POIR 613: Measurement Models and Statistical Computing

Pablo Barberá

School of International Relations
University of Southern California
`pablobarbera.com`

Course website:

pablobarbera.com/POIR613/

Today

1. Computational social science research: challenges and opportunities
2. Discussion: ethics of Big Data research.
   - Kramer et al 2014 (and "Editorial Expression of Concern")
   - Hargittai 2018
3. Good coding / programming practices

Logistics

1. Referee reports:
   - You should all have already signed up
   - Due day before class at 8pm
2. Class project:
   - One-paragraph idea due September 20

Computational Social Science

Shift in communication patterns

Computational Social Science

Two different approaches in the growing field of computational social science:

1. Big data as a new source of information
   - Behavior, opinions, and latent traits
   - Interpersonal networks
   - Elite behavior
   - Affordable online experiments
2. How big data and social media affect social behavior
   - Collective action and social movements
   - Political campaigns
   - Social capital and interpersonal communication
   - Political attitudes and behavior

Computational Social Science

Two different approaches in the growing field of computational social science:

1. Big data as a new source of information
   - Behavior, opinions, and latent traits
   - Interpersonal networks
   - Elite behavior
   - Affordable online experiments
2. How big data and social media affect social behavior
   - Collective action and social movements
   - Political campaigns
   - Social capital and interpersonal communication
   - Political attitudes and behavior

Behavior, opinions, and latent traits

- Digital footprints: check-ins, conversations, geolocated pictures, likes, shares, retweets, ...

$\rightarrow$ Non-intrusive measurement of behavior and public opinion

Regular Article

new media & society
2014, Vol. 16(2) 340–358

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1461444813480466
nms.sagepub.com
SAGE

Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France

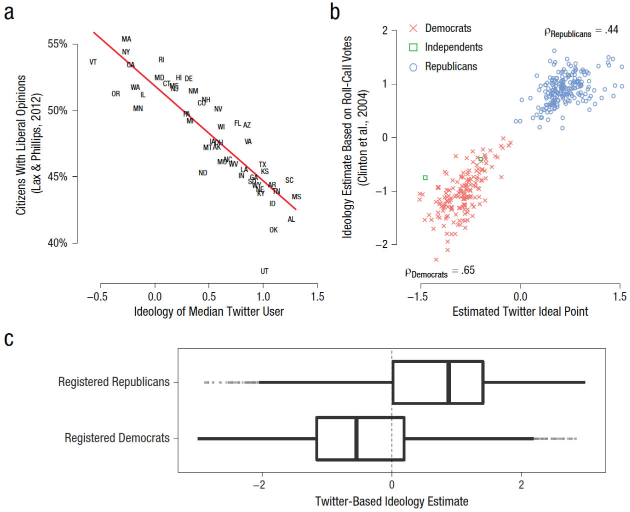
Andrea Ceron, Luigi Curini, Stefano M Iacus
Università degli Studi di Milano, Italy

Giuseppe Porro
Università degli Studi dell'Insubria, Italy

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

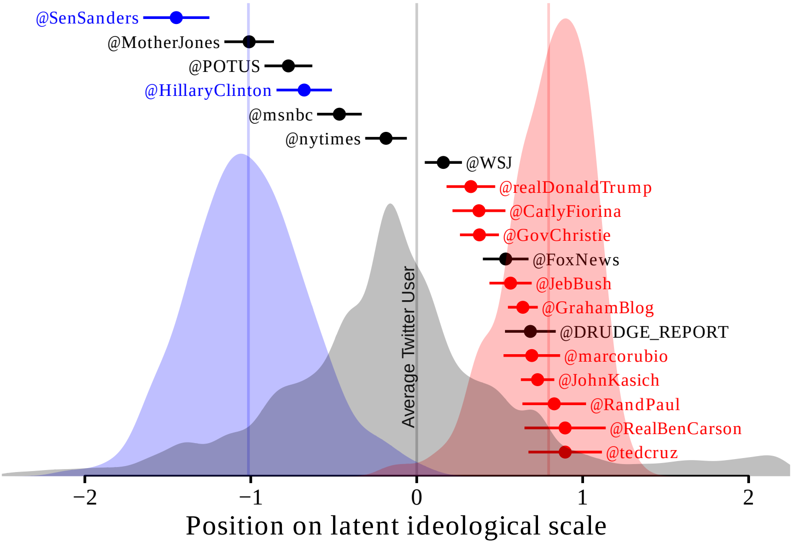
Behavior, opinions, and latent traits

→ Inference of latent traits: political knowledge, ideology, personal traits, socially undesirable behavior, . . .



Barberá, 2015 *Political Analysis*; Barberá et al, 2016, *Psychological Science*

Estimating political ideology using Twitter networks



Barberá "Who is the most conservative Republican candidate for president?" *The Monkey Cage / The Washington Post*, June 16 2015

Computational Social Science

Two different approaches in the growing field of computational social science:

1. Big data as a new source of information
   - Behavior, opinions, and latent traits
   - Interpersonal networks
   - Elite behavior
   - Affordable online experiments
2. How big data and social media affect social behavior
   - Collective action and social movements
   - Political campaigns
   - Social capital and interpersonal communication
   - Political attitudes and behavior

Interpersonal networks

- Political behavior is social, strongly influenced by peers

Bond et al, 2012, "A 61-million-person experiment in social influence and political mobilization", *Nature*

- Costly to measure network structure
- High overlap across online and offline social networks

OPEN ACCESS Freely available online PLOS ONE

Inferring Tie Strength from Online Directed Behavior

Jason J. Jones[1,2*], Jaime E. Settle[2], Robert M. Bond[2], Christopher J. Fariss[2], Cameron Marlow[3], James H. Fowler[1,2]

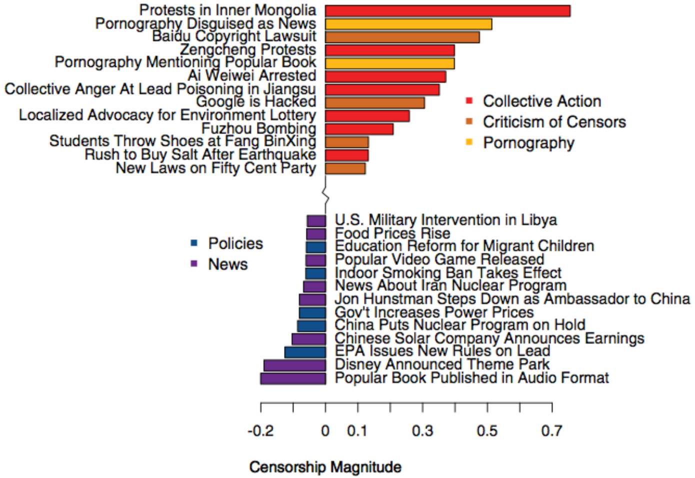
Computational Social Science

Two different approaches in the growing field of computational social science:

1. Big data as a new source of information
   - Behavior, opinions, and latent traits
   - Interpersonal networks
   - Elite behavior
   - Affordable online experiments
2. How big data and social media affect social behavior
   - Collective action and social movements
   - Political campaigns
   - Social capital and interpersonal communication
   - Political attitudes and behavior

Elite behavior

- Authoritarian governments' response to threat of collective action



King et al, 2013, "How Censorship in China Allows Government Criticism but Silences Collective Expression", *APSR*

- Estimation of conflict intensity in real time

Computational Social Science

Two different approaches in the growing field of computational social science:

1. Big data as a new source of information
    - Behavior, opinions, and latent traits
    - Interpersonal networks
    - Elite behavior
    - Affordable online experiments
2. How big data and social media affect social behavior
    - Collective action and social movements
    - Political campaigns
    - Social capital and interpersonal communication
    - Political attitudes and behavior

Affordable field experiments

Political Behavior
September 2017, Volume 39, Issue 3, pp 629–649 | Cite as

Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment

Authors Authors and affiliations

Kevin Munger

Original Paper
First Online: 11 November 2016

[redacted] · 13 Sep 2015
@[redacted] don't be a ni ger

Rasheed [redacted]
@Rasheed[redacted]

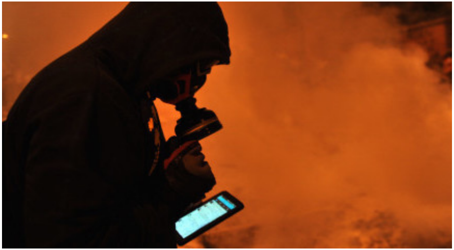
@[redacted] Hey man, just remember that there are real people who are hurt when you harass them with that kind of language

Computational Social Science

Two different approaches in the growing field of computational social science:

1. Big data as a new source of information
   - Behavior, opinions, and latent traits
   - Interpersonal networks
   - Elite behavior
   - Affordable online experiments
2. How big data and social media affect social behavior
   - Collective action and social movements
   - Political campaigns
   - Social capital and interpersonal communication
   - Political attitudes and behavior


MY BODY
NOT YOURS
LOVE
AND PROUD



#OccupyGezi

#OccupyWallStreet

#Euromaidan

#Indignados

slacktivism?

why the revolution will not be tweeted

When the sit-in movement spread from Greensboro throughout the South, it did not spread indiscriminately. It spread to those cities which had preexisting "movement centers" – a core of dedicated and trained activists ready to turn the "fever" into action.

The kind of activism associated with social media isn't like this at all. [. . .] Social networks are effective at increasing participation – by lessening the level of motivation that participation requires.

***Gladwell**, Small Change (New Yorker)*

You can't simply join a revolution any time you want, contribute a comma to a random revolutionary decree, rephrase the guillotine manual, and then slack off for months. Revolutions prize centralization and require fully committed leaders, strict discipline, absolute dedication, and strong relationships.

When every node on the network can send a message to all other nodes, confusion is the new default equilibrium.

***Morozov**, The Net Delusion: The Dark Side of Internet Freedom*

the critical periphery

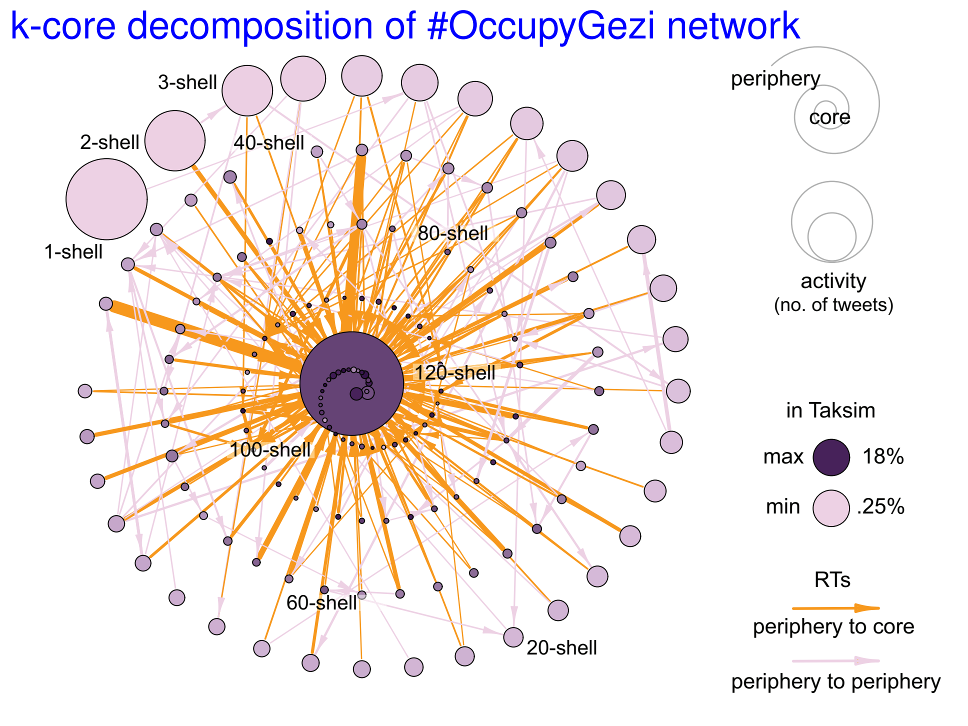
PLOS | ONE

RESEARCH ARTICLE

The Critical Periphery in the Growth of Social Protests

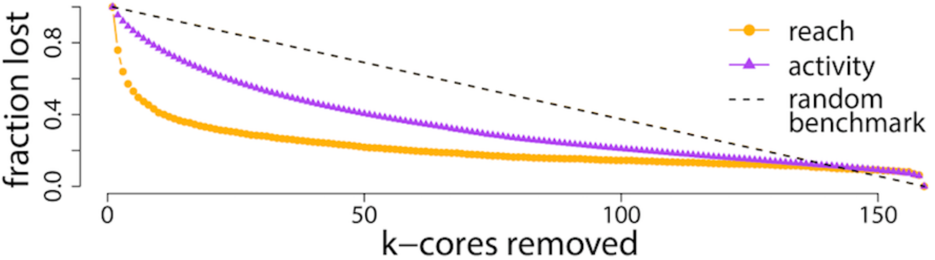
Pablo Barberá[1]*, Ning Wang[2], Richard Bonneau[3,4], John T. Jost[1,5,6], Jonathan Nagler[6], Joshua Tucker[6], Sandra González-Bailón[7]*

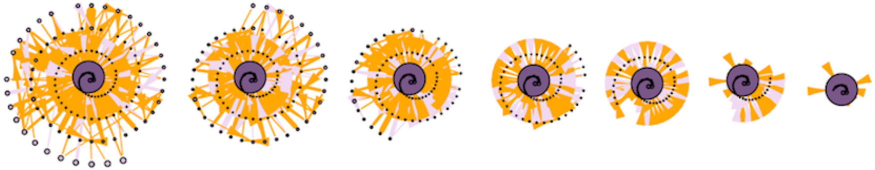
- Structure of online protest networks:
  1. Core: committed minority of resourceful protesters
  2. Periphery: majority of less motivated individuals
- Our argument: key role of peripheral participants
  1. Increase reach of protest messages (positional effect)
  2. Large contribution to overall activity (size effect)



k-core decomposition of #OccupyGezi network
1-shell
2-shell
3-shell
40-shell
80-shell
120-shell
100-shell
60-shell
20-shell
periphery
core
activity
(no. of tweets)
in Taksim
max
18%
min
.25%
RTs
periphery to core
periphery to periphery

Relative importance of core and periphery

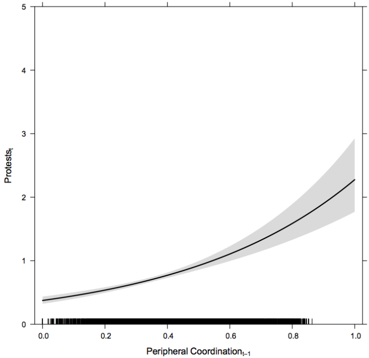




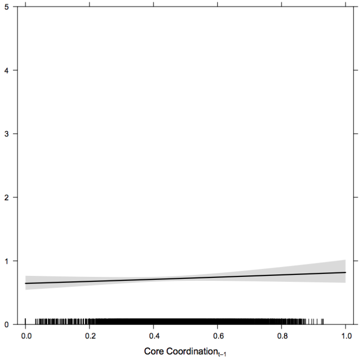
reach: aggregate size of participants' audience

activity: total number of protest messages published (not only RTs)

Peripheral mobilization during the Arab Spring



(a) Increase in protest as peripheral coordination increases



(b) Coordination does not come through core individuals

Steinert-Threlkeld (APSR 2017) “Spontaneous Collective Action”

Social media and democracy

FROM LIBERATION TO TURMOIL: SOCIAL MEDIA AND DEMOCRACY

Joshua A. Tucker, Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá

"How can one technology – social media – simultaneously give rise to hopes for liberation in authoritarian regimes, be used for repression by these same regimes, and be harnessed by antisystem actors in democracy? We present a simple framework for reconciling these contradictory developments based on two propositions: 1) that social media give voice to those previously excluded from political discussion by traditional media, and 2) that although social media democratize access to information, the platforms themselves are neither inherently democratic nor nondemocratic, but represent a tool political actors can use for a variety of goals, including, paradoxically, illiberal goals."

***Journal of Democracy**, 2017*

Computational Social Science

Two different approaches in the growing field of computational social science:

1. Big data as a new source of information
   - Behavior, opinions, and latent traits
   - Interpersonal networks
   - Elite behavior
   - Affordable online experiments
2. How big data and social media affect social behavior
   - Collective action and social movements
   - Political campaigns
   - Social capital and interpersonal communication
   - Political attitudes and behavior

Barack Obama
@BarackObama

Follow

Four more years.

RETWEETS 756,411 | FAVORITES 288,867

11:16 PM - 6 Nov 2012

Post Politics

By the end of the 2012 campaign, every Mitt Romney tweet had to be approved by 22 people

Political persuasion

Social media as a new campaign tool:

> *"Let me tell you about Twitter. I think that maybe I wouldn't be here if it wasn't for Twitter. [...] Twitter is a wonderful thing for me, because I get the word out... I might not be here talking to you right now as president if I didn't have an honest way of getting the word out."*
>
> ***Donald Trump**, March 16, 2017 (Fox News)*

- Diminished gatekeeping role of journalists
  - Part of a trend towards citizen journalism (Goode, 2009)
- Information is contextualized within social layer
  - Messing and Westwood (2012): social cues can be as important as partisan cues to explain news consumption through social media
- Real-time broadcasting in reaction to events
  - e.g. *dual screening* (Vaccari et al, 2015)
- Micro-targeting
  - Affects how campaigns perceive voters (Hersh, 2015), but unclear if effective in mobilizing or persuading voters

Computational Social Science

Two different approaches in the growing field of computational social science:

1. Big data as a new source of information
   - Behavior, opinions, and latent traits
   - Interpersonal networks
   - Elite behavior
   - Affordable online experiments
2. How big data and social media affect social behavior
   - Collective action and social movements
   - Political campaigns
   - Social capital and interpersonal communication
   - Political attitudes and behavior

Social capital

- Social connections are essential in democratic societies, but online interactions do not facilitate creation and strengthening of social capital (Putnam, 2001)
- Online networking sites facilitate and transform how social ties are established

American Politics Research
1–31

DOI: 10.1177/1532673X14557942
apr.sagepub.com
SAGE

Tweeting Alone? An Analysis of Bridging and Bonding Social Capital in Online Networks

Javier Sajuria[1], Jennifer vanHeerde-Hudson[1], David Hudson[1], Niheer Dasandi[1], and Yannis Theocharis[2]

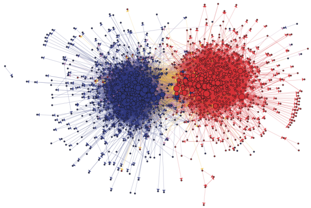
Computational Social Science

Two different approaches in the growing field of computational social science:

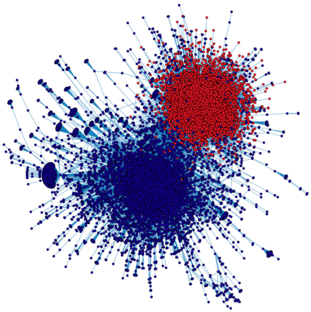
1. Big data as a new source of information
   - Behavior, opinions, and latent traits
   - Interpersonal networks
   - Elite behavior
   - Affordable online experiments
2. How big data and social media affect social behavior
   - Collective action and social movements
   - Political campaigns
   - Social capital and interpersonal communication
   - Political attitudes and behavior

Social media as echo chambers?

- communities of like-minded individuals (homophily, influence)



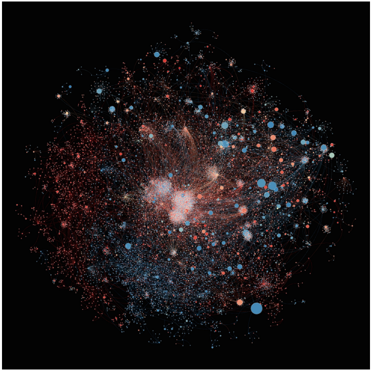
Adamic and Glance (2005)



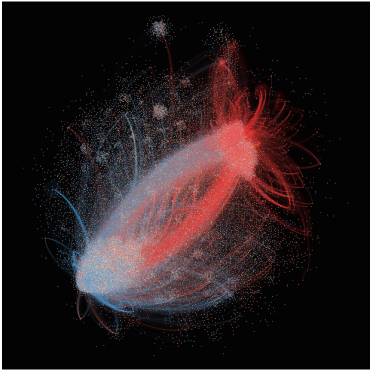
Conover et al (2012)

- ...generates selective exposure to congenial information
- ...reinforced by ranking algorithms – “filter bubble” (Parisier)
- ...increases political polarization (Sunstein, Prior)

Social media as echo chambers?



2013 SuperBowl

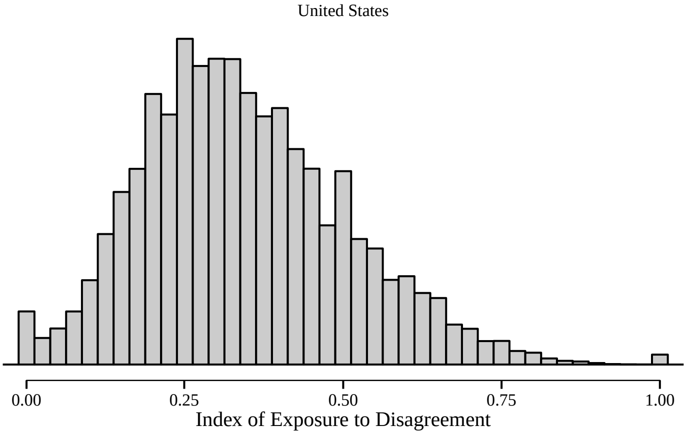


2012 Election

Barberá et al (2015) "Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?" *Psychological Science*

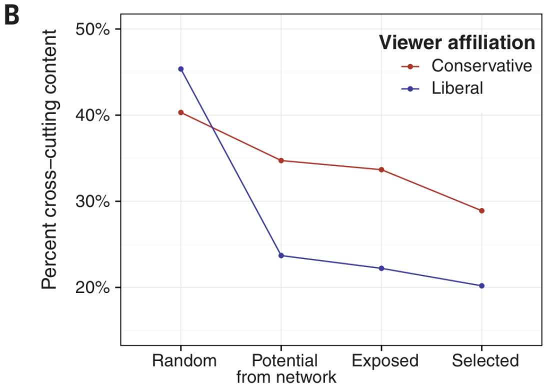
Measuring exposure to cross-cutting content

Most Twitter users are exposed to high levels of political disagreement



Social media as echo chambers?

Fig. 3. Cross-cutting content at each stage in the diffusion process. (**A**) Illustration of how algorithmic ranking and individual choice affect the proportion of ideologically cross-cutting content that individuals encounter. Gray circles illustrate the content present at each stage in the media exposure process. Red circles indicate conservatives, and blue circles indicate liberals. (**B**) Average ideological diversity of content (i) shared by random others (random), (ii) shared by friends (potential from network), (iii) actually appeared in users' News Feeds (exposed), and (iv) users clicked on (selected).



Bakshy, Messing, & Adamic (2015) "Exposure to ideologically diverse news and opinion on Facebook". *Science.*

Fake news?

- Guess et al (2018, 2019); Grinberg et al (2019): who consumes misinformation?
  - 25% Americans exposed to fake news sites in 2016; 6% of all news consumption; but heavily concentrated (1% saw 80%)
  - Older, conservative people more likely to be exposed
  - Fact-check does not reach consumers of misinformation

- Allcott and Gentzkow (2017): does it matter?
  - Survey experiment with real and placebo fake news stories
  - Most people do not remember seeing fake news stories
  - Unlikely to affect citizens' behavior

Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature

WILLIAM + FLORA
Hewlett Foundation

By Joshua A. Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan

SHARE

Computational Social Science

Two different approaches in the growing field of computational social science:

1. Big data as a new source of information
   - Behavior, opinions, and latent traits
   - Interpersonal networks
   - Elite behavior
   - Affordable online experiments
2. How big data and social media affect social behavior
   - Collective action and social movements
   - Political campaigns
   - Social capital and interpersonal communication
   - Political attitudes and behavior

Today

1. Computational social science research: challenges and opportunities
2. Discussion: ethics of Big Data research.
   - Kramer et al 2014 (and “Editorial Expression of Concern”)
   - Hargittai 2018
3. Good coding / programming practices

What are the most important challenges when working with Big Data?

Big data and social science: challenges

1. Big data, big bias?
2. The end of theory?
3. Spam and bots
4. The privacy paradox
5. Generalizing from online to offline behavior
6. Ethical concerns

1. Big data, big bias?

SOCIAL SCIENCES

Social media for large studies of behavior

Large-scale studies of human behavior in social media need to be held to higher methodological standards

By **Derek Ruths**[1]* *and* **Jürgen Pfeffer**[2]

On 3 November 1948, the day after Harry Truman won the United States presidential elections, the *Chicago Tribune* published one of the most famous erroneous headlines in newspaper history: "Dewey Defeats Truman" (*1*, *2*). The headline was informed by telephone surveys, which had inadver-

different social media platforms (*8*). For instance, Instagram is "especially appealing to adults aged 18 to 29, African-American, Latinos, women, urban residents" (*9*) whereas Pinterest is dominated by females, aged 25 to 34, with an average annual household income of $100,000 (*10*). These sampling biases are rarely corrected for (if even acknowledged).

Proprietary algorithms for public data. Platform-specific sampling problems, for example, the highest-volume source of pub-

The rise of "embedded researc searchers who have special rela with providers that give them ele cess to platform-specific data, al and resources) is creating a divic media research community. Such ers, for example, can see a platfor workings and make accommodat may not be able to reveal their c or the data used to generate their f

Ruths and Pfeffer, 2015, "Social media for large studies of behavior", *Science*

Big data, big bias?

Sources of bias (Ruths and Pfeffer, 2015; Lazer et al, 2017)

- Population bias
  - Sociodemographic characteristics are correlated with presence on social media
- Self-selection within samples
  - Partisans more likely to post about politics (Barberá & Rivero, 2014)
- Proprietary algorithms for public data
  - Twitter API does not always return 100% of publicly available tweets (Morstatter et al, 2014)
- Human behavior and online platform design
  - e.g. *Google Flu* (Lazer et al, 2014)

1. Big data, big bias?

Reducing biases and flaws in social media data

DATA COLLECTION

- 1. Quantifies platform-specific biases (platform design, user base, platform-specific behavior, platform storage policies)
- 2. Quantifies biases of available data (access constraints, platform-side filtering)
- 3. Quantifies proxy population biases/mismatches

METHODS

- 4. Applies filters/corrects for nonhuman accounts in data
- 5. Accounts for platform and proxy population biases
  a. Corrects for platform-specific and proxy population biases
  OR
  b. Tests robustness of findings
- 6. Accounts for platform-specific algorithms
  a. Shows results for more than one platform
  OR
  b. Shows results for time-separated data sets from the same platform
- 7. For new methods: compares results to existing methods on the same data
- 8. For new social phenomena or methods or classifiers: reports performance on two or more distinct data sets (one of which was not used during classifier development or design)

Issues in evaluating data from social media. Large-scale social media studies of human behavior should i address issues listed and discussed herein (further discussion in supplementary materials).

Ruths and Pfeffer, 2015, "Social media for large studies of behavior", *Science*

2. The end of theory?

> *Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.*
>
> ***Chris Anderson**, Wired, June 2008*

> *Correlations are a way of catching a scientist's attention, but the models and mechanisms that explain them are how we make the predictions that not only advance science, but generate practical applications.*
>
> ***John Timmer**, Ars Technica, June 2008*

(Big) social media data as a complement - not a substitute - for theoretical work and careful causal inference.

3. Spam and bots

"Follow your coordinators. We need to start tweeting, all at the same time, using the hashtag #ItsTimeForMexico. . . and don't forget to retweet tweets from the candidate's account..."

Unidentified PRI campaign manager

minutes before the May 8, 2012 Mexican Presidential debate

3. Spam and bots

Ferrara et al, 2016, *Communications of the ACM*

4. The privacy paradox

Online data present a paradox in the protection of privacy: Data are at once too revealing in terms of privacy protection, yet also not revealing enough in terms of providing the demographic background information needed by social scientists.

Golder & Macy*, Digital footprints, 2014*

5. Generalizing from online to offline behavior

What makes online behavior different:

- Platform affordances may distort behavior (e.g. anonymity encourages vitriol)
- Tools extend innate capacities (e.g. Dunbar's number)
- Asymmetries in data availability

6. Ethical concerns

1. Shifting notion of *informed consent*

PNAS

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer[a,1], Jamie E. Guillory[b,2], and Jeffrey T. Hancock[b,c]

[a]Core Data Science Team, Facebook, Inc., Menlo Park, CA 94025; and Departments of [b]Communication and [c]Information Science, Cornell University, Ithaca, NY 14853

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

Emotional states can be transferred to others via emotional contagion, leading people to experience the same emotions without their awareness. Emotional contagion is well established in laboratory experiments, with people transferring positive and negative emotions to others. Data from a large real-world social network, collected over a 20-y period suggests that longer-lasting moods (e.g., depression, happiness) can be transferred through networks [Fowler JH, Christakis NA (2008) *BMJ* 337:a2338], although the results are controversial. In an experiment with people who use Facebook, we test whether emotional contagion occurs

demonstrated that (*i*) emotional contagion occurs via text-based computer-mediated communication (7); (*ii*) contagion of psychological and physiological qualities has been suggested based on correlational data for social networks generally (7, 8); and (*iii*) people's emotional expressions on Facebook predict friends' emotional expressions, even days later (7) (although some shared experiences may in fact last several days). To date, however, there is no experimental evidence that emotions or moods are contagious in the absence of direct interaction between experiencer and target.

On Facebook, people frequently express emotions, which are

2. Most personal data can be de-anonymized

Ethics and Information Technology

December 2010, Volume 12, Issue 4, pp 313–325

"But the data is already public": on the ethics of research in Facebook

Authors | Authors and affiliations

Principles for ethical research with Big Data

From Salganik, Chapter 6:

1. **Respect for persons**: treating persons as autonomous and respecting their wishes (informed consent)
2. **Beneficence**: (1) do not harm, and (2) maximize possible benefits and minimize (probability and severity of) possible harms.
3. **Justice**: risks and benefits of research should be distributed fairly
4. **Respect for Law and Public Interest**: compliance with law and transparency-based accountability

For next week

1. Submit coding challenge
2. Readings for discussion:
   - Bond et al (2012)
   - King et al (2014)
   - Munger (2017)
   - Bail et al (2018)
3. No background readings