and the responsibilities of researchers in addressing these dynamics.

Biography:  
Dr. Emily Johnson is a sociologist and data scientist at the University of California, Berkeley. Her research focuses on the intersection of technology and society, with a particular emphasis on the impact of social media on political communication and public opinion. She has published numerous articles in leading academic journals and has presented her work at international conferences. Dr. Johnson is also a frequent speaker at industry events and has provided consulting services to several tech companies.

APRIL

10 2014

'G

C  
P

Justine  
racist tv

Justine Sacco,  
1 people living w

Adam R



The fa  
impac

Going  
kiddin

Reply

2,678  
RETWEETS

10:19 AM -

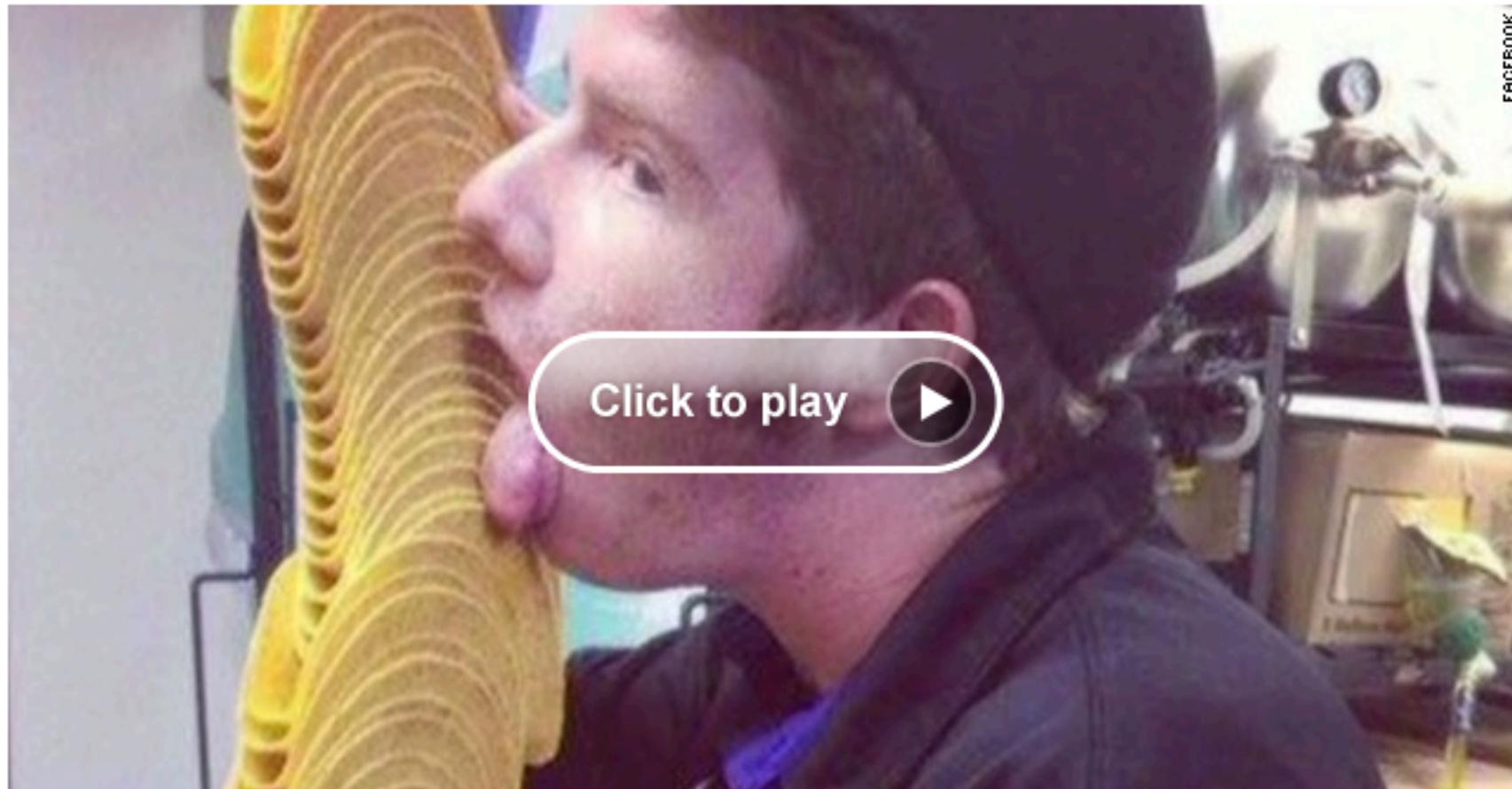
Nicole  
fire th  
an up  
quickl  
qualifi

you to  
create

# 10 people who learned social media can get you fired

By Ryan Broderick and Emanuella Grinberg, BuzzFeed/CNN

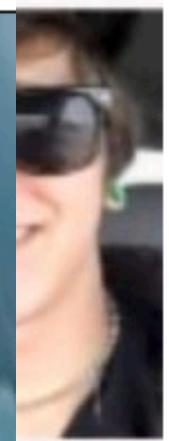
updated 1:30 PM EDT, Thu June 6, 2013



**Gross! Guy licks Taco Bell shells**

FACEBOOK

NT



10 to  
ios

ORY

I find it ironic that Detroit is known as [#motorcity](#) and yet no one here knows how to fucking drive

about 3 hours ago via web



@ChryslerAutos

Chrysler Autos

Our apologies - our account was compromised earlier and we are taking steps to resolve it.

1 hour ago via web [☆ Favorite](#) [Retweet](#)

Retweeted by [piercecp](#) and others



**Connor Riley**  
@theconnor

Cisco just offered me a job! Now I have to weigh the utility of a fatty paycheck against the daily commute to San Jose and hating the work.

@theconnor Who is the hiring manager. I'm sure they would love to know that you will hate the work. We here at Cisco are versed in the web.

about 18 hours ago from TweetDeck in reply to theconnor

# Twitter restrictions on Elon Musk are set to take hold. But is Tesla willing to reel him in?

Originally published December 27, 2018 at 4:09 pm | Updated December 27, 2018 at 10:22 pm



Elon Musk

@elonmusk

Am consider  
\$420. Fundi

12:48 PM - 7 Aug 2018

15,913 Retweets 87,67

6.3K 16K



Tweet your i

Elon Musk @elonmu

Shareholders could eith

1.3K 2.2K



ng secured'

ecutors claim

Tesla's board must now "oversee" CEO Elon Musk's public communications, including his Twitter feed. Musk is shown speaking at the unveiling for the Boring Co. test tunnel in Hawthorne, Calif., on Dec. 18. (Robyn Beck/Pool Photo via AP)

Under a fraud settlement Tesla CEO Elon Musk signed with the SEC, Tesla's board must now "oversee" Musk's public communications. Whether the oversight will fundamentally alter his Twitter activity is open to doubt.

What are the ethical  
considerations to  
conduct research on  
Social Media?

# Is the data really public?

Social media data, specially when published in ‘public’ profiles (e.g. tweets) is widely seen as public

However, because this data involves individual persons, consideration of principles related to research on human subjects may be necessary even if it is not immediately apparent how and where persons are involved in the research data.

# Is the data really public?

Ethical issues may arise and need to be addressed during all steps of the research process, from planning, research conduct, publication, and dissemination.

Social media data can be de-anonymized (Zimmer, 2010)

Rule of thumb: the greater the vulnerability of the community / author / participant, the greater the obligation of the researcher to protect the community / author / participant.

# Is the data really public?

Example: ethical considerations studying uncivil and intolerant comments online

Concern: topic is potentially sensitive/harmful to participants insofar as being publicly connected to uncivil/intolerant messages can affect them personally

Solution: anonymized individuals (even if people use pseudonyms), don't use direct quotations, analysis conducted on aggregated data only

# Is the data really public?

Example: Facebook massive-scale contagion experiment

## **Experimental evidence of massive-scale emotional contagion through social networks**



Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock

PNAS June 17, 2014 111 (24) 8788-8790; first published June 2, 2014 <https://doi.org/10.1073/pnas.1320040111>

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)

# Is the data really public?

Example: Facebook massive-scale contagion experiment

## Ethical issues:

- violated principles of informed consent (assumed ToS as consent)
- Participants did not have the option to opt out

# Am I violating terms of service?

Before collecting and analyzing data, you must read terms of service to make sure you are not violating them.

Examples:

1. Scraping data from Twitter instead of using the API violates terms of service
2. Sharing data collected through the API with other researchers is a violation of ToS [Facebook/Twitter]
3. Analyzing deleted tweets is against Twitter's ToS

Challenge: terms of service are fluid and change fast.

# Hands on: Collecting Twitter Data



# Types of data

## Timeline & user data:

- Collect up to 3200 tweets from one or more user
- Can include replies / retweets
- Collect networks (followers/follows)

## Search data:

- Collect a random sample of tweets based on hashtags, keywords, geolocation
- Search for users based on their public bios

# Let's get to work!

