

RECSM Summer School: Social Media and Big Data Research

Pablo Barberá

London School of Economics

www.pablobarbera.com

Course website:

pablobarbera.com/social-media-upf







George Takei

March 28 at 10:10pm ·

Who's with me.



Like · Comment · Share

408,735 people like this.

66,990 shares



Bon Alimago
@karma_thief

I need a hug. I have never been so traumatized by a television show.
#gameofthrones

More

RETWEETS 356 FAVORITES 110



10:06 PM - 2 Jun 2013



Google

how do I convert to

how do I convert to judaism

how do I convert to islam

how do I convert to catholicism

how do I convert to pdf



VIA 9GAG.COM

Press Enter to search.



Justin Bieber
@justinbieber

I make music. I love music.

More

RETWEETS 54,213 FAVORITES 59,205



10:09 PM - 7 Apr 2014



dustin curtis

@dcurtis



Follow

"At any moment, Justin Bieber uses 3% of our infrastructure. Racks of servers are dedicated to him. - A guy who works at Twitter

RETWEETS

1,528

FAVORITES

267



8:56 PM - 6 Sep 2010



...



Dmitry Medvedev @MedvedevRussiaE

Follow

The harmonious development of Crimea and Sevastopol as part of our state is one of the main objectives of the Russian Government

Reply Retweet Favorite More

RETWEETS
144

FAVORITES
57



10:39 AM - 21 Mar 2014



The New York Times

April 2

"Much of the foreign media coverage has distorted the reality of my country and the facts surrounding the events," writes Nicolás Maduro, the president of Venezuela, in Opinion: <http://nyti.ms/1gP5o2I>

Like · Comment · Share

57

262 people like this.

Top Comments ▾



Elizabeth Warren shared a link.
January 16

I'm not giving up on our fight to extend unemployment benefits. Watch my interview with Now With Alex Wagner about why we need to keep fighting.



Warren: This is the moment to back on economy
www.msnbc.com

President Obama faces one huge problem with his effort to improve the economy: an opposition party

Like · Comment · Share

15,483 720 1,041



Jackie Walorski
@RepWalorski

Follow

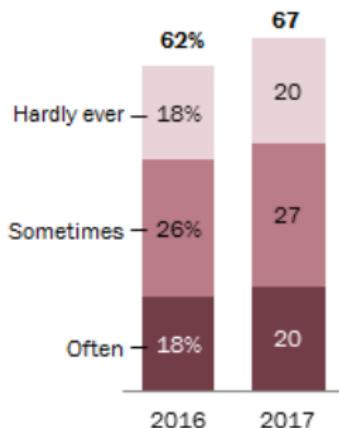
Today, a representative from my office will be meeting with constituents in Goshen. For more details, visit walorski.house.gov/services/upcom...

Reply Retweet Favorite More

11:22 AM - 8 Apr 2014

In 2017, two-thirds of U.S. adults get news from social media

% of U.S. adults who get news from social media sites ...



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

PEW RESEARCH CENTER

- ▶ 62% of Americans get news on social media (Pew)
- ▶ 27% of online EU citizens use social media to get news on national political matters (Eurobarometer, Fall 2012)
- ▶ Social media: top source of news for U.S. young adults (Pew)



Shift in communication patterns



Hello!

About me

- ▶ Assistant Professor of Computational Social Science at the [London School of Economics](#)
 - ▶ Previously Assistant Prof. at [Univ. of Southern California](#)
 - ▶ PhD in Politics, [New York University](#) (2015)
 - ▶ Data Science Fellow at [NYU](#), 2015–2016
- ▶ [My research:](#)
 - ▶ Social media and politics, comparative electoral behavior
 - ▶ Text as data methods, social network analysis, Bayesian statistics
 - ▶ Author of R packages to analyze data from social media
- ▶ [Contact:](#)
 - ▶ P.Barbera@lse.ac.uk
 - ▶ www.pablobarbera.com

This course

Two central questions:

1. **What** type of social science questions can I answer with social media data?
2. **How** would I answer those questions? What methods and tools would I use?

Today: research opportunities and challenges

- ▶ New and old social science questions
- ▶ Limits of Big Data
- ▶ Introduction to social media data analysis

Tomorrow

- ▶ Automated classification of social media text

Wednesday

- ▶ Discovery in large-scale social media data

Course philosophy

How to learn the techniques in this course?

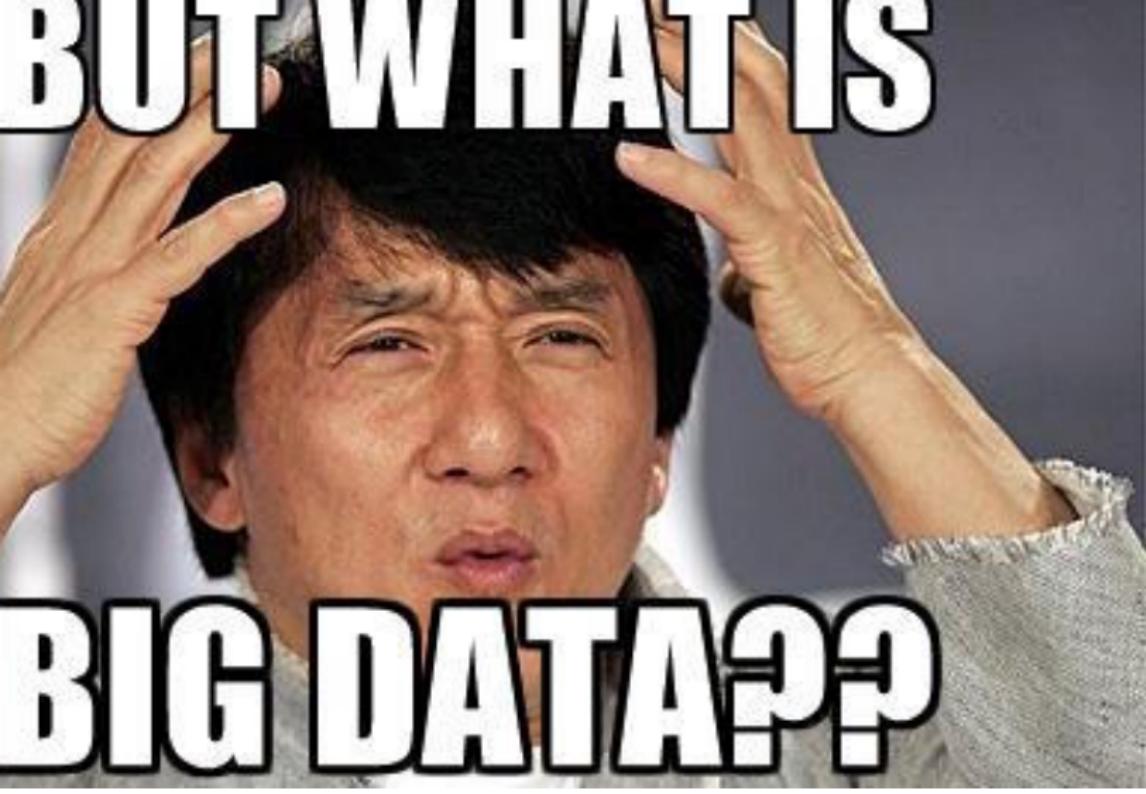
- ▶ Lecture approach: not ideal for learning how to code
- ▶ You can only **learn by doing**.
- We will cover each concept three times during each session
 1. Introduction to the topic (30 minutes)
 2. Guided coding session (30 minutes)
 3. Coding challenges (30 minutes)
 - Repeat twice per day
- ▶ You're encouraged to continue working on the coding challenges after class. Solutions will be posted the following day.
- ▶ Warning! We will **move fast**.

Your turn!



1. Name?
2. Affiliation? Background?
3. Summarize your research interests in 5 words

Social Media & Big Data Research: Opportunities and Challenges



BUT WHAT IS

BIG DATA??

The Three V's of Big Data

Dumbill (2012), Monroe (2013):

1. **Volume**: 6 billion mobile phones, 1+ billion Facebook users, 500+ million tweets per day...
2. **Velocity**: personal, spatial and temporal granularity.
3. **Variability**: images, networks, long and short text, geographic coordinates, streaming...

Big data: data that are so large, complex, and/or variable that the tools required to understand them must first be invented.

Computational Social Science

"We have [life in the network](#). We check our emails regularly, make mobile phone calls from almost any location ... make purchases with credit cards ... [and] maintain friendships through online social networks ... These transactions leave digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations and societies".

Lazer et al (2009) Science

Digital trace data

What are the main advantages of using social media data to study human behavior?

1. Unobtrusive data collection at scale, e.g. in study of networks, censorship
2. Homogeneity in data format across actors, countries, and over time, e.g. in study of political rhetoric
3. Temporal and spatial data granularity, e.g. in study of geographic segregation
4. Increasing representativeness of social media users, e.g. in study of political elites

Social media research

Two different approaches in the growing field of social media research:

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

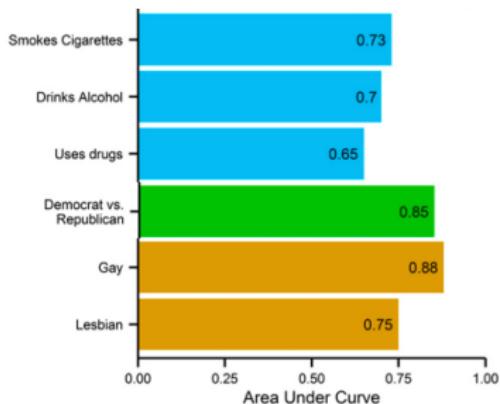
Social media research

Two different approaches in the growing field of social media research:

1. Social media as a new source of information
 - ▶ Behavior, opinions, and latent traits
 - ▶ Interpersonal networks
 - ▶ Elite behavior
 - ▶ Affordable field experiments
2. How social media affects 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, ...
- Non-intrusive measurement of behavior and public opinion
Beauchamp (AJPS 2016): “Predicting and Interpolating State-level Polls using Twitter Textual Data”
- Inference of latent traits: political knowledge, ideology, personal traits, socially undesirable behavior, ...

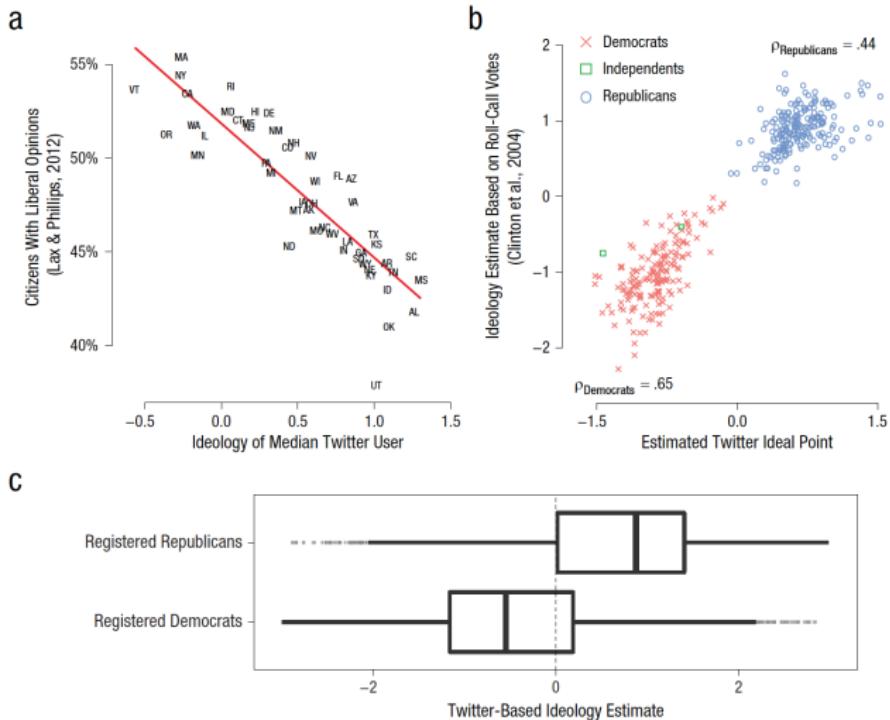


Kosinski et al, 2013, “Private traits and attributes are predictable from digital records of human behavior”, PNAS (also personality, PNAS 2015)

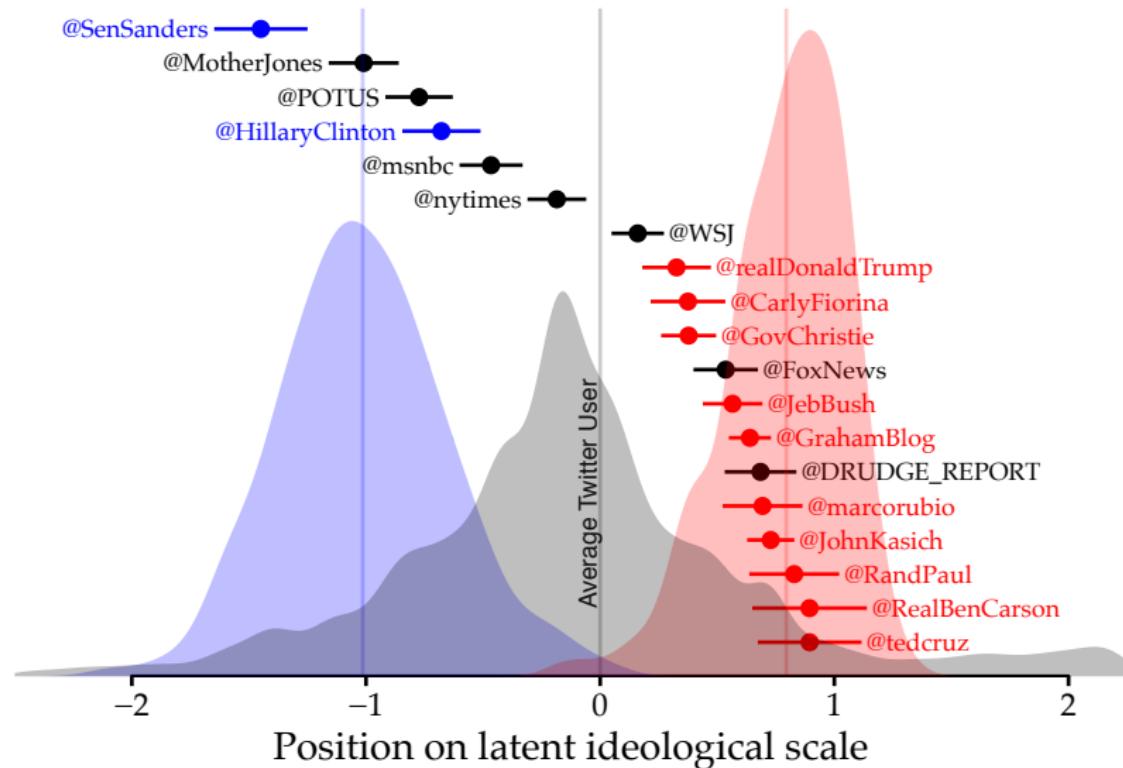
Fig. 2. Prediction accuracy of classification for dichotomous/dichotomized attributes expressed by the AUC.

Behavior, opinions, and latent traits

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



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

Social media research

Two different approaches in the growing field of social media research:

1. Social media as a new source of information
 - ▶ Behavior, opinions, and latent traits
 - ▶ **Interpersonal networks**
 - ▶ Elite behavior
 - ▶ Affordable field experiments
2. How social media affects 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

Today is Election Day [What's this?](#) • [close](#)

 Find your polling place on the U.S. Politics Page and click the "I Voted" button to tell your friends you voted.

I Voted


0 1 1 5 5 3 7 6
People on Facebook Voted

 **f** Jaime Settle, Jason Jones, and 18 other friends have voted.

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. Settle², Robert M. Bond², Christopher J. Fariss², Cameron Marlow³,

bioRxiv preprint doi: [https://doi.org/10.1101/12](#)

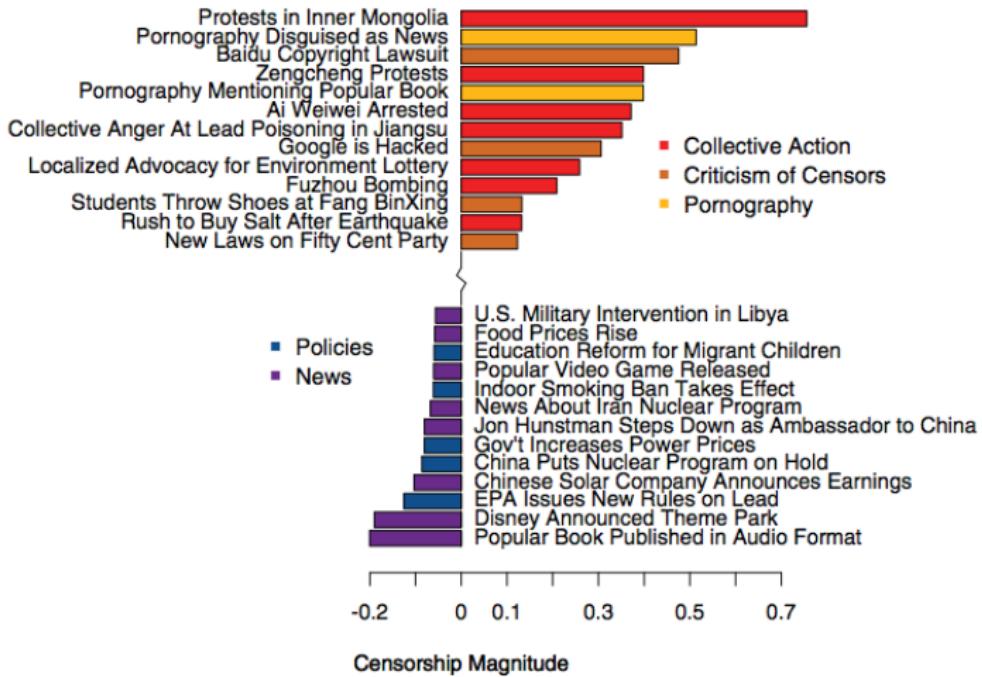
Social media research

Two different approaches in the growing field of social media research:

1. Social media as a new source of information
 - ▶ Behavior, opinions, and latent traits
 - ▶ Interpersonal networks
 - ▶ Elite behavior
 - ▶ Affordable field experiments
2. How social media affects 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

Social media research

Two different approaches in the growing field of social media research:

1. Social media as a new source of information
 - ▶ Behavior, opinions, and latent traits
 - ▶ Interpersonal networks
 - ▶ Elite behavior
 - ▶ Affordable field experiments
2. How social media affects 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

 2.7k Shares

 12k Downloads

 3 Citations

Social media research

Two different approaches in the growing field of social media research:

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





#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

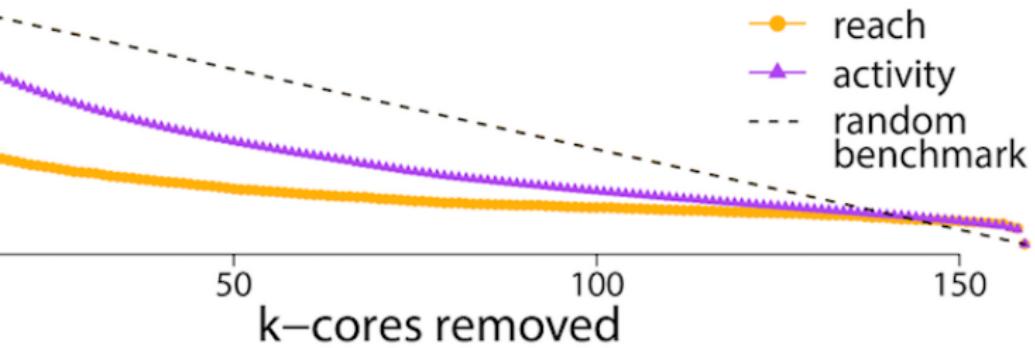


RESEARCH ARTICLE

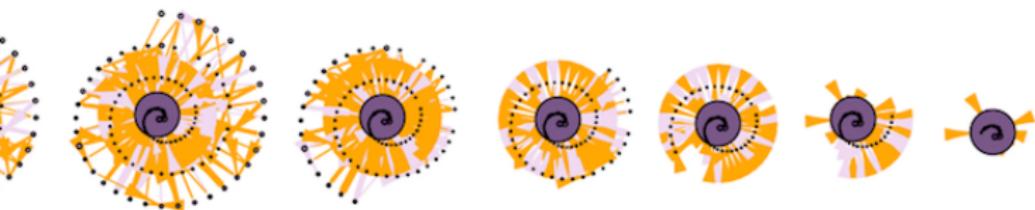
The Critical Periphery in the Growth of Social Protests

Pablo Barberá^{1*}, Ning Wang², Richard Bonneau^{3,4}, John T. Jost^{1,5,6}, Jonathan Nagler⁶, Joshua Tucker⁶, 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)



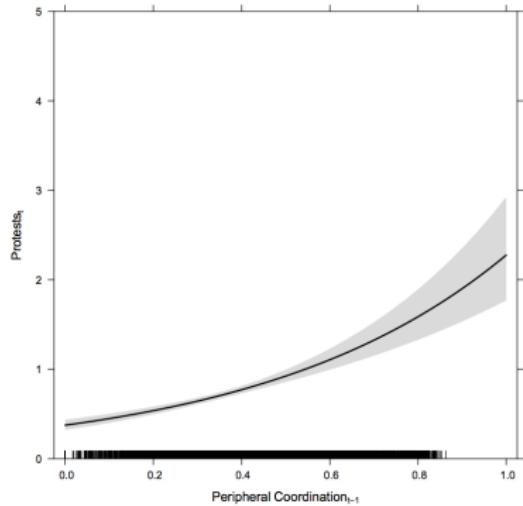
k -cores removed



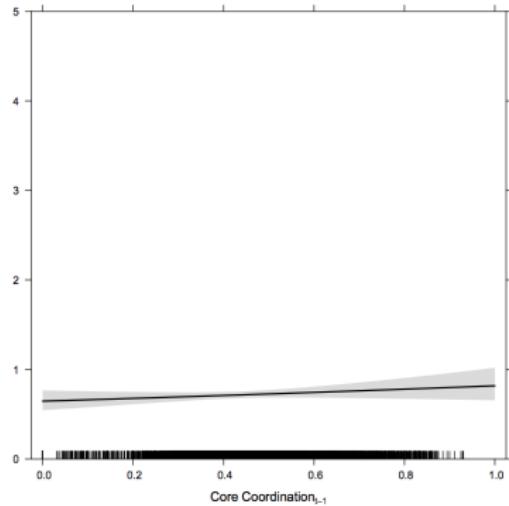
aggregate size of participants' audience

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 research

Two different approaches in the growing field of social media research:

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



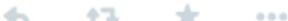
Barack Obama

@BarackObama



Follow

Four more years.



RETWEETS

756,411

FAVORITES

288,867



11:16 PM - 6 Nov 2012

Sections ≡

The Washington Post

Search



Sign In

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

Social media research

Two different approaches in the growing field of social media research:

1. Social media as a new source of information
 - ▶ Behavior, opinions, and latent traits
 - ▶ Interpersonal networks
 - ▶ Elite behavior
 - ▶ Affordable field experiments
2. How social media affects 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

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

American Politics Research

1–31

© The Author(s) 2014

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/1532673X14557942

apr.sagepub.com



**Javier Sajuria¹, Jennifer vanHeerde-Hudson¹,
David Hudson¹, Niheer Dasandi¹, and Yannis
Theocharis²**

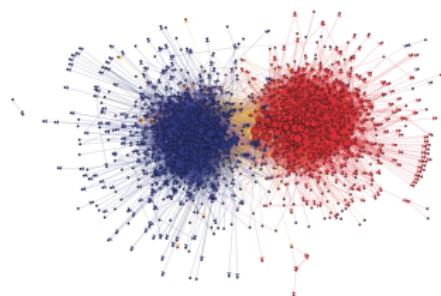
Social media research

Two different approaches in the growing field of social media research:

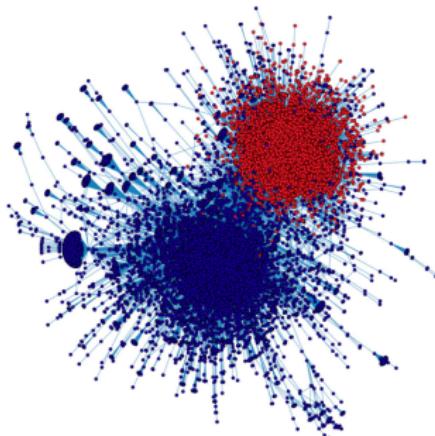
1. Social media as a new source of information
 - ▶ Behavior, opinions, and latent traits
 - ▶ Interpersonal networks
 - ▶ Elite behavior
 - ▶ Affordable field experiments
2. How social media affects social behavior
 - ▶ Collective action and social movements
 - ▶ Political campaigns
 - ▶ Social capital and interpersonal communication
 - ▶ **Political attitudes and behavior**

Social media as echo chambers?

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



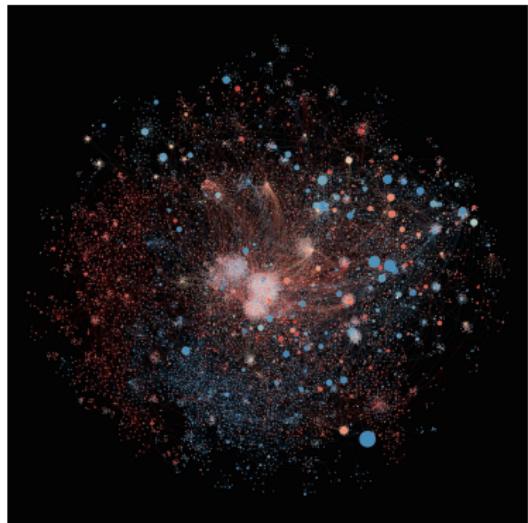
Adamic and Glance (2005)



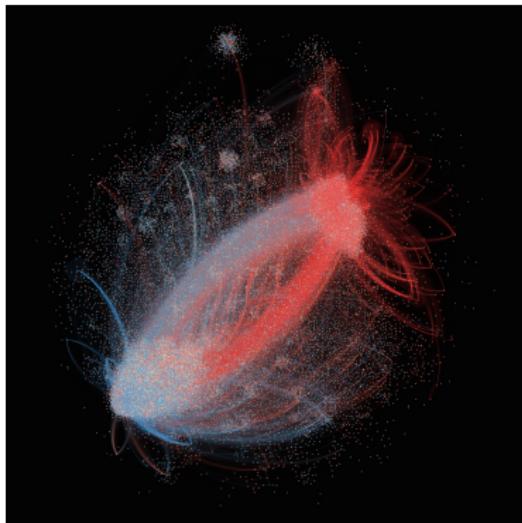
Conover et al (2012)

- ▶ ...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



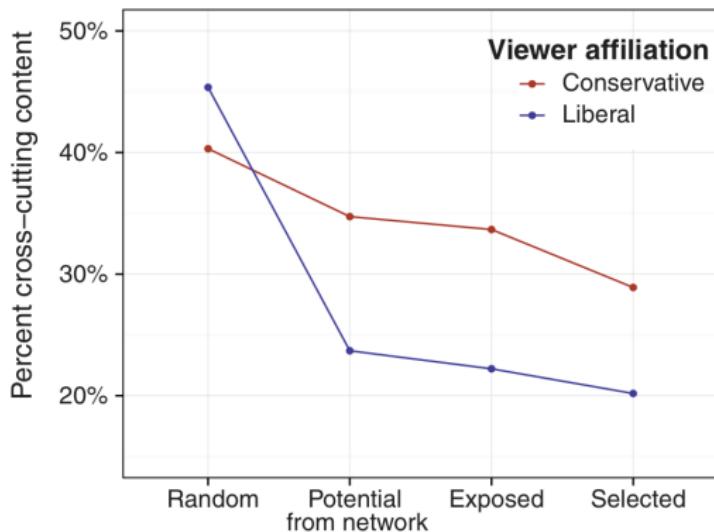
2012 Election

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

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).

B



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

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

Social media research

Two different approaches in the growing field of social media research:

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

Analyzing Social Media Data: First Steps

Why we're using R

- ▶ Becoming *lingua franca* of statistical analysis in academia
- ▶ What employers in private sector demand
- ▶ It's free and open-source
- ▶ Flexible and extensible through *packages* (over 10,000 and counting!)
- ▶ Powerful tool to conduct automated text analysis, social network analysis, and data visualization, with packages such as quanteda, igraph or ggplot2.
- ▶ Command-line interface and scripts favors reproducibility.
- ▶ Excellent documentation and online help resources.

R is also a full programming language; once you understand how to use it, you can learn other languages too.

RStudio Server

RStudio interface showing a script, workspace, and plot.

Script Editor:

```
1 library(ggplot2)
2 source("plots/formatPlot.R")
3
4 view(diamonds)
5 summary(diamonds)
6
7 summary(diamonds$price)
8 aveSize <- round(mean(diamonds$carat), 4)
9 clarity <- levels(diamonds$clarity)
10
11 p <- qplot(carat, price,
12           data=diamonds, color=clarity,
13           xlab="Carat", ylab="Price",
14           main="Diamond Pricing")
15
```

Console:

```
Min. : 0.000   Min. : 0.000   Min. : 0.000
1st Qu.: 4.710   1st Qu.: 4.720   1st Qu.: 2.910
Median : 5.700   Median : 5.710   Median : 3.530
Mean   : 5.731   Mean   : 5.735   Mean   : 3.539
3rd Qu.: 6.540   3rd Qu.: 6.540   3rd Qu.: 4.040
Max.  :10.740   Max.  :58.900   Max.  :31.800
> summary(diamonds$price)
   Min. 1st Qu. Median  Mean 3rd Qu.  Max.
   326    950   2401   3933   5324  18820
> aveSize <- round(mean(diamonds$carat), 4)
> clarity <- levels(diamonds$clarity)
> p <- qplot(carat, price,
+            data=diamonds, color=clarity,
+            xlab="Carat", ylab="Price",
+            main="Diamond Pricing")
>
> format.plot(p, size=24)
> |
```

Workspace:

- Data: diamonds (53940 obs. of 10 variables)
- Values: aveSize (0.7979)
- clarity (character [8])
- p (ggplot [8])

Functions:

- format.plot(plot, size)

Plots:

Diamond Pricing

The plot displays a strong positive linear trend, indicating that larger diamonds tend to have higher prices. The color coding by clarity grade (I1, SI2, SI1, VS2, VS1, VVS2, VVS1, IF) shows that clarity has a significant impact on price, with higher clarity diamonds (lower numbers) generally having higher prices.

Course website

[pablobarbera / big-data-upf](#) Private

Unwatch 1 Star 0 Fork 0

Code Issues 0 Pull requests 0 Projects 0 Wiki Settings Insights

RECSM-UPF Summer School: Social Media and Big Data Research Edit

Add topics

2 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

		Latest commit 9c73b88 10 hours ago
	pablobarbera day 1 updates	
	00-setup.Rmd	day 1 updates 10 hours ago
	01a-scraping-tables.Rmd	initial commit 14 hours ago
	01b-scraping-unstructured-data.Rmd	initial commit 14 hours ago
	01c-apis.Rmd	day 1 updates 10 hours ago
	README.md	initial commit 14 hours ago
	packages.r	day 1 updates 10 hours ago
	test.r	day 1 updates 10 hours ago

README.md

Summer School: Social Media and Big Data Research

Sponsored by

- Barcelona Summer School in Survey Methodology

pablobarbera.com/social-media-upf

RECSM Summer School: Social Media and Big Data Research

Pablo Barberá

London School of Economics

www.pablobarbera.com

Course website:

pablobarbera.com/social-media-upf