Birds of the same feather tweet together?
What social media data reveals about political behavior.

Pablo Barberá
Center for Data Science
New York University
www.pablobarbera.com

#iwsgrcp16

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

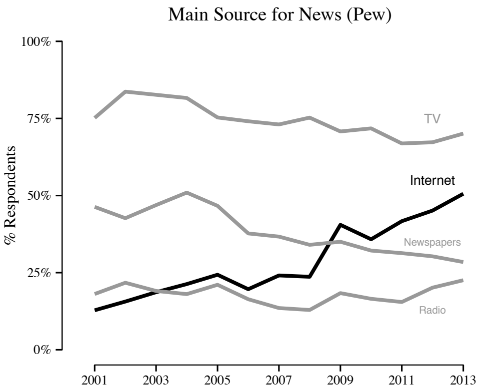
Jackie Walorski @RepWalorski
Follow

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

Sources of Political Information

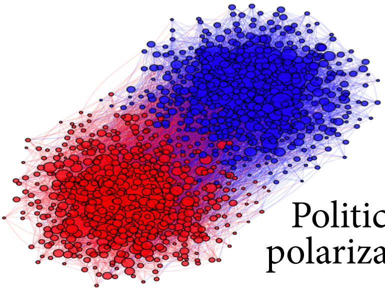


Data: Pew Research Center. Respondents were allowed to name up to two sources.

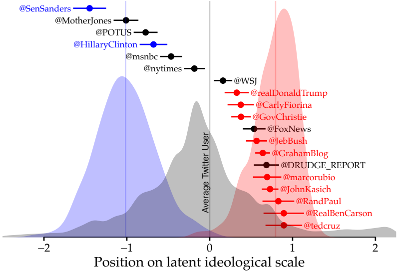
- 41% of Americans see news on social media every day (Pew)
- 27% of online EU citizens use social media to get news on national political matters (Eurobarometer, Fall 2012)
- Social media: top source of news for U.S. young adults (Pew)

Shift in communication patterns

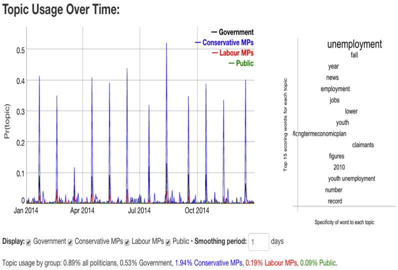
What social media data reveals about...



Political polarization



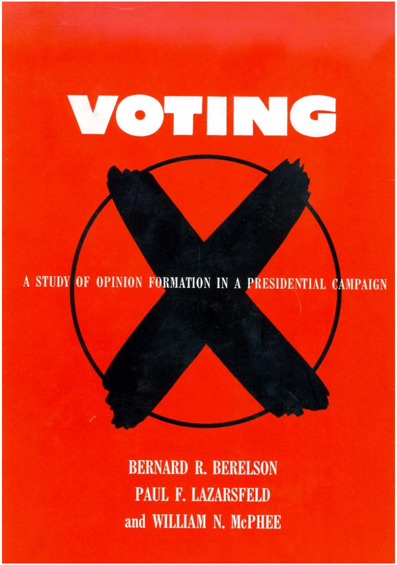
Latent individual traits



Issue salience

Political behavior is social

- Opinion formation as a *social process* (Berelson et al, 1954)



Political behavior is social

- Opinion formation as a *social process* (Berelson et al, 1954)
- *Voting is contagious* (Nickerson, 2008)

American Political Science Review Vol. 102, No. 1 February 2008

DOI: 10.1017/S0003055408080039

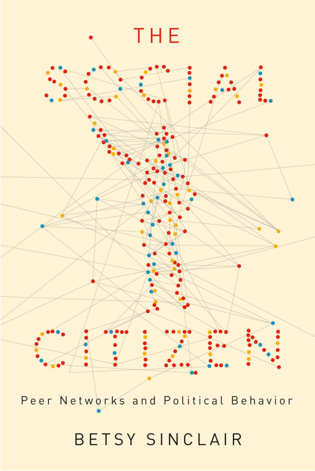
Is Voting Contagious? Evidence from Two Field Experiments

DAVID W. NICKERSON *University of Notre Dame*

Members of the same household share similar voting behaviors on average, but how much of this correlation can be attributed to the behavior of the other person in the household? Disentangling and isolating the unique effects of peer behavior, selection processes, and congruent interests is a challenge for all studies of interpersonal influence. This study proposes and utilizes a carefully designed placebo-controlled experimental protocol to overcome this identification problem. During a face-to-face canvassing experiment targeting households with two registered voters, residents who answered the door were exposed to either a Get Out the Vote message (treatment) or a recycling pitch (placebo). The turnout of the person in the household not answering the door allows for contagion to be measured. Both experiments find that 60% of the propensity to vote is passed onto the other member of the household. This finding suggests a mechanism by which civic participation norms are adopted and couples grow more similar over time.

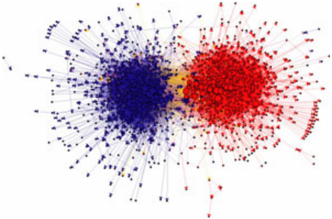
Political behavior is social

- Opinion formation as a *social process* (Berelson et al, 1954)
- *Voting is contagious* (Nickerson, 2008)
- The *social citizen* (Sinclair, 2012)

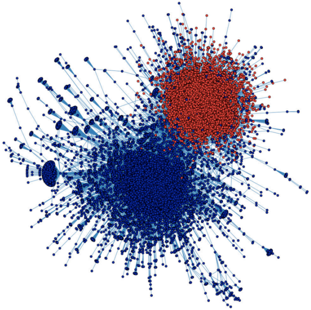


Social media as echo chambers?

- communities of like-minded individuals (homophily)



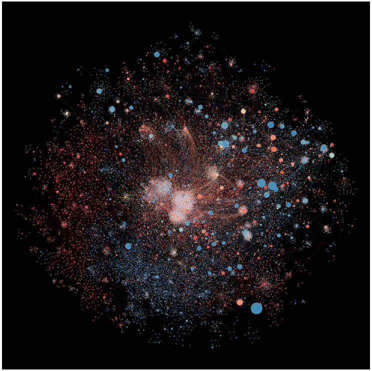
Adamic and Glance (2005)



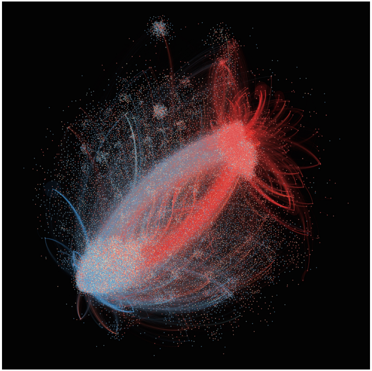
Conover et al (2012)

- ...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

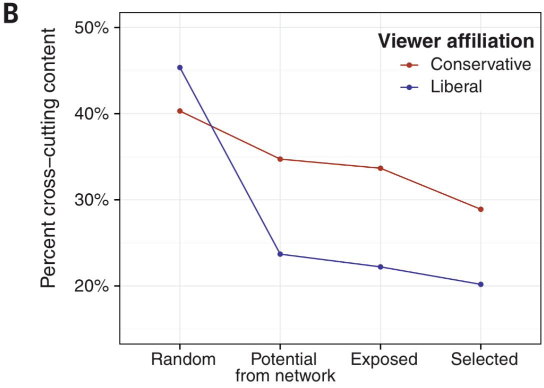


2012 Election

Barberá, Jost, Nagler, Tucker, & Bonneau (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).



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

Beyond the echo chamber

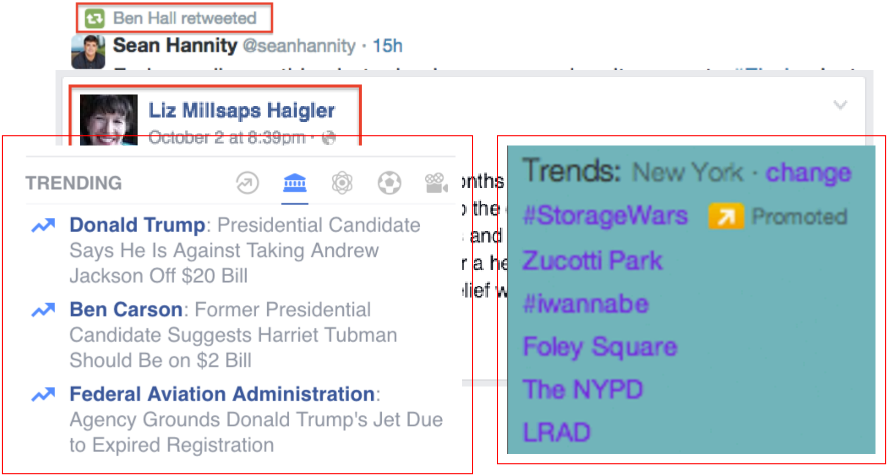
Social media usage induces political moderation

1. Inadvertent exposure to political messages
    - “Your friends deliver the news” (Adamic)

Inadvertent Exposure

Inadvertent Exposure

Inadvertent Exposure



Beyond the echo chamber

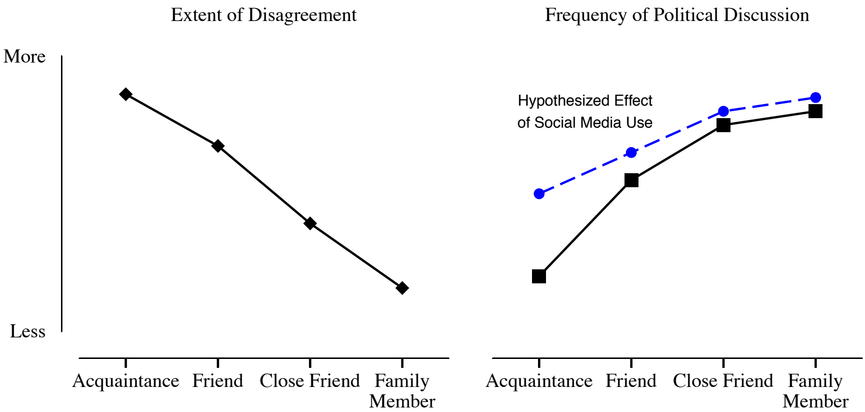
Social media usage induces political moderation

Why?

1. Inadvertent exposure to political messages
   - "Your friends deliver the news" (Adamic)
   - Less selective exposure

2. More frequent interactions beyond immediate personal network
   - "The strength of weak ties" in providing novel information (Granovetter, 1973; Bakshy et al, 2012)

The strength of weak ties



Source: Mutz (2006)

- Extent of disagreement with weak ties is greater.
- Bakshy et al (2012): weak ties are collectively more influential on social media

Beyond the echo chamber

Social media usage induces political moderation

Why?

1. Inadvertent exposure to political messages
   - "Your friends deliver the news" (Adamic)
   - Less selective exposure

2. More frequent interactions beyond immediate personal network
   - "The strength of weak ties" in providing novel information (Granovetter, 1973; Bakshy et al, 2012)

...increases exposure to dissonant views

...and therefore mass political polarization decreases.

Research Design

Question: do individuals exposed to diverse social media networks *become* moderate over time?

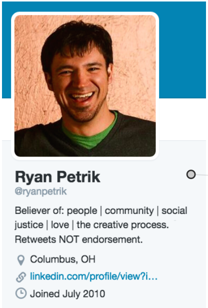
Outcome variable: change in ideological positions

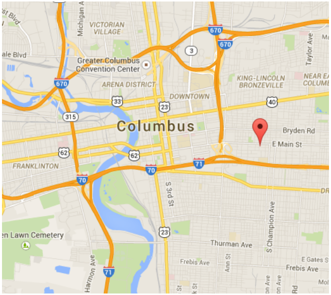
Independent variable: exposure to dissonant opinions

Focus on Twitter:

- Individuals "follow" other users (directed links)
  - → Following political accounts is informative about ideology
  - → Observe personal network (excluding political accounts) to measure potential exposure to dissonant political messages
- Networks are dynamic: we can observe change
  - → Exploit panel structure of dataset to identify causal effects
- Most accounts are public and use real names
  - → Possible to match with individual voting records

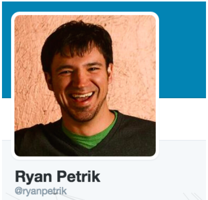
Concerns about representativeness.

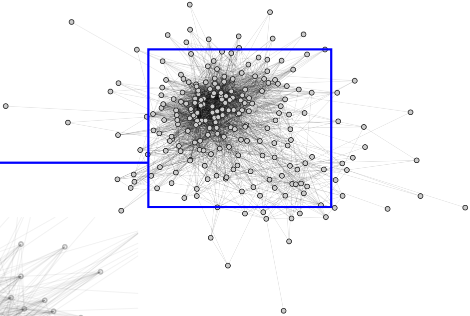


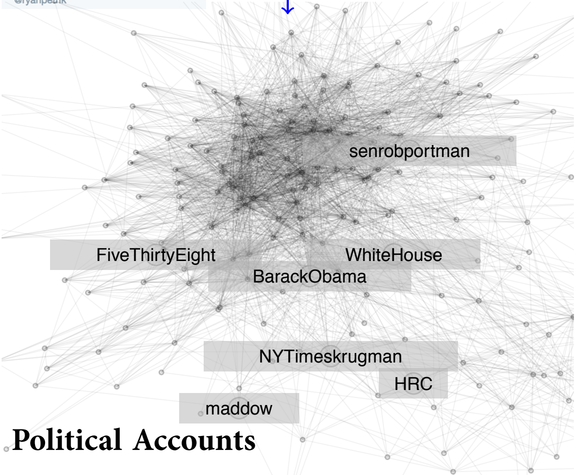


Voter file records:

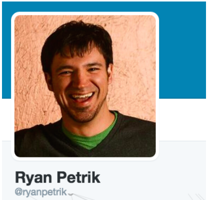
Voted in 2012 Presidential election
Affiliated as a Democrat
Age 29; Lives in Franklin county

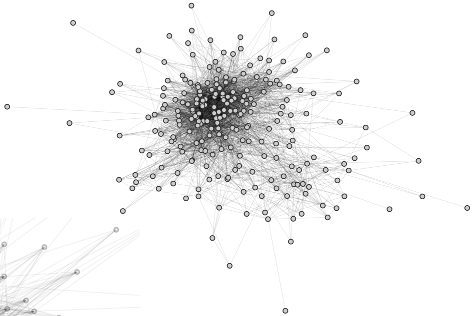


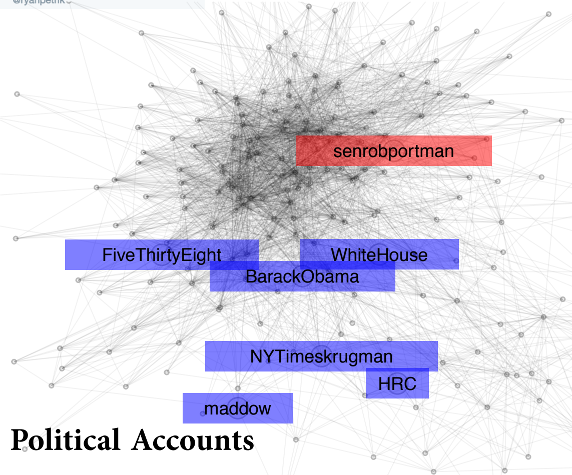




Political Accounts



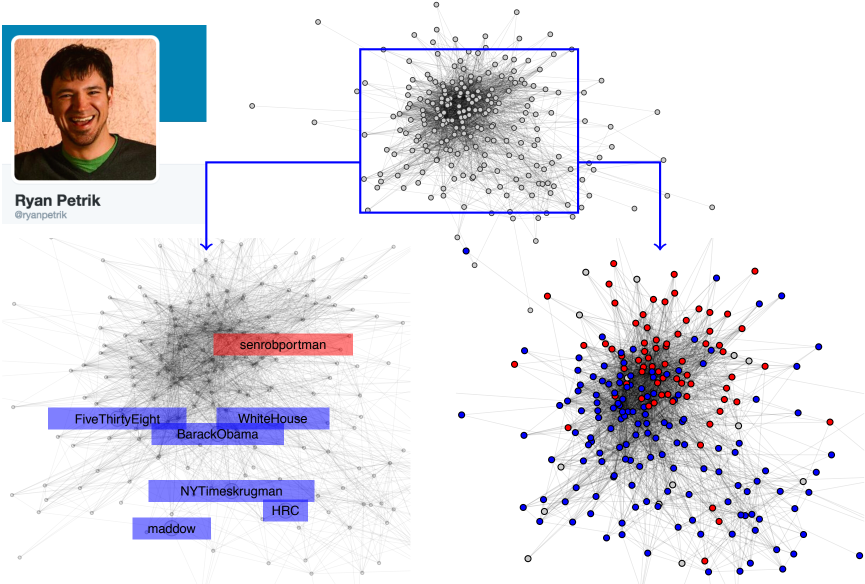




Political Accounts

	BarackObama	WhiteHouse	GOP	maddow	FoxNews	HRC	…	pol. account m
ryanpetrik	1	1	0	1	0	1	…	
user 2	0	0	1	0	1	0	…	
user 3	0	0	1	0	1	0	…	
user 4	1	1	0	0	0	1	…	
user 5	0	1	0	0	0	1	…	
…								
user n	0	1	1	0	0	0	…	

Estimated ideology: $\theta_i = -1.05$



Political Accounts
Estimated ideology: $\theta_i = -1.05$

Social Ties ("friends")
32% Conservative

Data

- m = list of 620 popular political accounts in U.S.
  - $\rightarrow$ Legislators, president, candidates, other political figures, media outlets, journalists, interest groups…
- n = followers of at least one of these accounts
  - $\rightarrow$ 30.8M users ($\sim$75% of U.S. users)
- t = January 2013 and July 2014

Observing Communication Networks

Personal networks

- List of users each individual follows (social ties)
- Political accounts and *verified* users are excluded

  - Ensure independence wrt estimation of ideology
  - Focus on information shared by social ties

Sample:

- 75K active users in the U.S. matched with voter files

Exposure to Dissonant Opinions

Index of exposure to disagreement for user i:

$$\frac{u_{iC}}{u_{iC}+u_{iL}} \text{ if user } i \text{ is liberal}$$
$$\frac{u_{iL}}{u_{iC}+u_{iL}} \text{ if user } i \text{ is conservative}$$

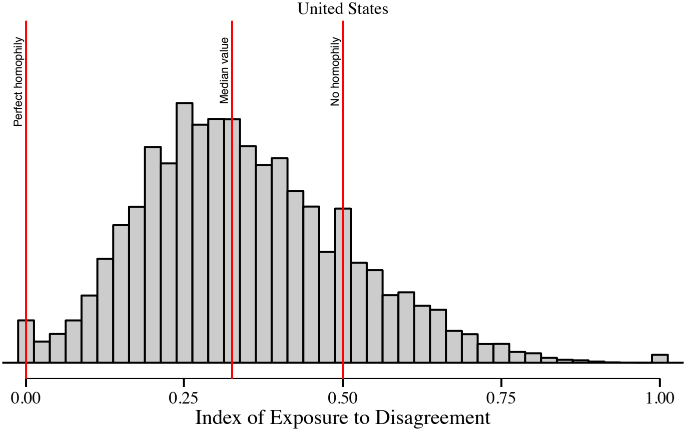
where user i is liberal if $\theta_i < 0$ and conservative if $\theta > 0$

and u_{iC} (u_{iL}) is the count of conservative (liberal) users in user i's personal network

→ Measure of the proportion of individuals in a user's network that disagree with her ideological position

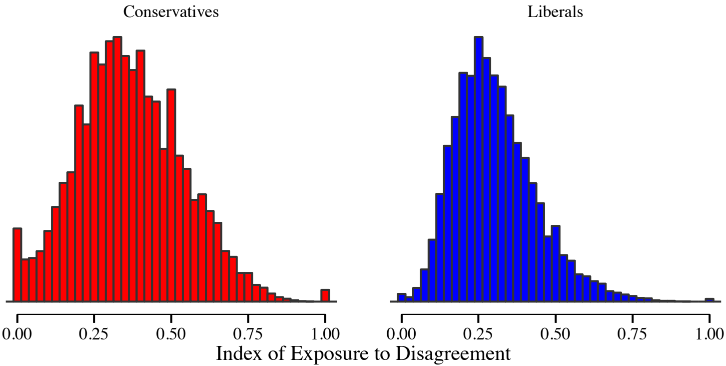
Exposure to Dissonant Opinions

Most Twitter users are exposed to dissonant opinions



Exposure to Dissonant Opinions

Conservatives are more exposed to dissonant opinions (on average)



Findings

Do social media users exposed to dissonant opinions tend to become more politically moderate over time?

Panel design:

- $\theta_{i,t=2013}$ and $\theta_{i,t=2014}$: ideology estimates in 2013 and 2014
- D_i: index of exposure to disagreement in 2013
- Causal identification

Regression model:

$$-(|\hat{\theta}_{i,t=2014}| - |\hat{\theta}_{i,t=2013}|) = \psi_0 + \psi_1 D_i + \mathbf{X}\xi + \epsilon_i$$

Control variables, **X**:

- Network controls: network size, political interest, activity level, number of followers.
- Offline behavior controls: turnout, party affiliation, age, state-fixed effects

Findings

Table: OLS Regressions of Change in Political Moderation on Exposure to Disagreement in 2013

	United States	
Exposure to Disagreement	0.20*	0.20*
	(0.00)	(0.00)
Intercept	0.21*	0.20*
	(0.01)	(0.01)
Network controls	✓	✓
Offline controls		✓
N	74,515	74,515
R^2	0.09	0.09
Resid. sd	0.23	0.23

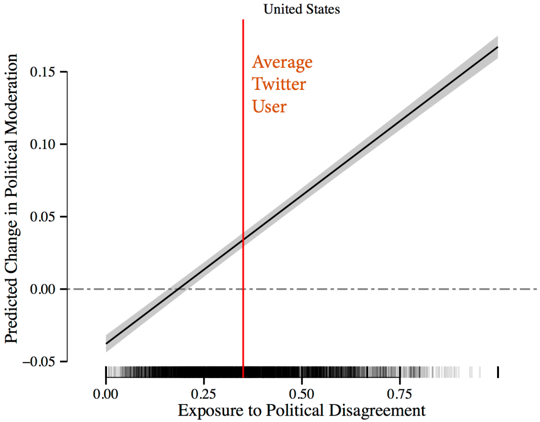
Note: * significant at $p < 0.05$. Standard errors in parentheses.

Network controls: network size, political interest, activity level, number of followers.

Offline controls: turnout, party affiliation, age, state-fixed effects

Findings

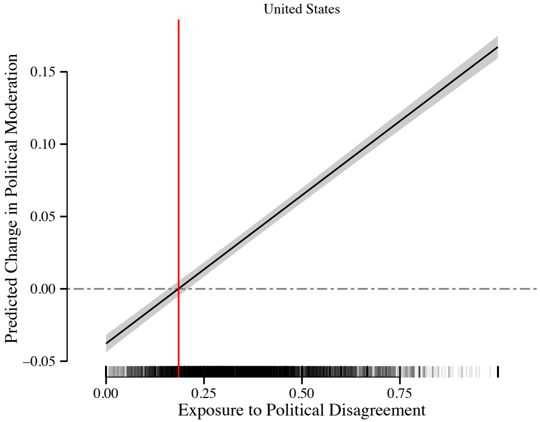
Predicted change in moderation (2013 to 2014) for average individual, conditional on exposure to disagreement



Social media users exposed to dissonant opinions tend to become more politically moderate over time

Findings

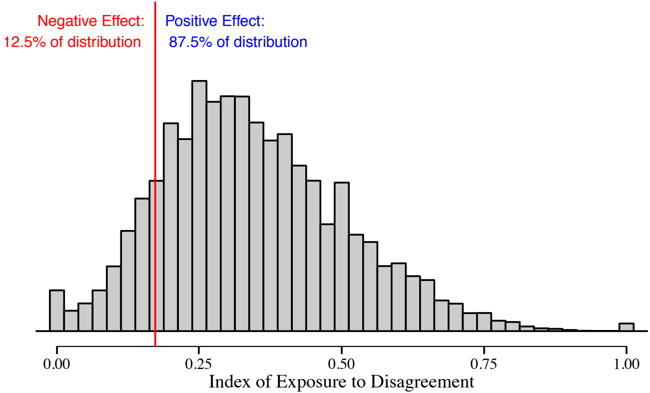
Predicted change in moderation (2013 to 2014) for average individual, conditional on exposure to disagreement



Social media users exposed to dissonant opinions tend to become more politically moderate over time

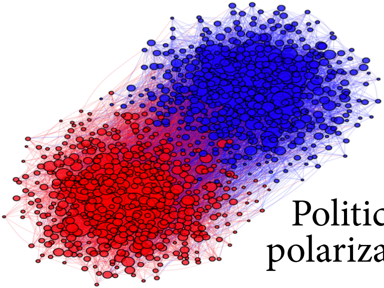
Findings

Predicted change in moderation (2013 to 2014) for average individual, conditional on exposure to disagreement

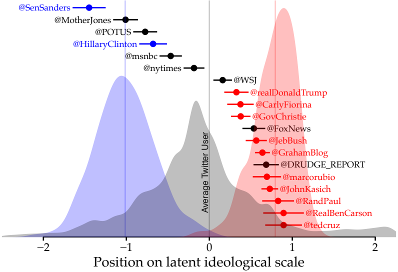


Social media users exposed to dissonant opinions tend to become more politically moderate over time

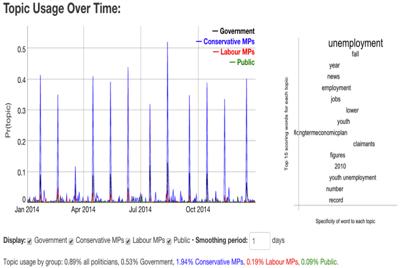
What social media data reveals about...



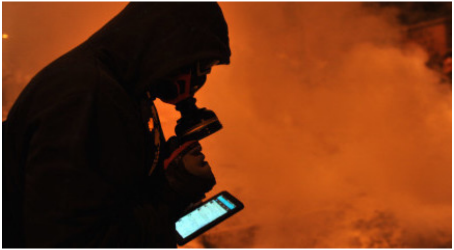
Political polarization



Latent individual traits



Issue salience



#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.

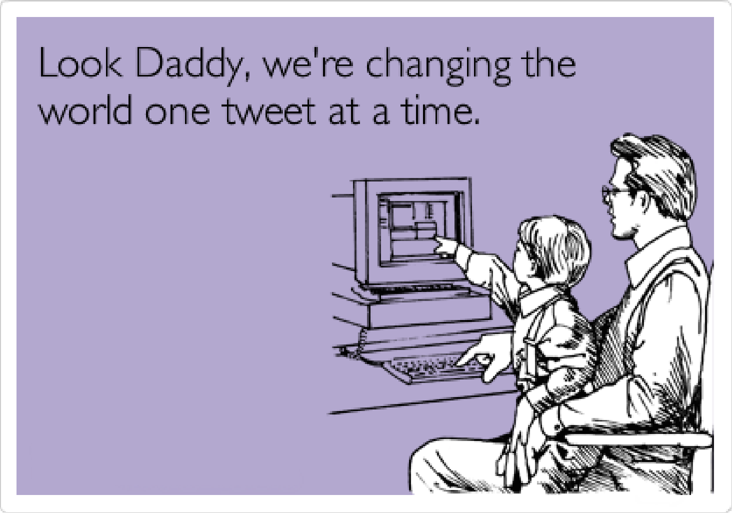
***Gladwell**, Small Change (New Yorker)*

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*

our argument



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 contribution: key role of peripheral participants
  1. Increase reach of protest messages (positional effect)
  2. Large contribution to overall activity (size effect)

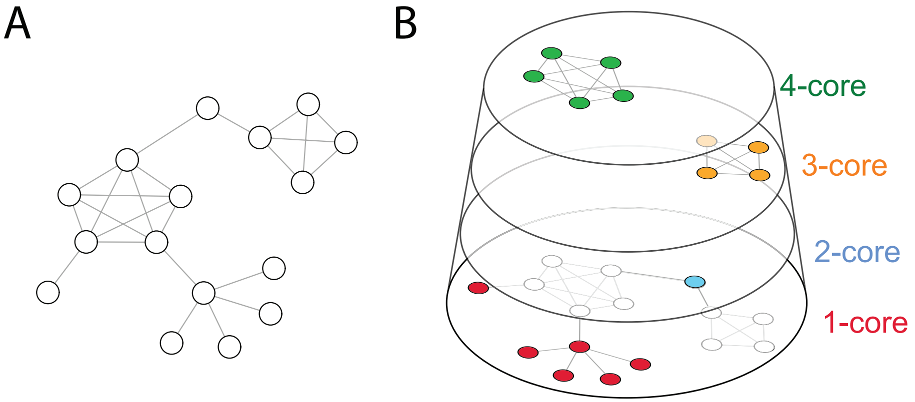
related work

1. Collective action
   - Resource mobilization theory (Jenkins 1983)
     - → But how does a spark turn into a protest wildfire? (Biggs, 2005)
   - Critical mass theory
     - Granovetter 1978, Marwell and Oliver 1993, Schelling 1978
     - Interdependent decisions; feedback mechanisms
     - → Mostly simulations and mathematical models
     - → Effect of core-periphery structure on attaining critical mass has been disregarded
2. Diffusion of innovation
   - Emphasis on early adopters, opinion leaders, social influence bias
   - → But both influence and susceptibility drive contagion (Aral 2012)

network hierarchy

- **Motivation**
  - Analysis of hierarchical properties of large scale networks
- **Network core:**
  1. *Centrality*: high relative importance in network
  2. *Connectivity*: many possible distinct paths between individuals (not captured by simple topological measures)
- **k-core decomposition**
  - Algorithm to partition a network in nested shells of connectivity
  - The k-core of a graph is the maximal subgraph in which every node has at least degree k
  - Many applications; scales well to large networks: $\mathcal{O}(n + e)$

k-core decomposition



case selection

1. Gezi Park protests (Turkey)
    - May–June 2013
    - 30,019,710 tweets sent by 2,908,926 users
2. "United for Global Change"
    a) Occupy Wall Street
    b) Indignados Movement (Spain)
    - April–May 2012
    - 606,625 tweets sent by 125,219 users
3. "Placebo" networks:
    a) Oscars, March 2014
    b) Discussions about minimum wage, 2014

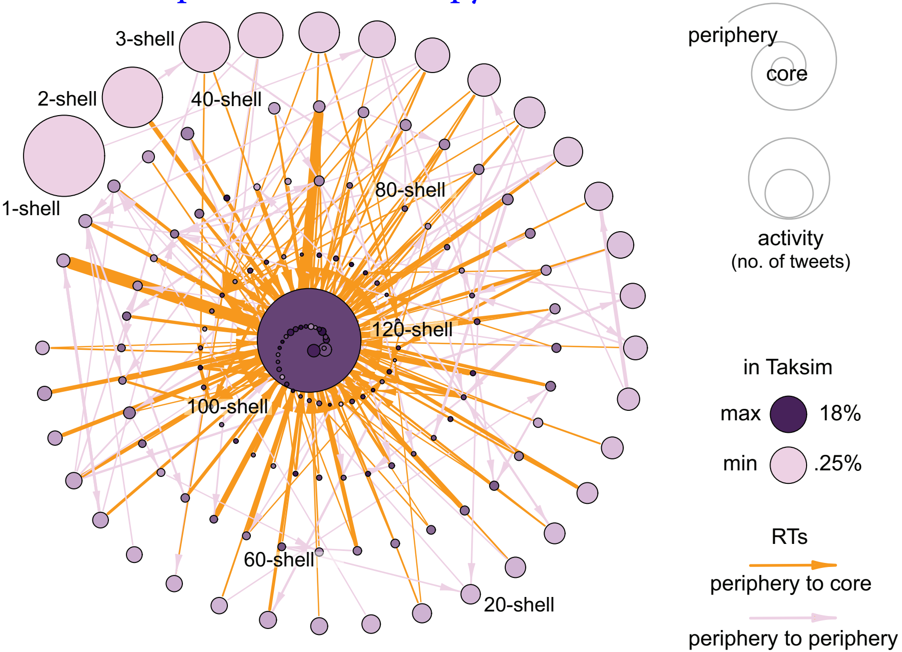
summary statistics

Table: Summary statistics for five retweet networks (largest weakly connected component)

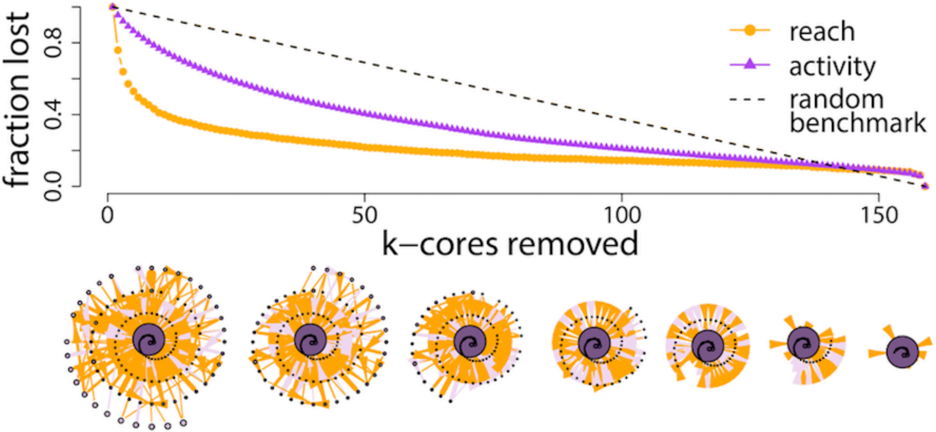
	Gezi	Occupy	Indignados
Nodes	1,935,911	30,708	49,534
Edges	15,761,311	80,967	124,519
Max indegree	181,387	2,092	3,898
Clustering	0.091	0.147	0.125

	Oscars	Min. wage
Nodes	2,800,880	721,660
Edges	3,925,396	1,310,384
Max indegree	918,968	96,669
Clustering	0.066	0.094

k-core decomposition of #OccupyGezi network



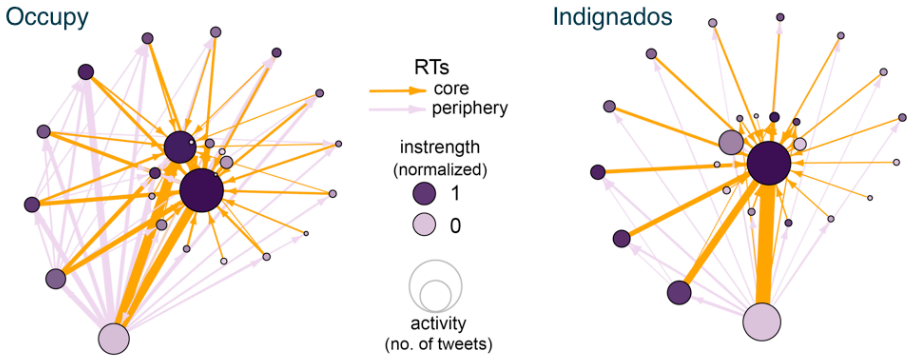
Relative importance of core and periphery



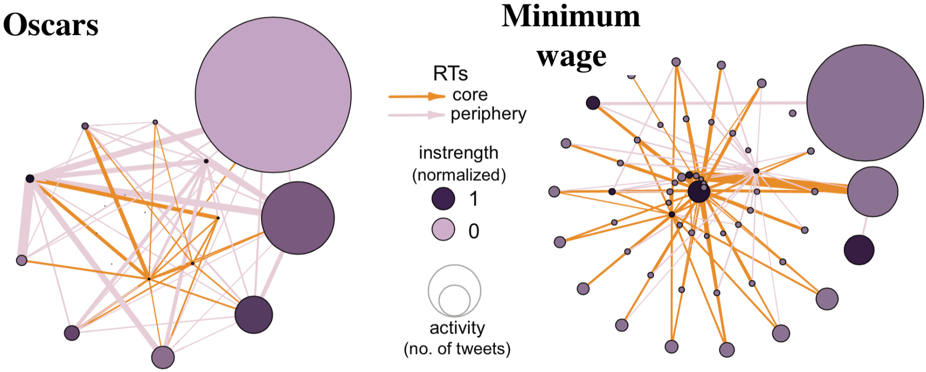
reach: aggregate size of participants' audience

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

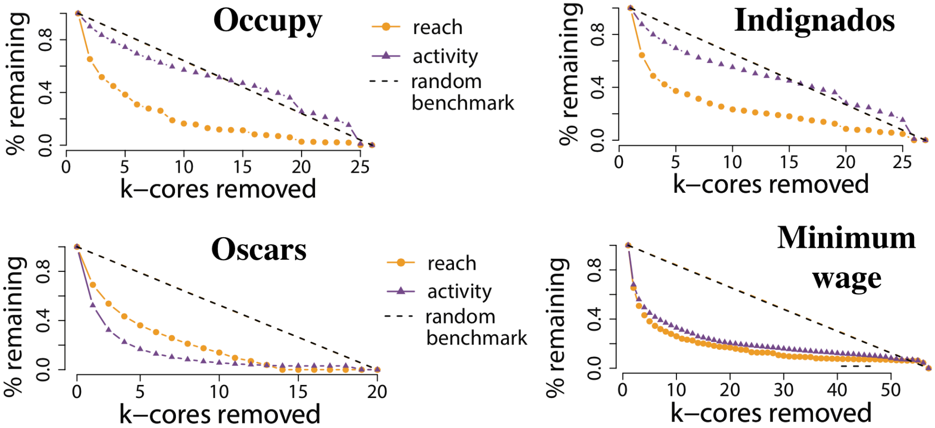
k-core decomposition of Occupy & Indignados networks



k-core decomposition of Oscars and Min.Wage networks



Relative importance of core and periphery



reach: aggregate size of participants' audience

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

conclusions

- "Slacktivists" are crucially important as a collective:
  1. Amplify visibility of high-risk activism
  2. Generate content at levels comparable to core

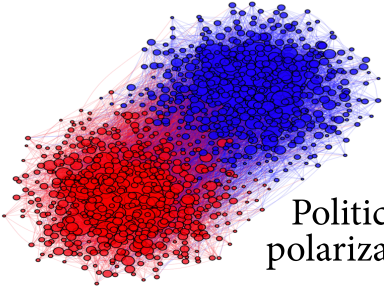
  → **BOTH** core and periphery are necessary!

- Ongoing work:
  - Influence of mass media censorship or lack of coverage
  - Categorize protest networks
  - Other cases: Black Lives Matter, 2016 U.S. presidential campaign, Egypt, Venezuela, Ukraine...

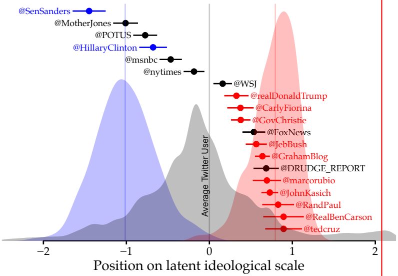
- Implications for study of cascading behavior and more general studies of epidemic behavior:

  → *both* core and periphery explain success of diffusion

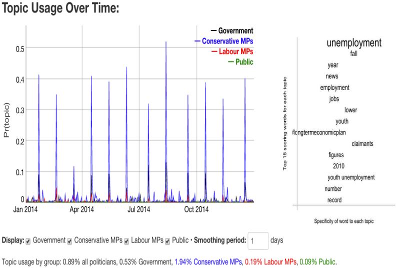
What social media data reveals about...



Political polarization



Latent individual traits



Issue salience

Behavior, opinions, and latent traits

- Digital footprint: check-ins, conversations, geolocated pictures, likes, shares, retweets, …

→ Non-intrusive measurement of behavior and opinion

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

Behavior, opinions, and latent traits

- Digital footprint: check-ins, conversations, geolocated pictures, likes, shares, retweets, …

→ Non-intrusive measurement of behavior and opinion

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 platfo workings and make accommoda 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*

Behavior, opinions, and latent traits

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

→ Non-intrusive measurement of behavior and opinion

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

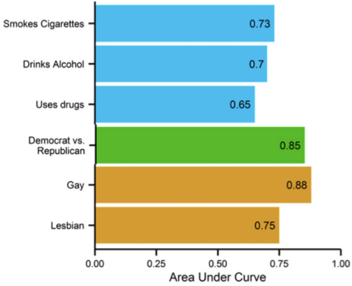


Fig. 2. Prediction accuracy of classification for dichotomous/dichotomized attributes expressed by the AUC.

Kosinki et al, 2013, "Private traits and attributes are predictable from digital records of human behavior", *PNAS* (also personality, PNAS 2015)

Estimating ideology with Twitter networks

- **Assumption:** individuals prefer to *follow* political accounts they perceive to be ideologically close.
- Data: “following” decisions, a matrix of binary choices ($\mathbf{Y_{ij}}$).
- Spatial following model (Barberá, 2014, *Political Analysis*): Probability that user i follows political account j in period t is

$$P(y_{ijt} = 1) = \text{logit}^{-1}\left(\alpha_j + \beta_i - \gamma(\theta_{it} - \phi_j)^2\right) \ ,$$

- with latent variables:

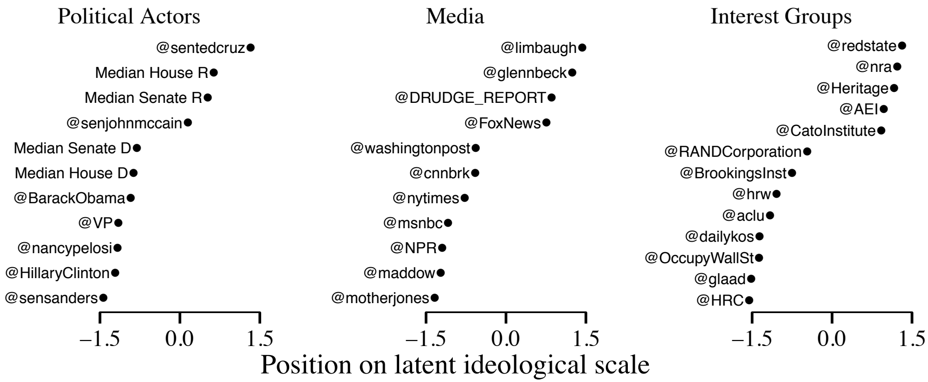
  θ_{it} measures *ideology* of user i at time t

  ϕ_j measures *ideology* of political account j
- and:

  α_j measures *popularity* of politician j

  β_i measures *political interest* of user i

Twitter-Based Ideal Points



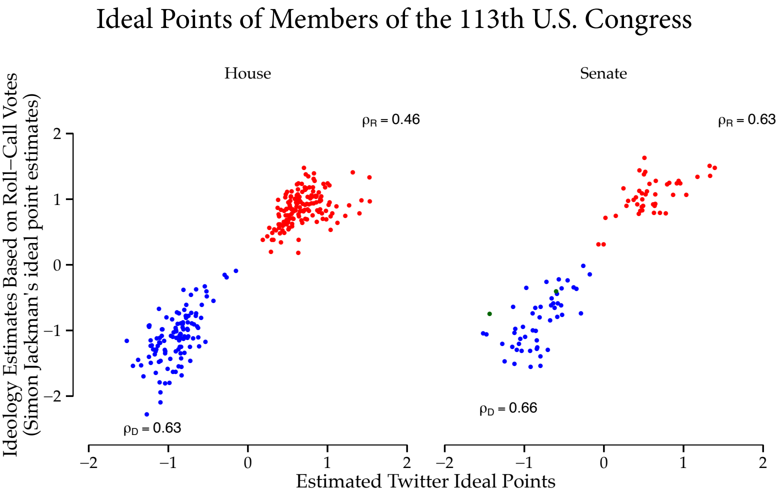
Validation

This method able to correctly classify and scale Twitter users on the left-right dimension:

1. Political elites
   - Correlation with measures based on roll-call votes.
2. Ordinary citizens
   - Individual and aggregate-level survey responses
   - Voting registration files
   - Campaign contribution records

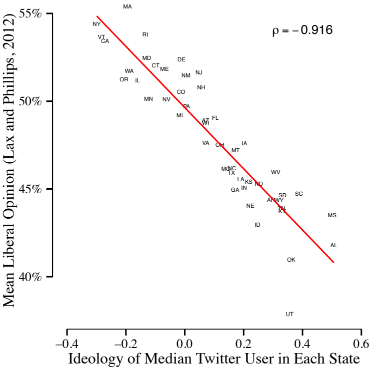
It is also able to predict change over time.

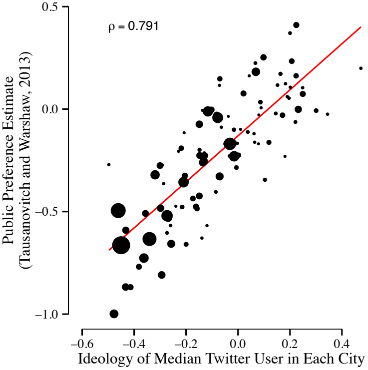
Political elites



Ordinary Users

Comparison with ideology estimates from aggregated surveys (Lax and Phillips, 2012; Tausanovitch and Warshaw, 2013)





Matching Twitter Accounts with Voting Records

Geographic location for Twitter users:

- 1.2 billion geolocated tweets ($\sim$8TB) from July 2013 to June 2014 $\rightarrow$ 250M in the U.S. (4.4M unique users)
- Use shape files to identify county and zipcode in U.S.

Voting registration records:

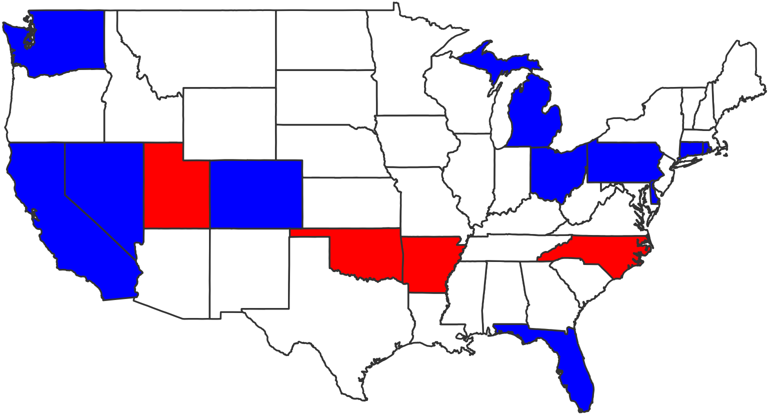
FIRST	LAST	VOTERID	COUNTY	PARTY	2012	GENDER …	
angela	myers	610901468	franklin	REP	X	F	…
ryan	petrik	610901998	franklin	DEM	X	M	…
…							

	RESIDENTIAL ADDRESS	ZIP	RACE	…
…	123 Main St, Columbus Oh	08001	W	…
…	77 Canal St, Columbus Oh	08009	W	…

Matching process:

- Perfect *and* unique matches of first/last name at county level
- If duplicated, match at zipcode level.

Matching Twitter Accounts with Voting Records

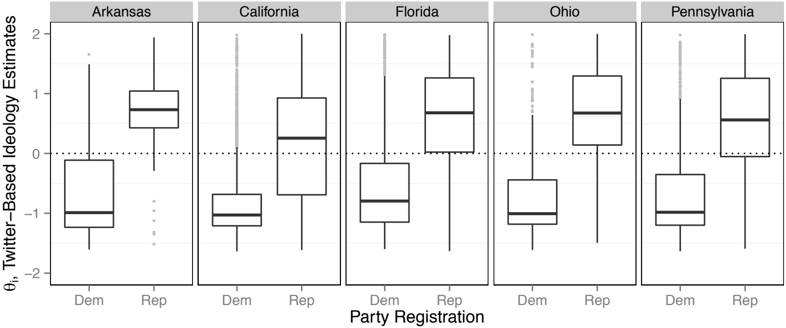


Code: github.com/pablobabera/voter-files

15 states, 77M registered voters (35-50% of U.S. total)

Ordinary Users

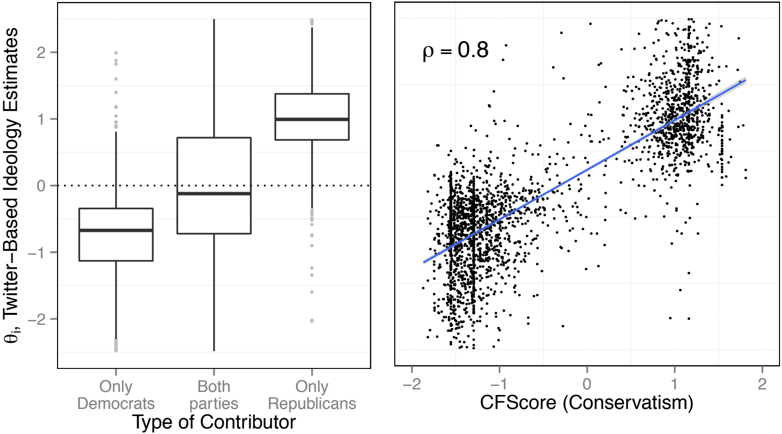
Republicans are more conservative than Democrats



Predictive accuracy for party affiliation is 83%

Campaign Contributions

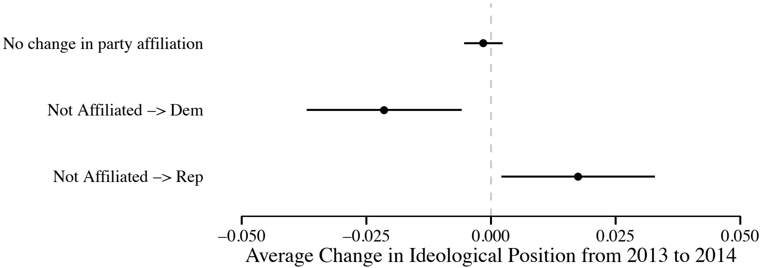
Voters who contributed to Democratic candidates only are more liberal than those contributing only to Republicans.



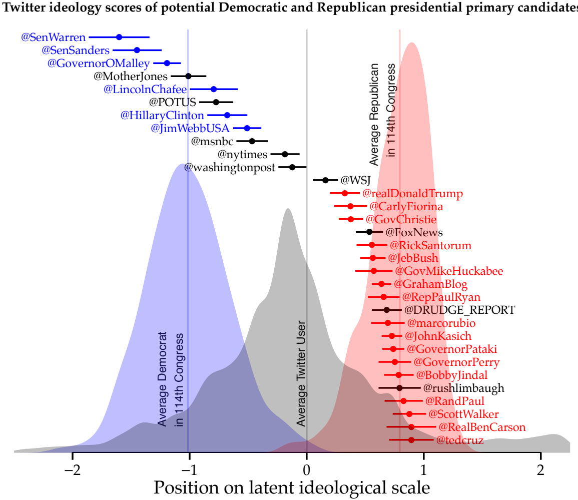
Data: campaign contribution records from Bonica (2014), matched with voting registration file in Ohio

Dynamic Validation

Changes in party affiliation (Ohio voters) from 2012 to 2014 are associated with changes in Twitter ideal points

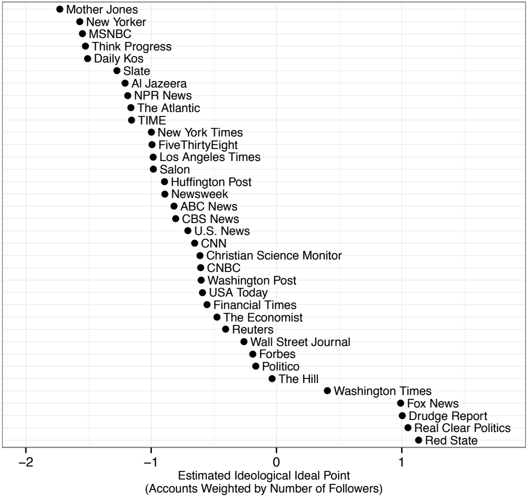


Application: Ideology of Presidential Candidates



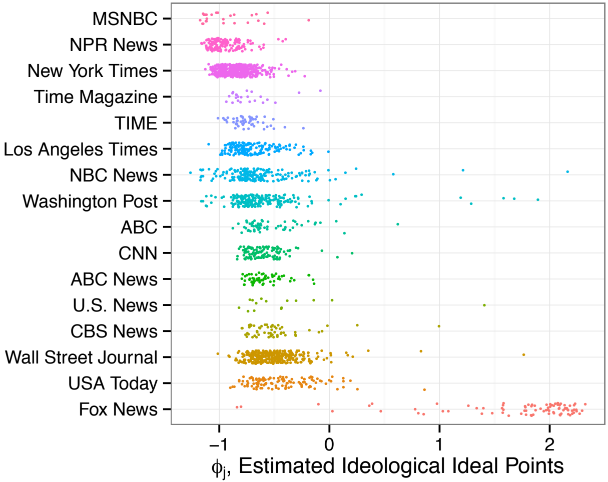
Barberá "Who is the most conservative Republican candidate for president?"
The Washington Post, June 16 2015

Application: Ideology of Media Outlets and Journalists



Barberá & Sood (2014) "Follow Your Ideology: A Measure of Ideological Location of Media Sources", *MPSA paper*

Application: Ideology of Media Outlets and Journalists



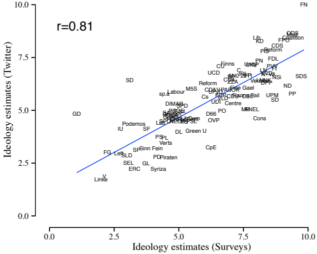
Barberá & Sood (2014) "Follow Your Ideology: A Measure of Ideological Location of Media Sources", *MPSA paper*

Application: Multidimensional Policy Spaces in Europe

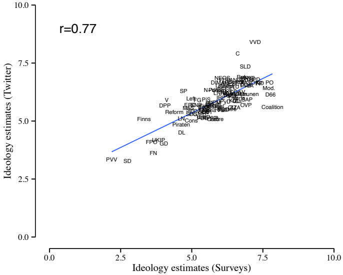
$$P(y_{ij}=1)=\text{logit}^{-1}\left(\alpha_i+\beta_j-\sum_{d=1}^{D}\gamma_d(\theta_{ik}-\phi_{jk})^2\right)$$

Estimated ideological positions for 120 parties in 28 European countries

Left-Right Dimension

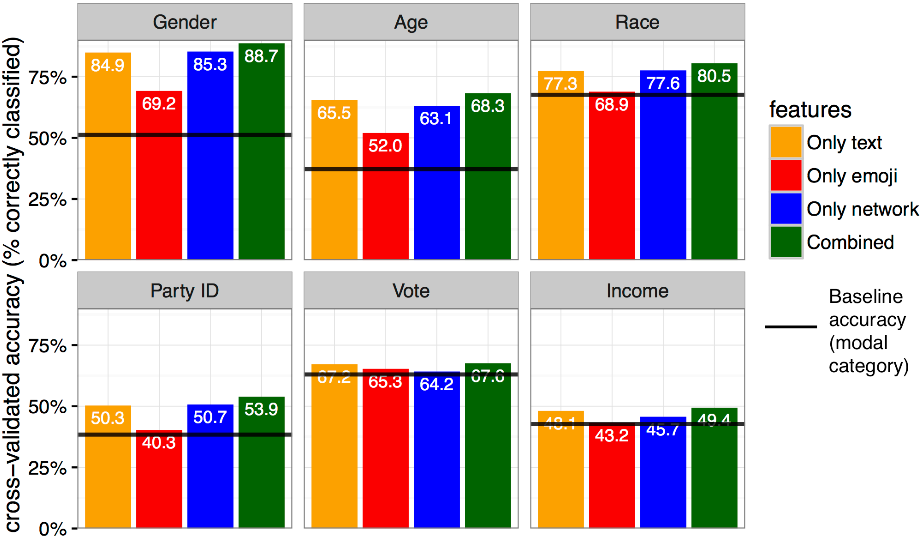


Pro/Anti-European Union Dimension



Barberá, Popa, & Schmitt (2015) “Analyzing the Common Multidimensional Political Space for Voters, Parties, and Legislators in Europe”, *MPSA paper*

Application: Predicting Sociodemographic Traits



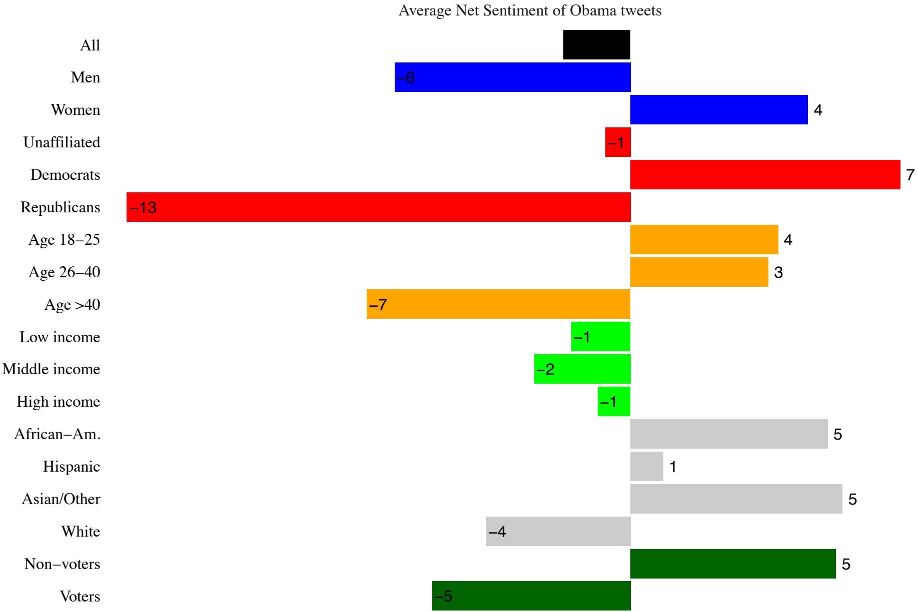
Application: Measuring Public Opinion with Twitter Data

Why can't we predict elections with Twitter data?

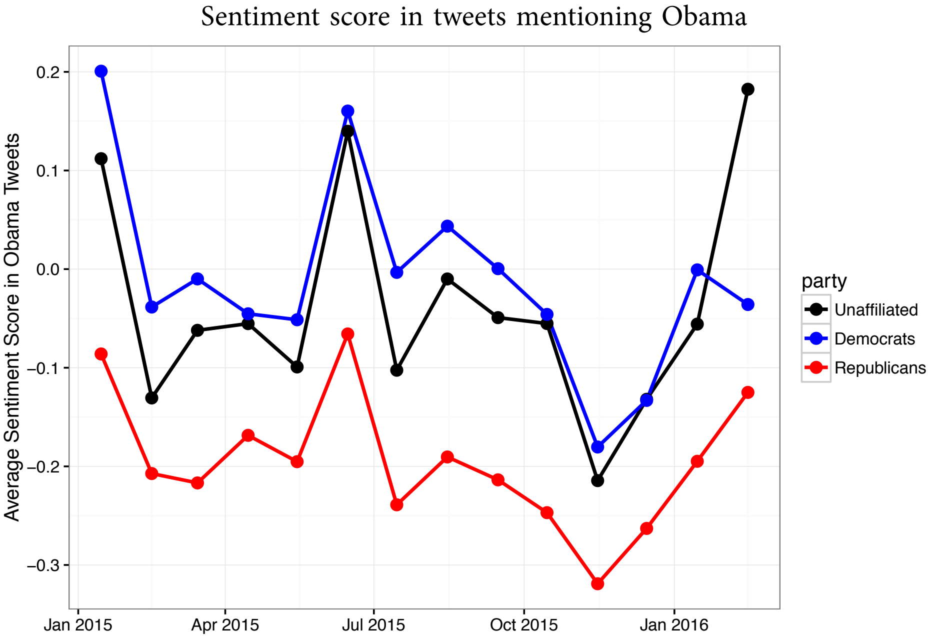
1. Sampling bias: who is on Twitter?
2. Nonresponse bias: who is tweeting about politics?

Importing methods from survey research:

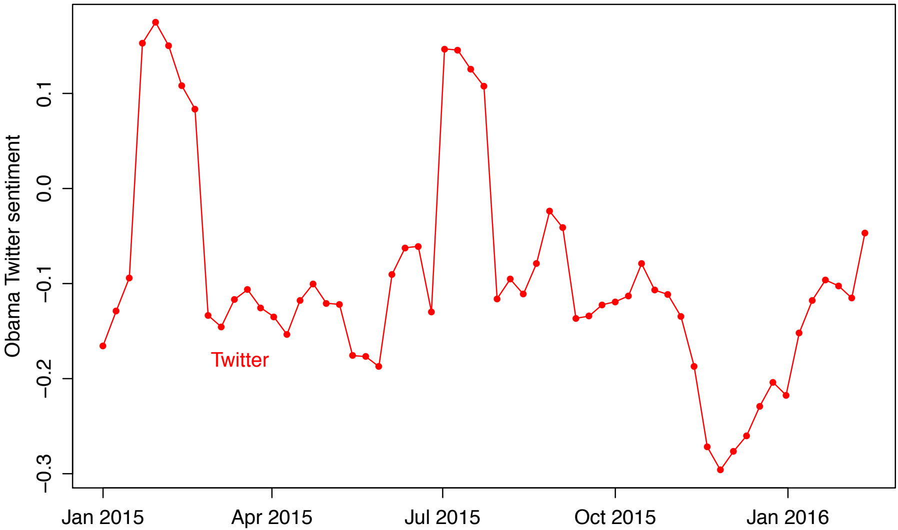
1. Post-stratification
   - Use sociodemographic information to compute sampling weights and adjust public opinion estimates
2. Panel design:
   - Same set of users across different issues
   - Use prior behavior to detect biases and "spiral of silence"
   - Sentiment analysis applied to tweets aggregated by user



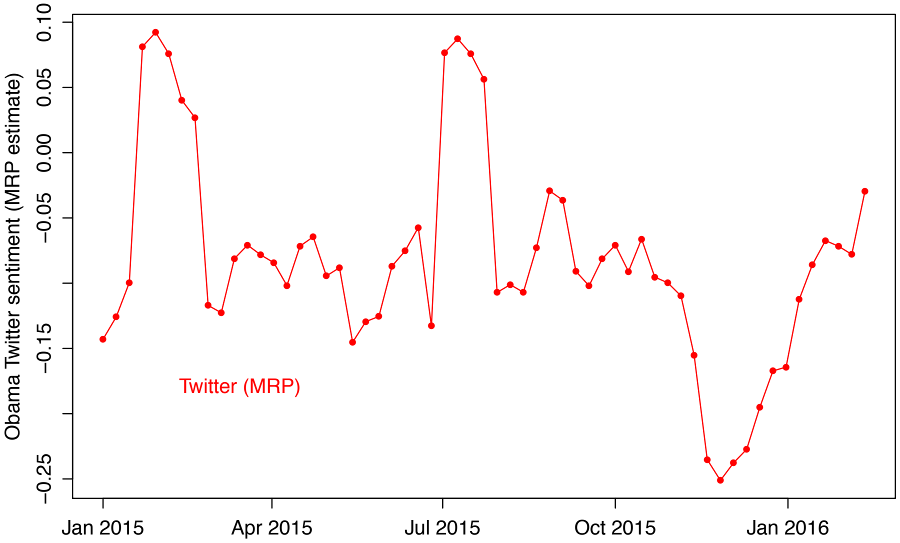
Average Net Sentiment of Obama tweets
All
Men
Women
Unaffiliated
Democrats
Republicans
Age 18–25
Age 26–40
Age >40
Low income
Middle income
High income
African–Am.
Hispanic
Asian/Other
White
Non–voters
Voters
–6
4
–1
7
–13
4
3
–7
–1
–2
–1
5
1
5
–4
5
–5


Sentiment score in tweets mentioning Obama
Average Sentiment Score in Obama Tweets
0.2
0.1
0.0
−0.1
−0.2
−0.3
Jan 2015
Apr 2015
Jul 2015
Oct 2015
Jan 2016
party
Unaffiliated
Democrats
Republicans

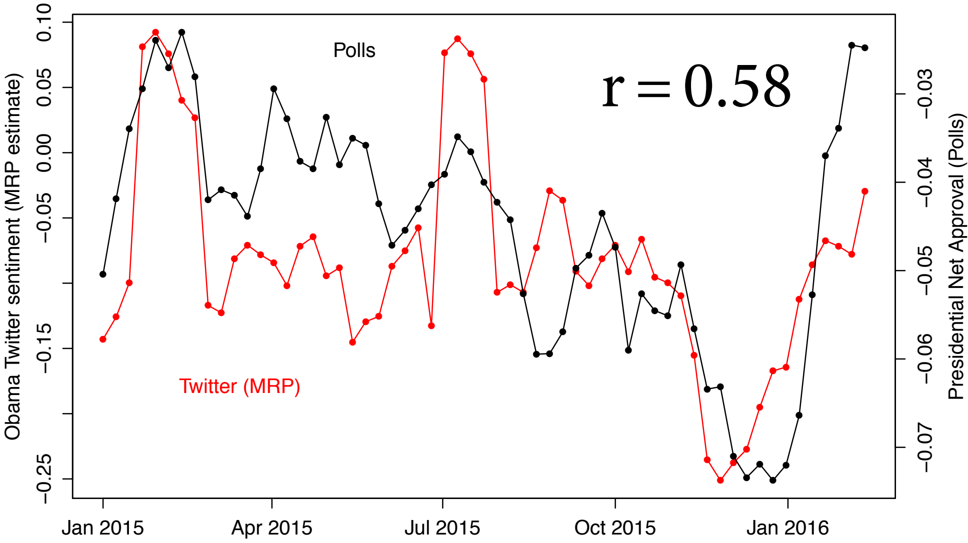
Sentiment score in Obama tweets



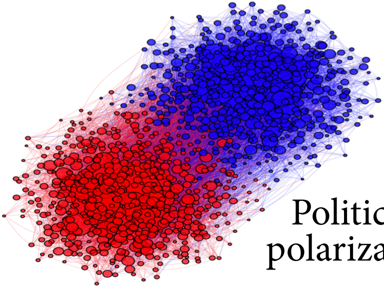
Sentiment score in Obama tweets



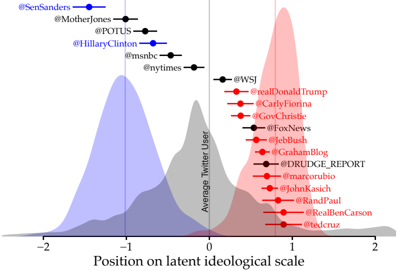
Sentiment score in Obama tweets



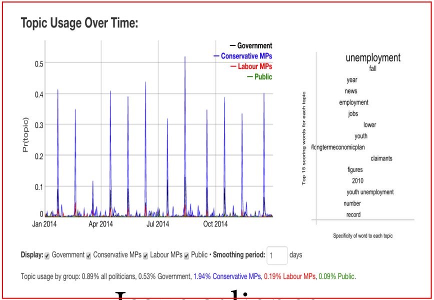
What social media data reveals about...



Political polarization



Latent individual traits



Issue salience

Political Representation

Citizens' political preferences → Politicians' positions → Public policy

assumes that...

Citizens' attention to issues → Politicians' attention to issues → Policy priorities

Agenda-setting

$\boxed{\text{Citizens' attention to issues} \longrightarrow \text{Politicians' attention to issues}}$

- "A Comparative Study of the Quality of Political Representation Using Social Media Data," joint work with Jørgen Bølstad.
- Goal: what explains citizens' capacity to set the agenda? The role of institutions and party-level characteristics.
- Use social media data to measure citizens' and politicians' attribution of salience to different issues in 6 European countries
- Preliminary findings:
  1. Political congruence (static) is higher in PR systems, but majoritarian systems allow greater responsiveness (dynamic)
  2. Parties are more responsive to the public in issues they own

Measuring Issue Attention

How to measure issue salience for **both** governments and voters?

- Governments, parties: manifestos, speeches
- Voters: “most-important problem” question

$\rightarrow$ Comparability issues (see e.g. Wlezien, 2005)

Left-right positions? (McDonald and Budge, 2005; Golder and Stramski, 2010) $\rightarrow$ not exactly what we want

Our proposed approach: social media data

Measuring Issue Salience

Social media data (Twitter)

- \+ Governments, parties, and voters are active Twitter users.
- \+ Data availability, granularity, comparability.
- – Sampling issues, different uses across countries.

Case selection

Theoretical expectations

Country	Government	Instit.	Congr.	Responsiv.
Denmark	Coalition	Proportional	High	Low
Germany	Coalition	Proportional	High	Low
Italy	Coalition	Proportional	High	Low
Spain	Single-party	Proportional	Medium	Medium
United Kingdom	Coalition	Majoritarian	Medium	Medium
France	Single-party	Majoritarian	Low	High

(More to come soon...)

Data

1. Government: institutions, ministers **in 2014**.
2. Gov. party and Opp. party: MPs for 1st and 2nd largest parties
3. "Informed citizens" (follow 1+ of 5 major media outlets)

		Gov.	Gov.Party	Opp.Party	Citizens
Denmark	Accounts	9	26	21	5,000
	Tweets	671	3,649	1,751	487,197
Germany	Accounts	36	53	75	5,000
	Tweets	17,227	23,075	26,531	810,013
Italy	Accounts	24	263	38	5,000
	Tweets	6,521	63,266	7,390	549,723
Spain	Accounts	20	62	80	5,000
	Tweets	17,054	34,568	49,910	1,234,855
UK	Accounts	42	200	196	5,000
	Tweets	40,540	105,442	130,464	682,383
France	Accounts	38	197	136	5,000
	Tweets	36,777	77,789	78,195	805,606

From Tweets to Issues

Existing methods: manual coding (e.g. Policy Agendas Project), dictionaries, supervised machine learning, unsupervised methods.

Our approach: topic modeling of tweets.

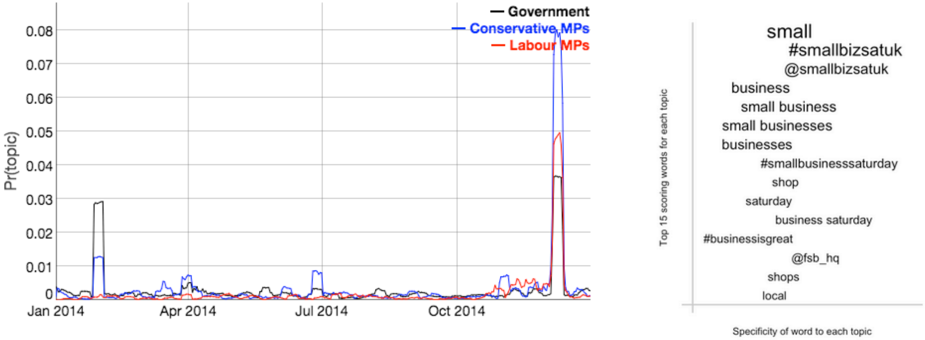
- Language-agnostic, does not require a priori judgment of relevant issues, flexible to idiosyncrasies of Twitter language, lower cost.
- Latent Dirichlet Allocation (LDA), probabilistic model of word occurrences:
  - Topic = distribution over words
  - Document = random mixture over latent topics

  → Document: aggregation of tweets by day and group
- Procedure:
  1. Estimate LDA with tweets from political accounts
  2. Apply LDA parameters to citizens' tweets
- K (number of topics) is set to $K = 75$ topics

Validation

j.mp/resp-lda-demo

Topic Usage Over Time:



Display: ☑ Government ☑ Conservative MPs ☑ Labour MPs ☐ Public • **Smoothing period:** 7 days

Topic usage by group: 0.30% all politicians, 0.30% Government, 0.37% Conservative MPs, 0.22% Labour MPs, 0.18% Public.

UK Prime Minister
@Number10gov

.@UKTi supported over 40k small businesses grow. Support your local business today bit.ly/1vnAWns @SmallBizSatUK #SmallBizSatUK

3:01 PM - 6 Dec 2014

89 62

Guy Opperman MP
@GuyOpperman

Small Business Saturday - try and support small businesses today and in the run up to Christmas: All of us can... bit.ly/1zxsbXj

9:25 AM - 6 Dec 2014

4 2

Mike Gapes
@MikeGapes

Support our local small businesses this Saturday by making a special effort to shop local #SmallBizSatUK

12:09 PM - 5 Dec 2014

1 1

Chris Heaton-Harris
@chhcalling

Celebrate Small Business Saturday this Saturday (6th December)!

This Saturday I'm hosting a small business... fb.me/2DLMUAtug

11:32 AM - 2 Dec 2014

1 2

Mark Spencer MP
@Mark_Spencer

Buying my fruit at Farm Direct in Hucknall supporting small business Saturday, shop local, use them or lose them pic.twitter.com/IHv9Ws8Vnr

12:38 PM - 6 Dec 2014

4 3

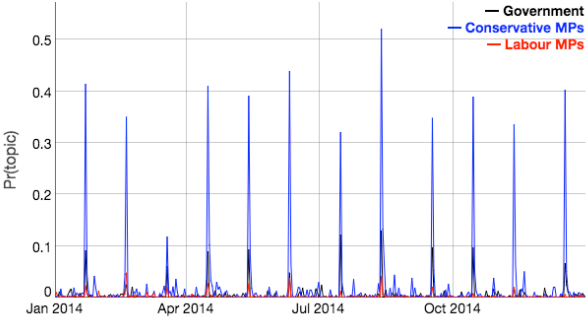
David Morris MP
@Davidmpmorris

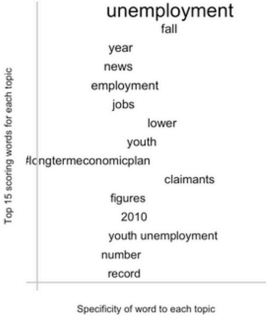
Today is Small business Saturday, celebrating all of the small businesses across the country. Shop small buy local! pic.twitter.com/KSdo7c2pdj

9:44 AM - 6 Dec 2014

10 3

Topic Usage Over Time:





Display: Government Conservative MPs Labour MPs Public • **Smoothing period:** 1 days

Topic usage by group: 0.89% all politicians, 0.53% Government, 1.94% Conservative MPs, 0.19% Labour MPs, 0.09% Public.

Sample of representative tweets by politicians:

Nicky Morgan @NickyMorgan01 — Follow

Today's employment figs show 680 JSA claimants in Loughborough in November-1.4% of econ active pop aged 16-64 & 364 lower than November 2013

1:53 PM - 17 Dec 2014

1

CCHQ Press Office @CCHQPress — Follow

BIGGEST annual drop in unemployment for 26 YRS, RECORD fall in youth unemployment, RECORD number of women in work - #LongTermPlan is working

11:19 AM - 17 Sep 2014

44 6

CCHQ Press Office @CCHQPress — Follow

FULL-TIME employment UP 1.3 million. RECORD fall in youth unemployment. RECORD number of women in work. #LongTermPlan is working

11:02 PM - 17 Sep 2014

21 2

Matt Hancock @MattHancockMP — Follow

Over half a million more full time jobs last year - shows the long term economic plan is working, giving more families economic security

11:18 AM - 16 Apr 2014

7 1

Sajid Javid @sajidjavid — Follow

Good news: Today's jobs numbers show unemployment continues to fall in #Bromsgrove. Now 1.7%. Youth uemployment @ lowest level in over 5 yrs

12:38 PM - 22 Jan 2014

1 1

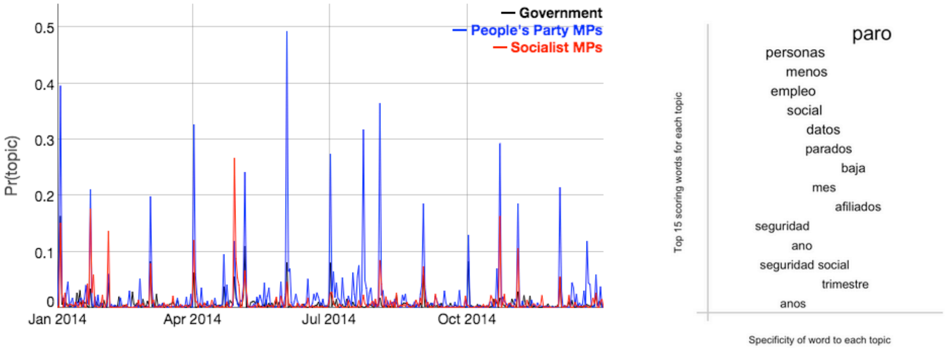
Kris Hopkins @krishopkins2015 — Follow

Biggest fall in youth unemployment since records began 30 yrs ago - #LongTermPlan is helping secure more jobs & opportunities for youngsters

10:01 PM - 13 Aug 2014

1

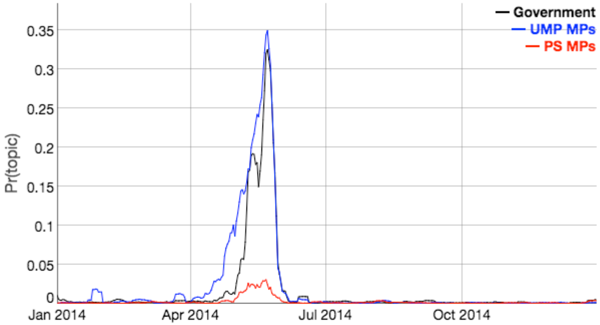
Topic Usage Over Time:

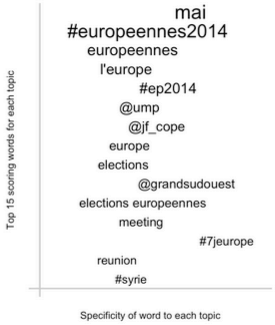


Display: ☑ Government ☑ People's Party MPs ☑ Socialist MPs ☐ Public • **Smoothing period:** 1 days

Topic usage by group: 1.30% all politicians, 0.75% Government, 2.19% People's Party MPs, 0.97% Socialist MPs, 0.24% Public.

Topic Usage Over Time:





Display: Government UMP MPs PS MPs Public • **Smoothing period:** 7 days

Topic usage by group: 1.34% all politicians, 1.59% Government, 2.15% UMP MPs, 0.28% PS MPs, 0.32% Public.

Analysis

Classify topics into four issue areas:

1. Economic policy
2. Social Policy
3. Defense, security, and nationalism
4. European politics and foreign affairs

Analysis:

- **Congruence**: collapse topics by group, and compute correlation coefficients to measure similarity of topic distributions
- **Responsiveness**: Granger causality framework with panel-variant of vector autoregressive model (VAR):

$$\Phi_{i,j,t} = \alpha_j + \sum_{i} \sum_{p=1}^{7} \beta_{i,p} \Phi_{i,j,t-p} + \varepsilon_{i,j,t} \tag{1}$$

where $\Phi_{i,j,t}$ is topic proportion for group i, topic k, at time t; and p indicates lag

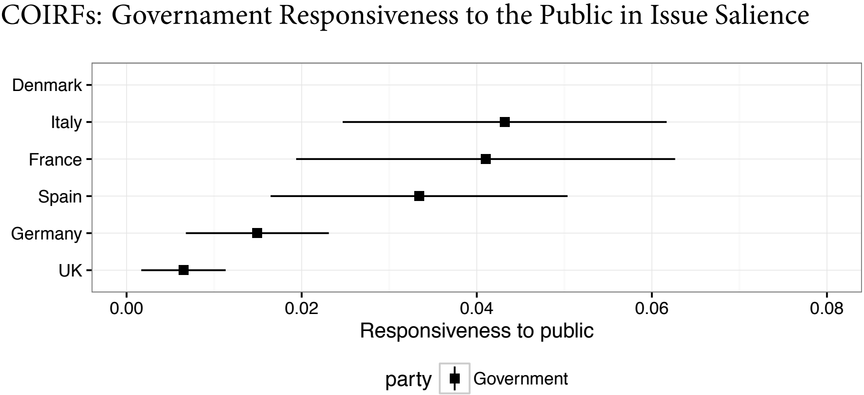
Congruence

Correlations of Issue Priorities (Political Issues)

	Government	Gov Party	Opposition
Germany	.306	-.333	.044
UK	.173	-.020	.053
Denmark	-.169	-.276	.283
Italy	-.349	.056	-.049
Spain	-.365	.042	-.005
France	-.429	-.241	.520

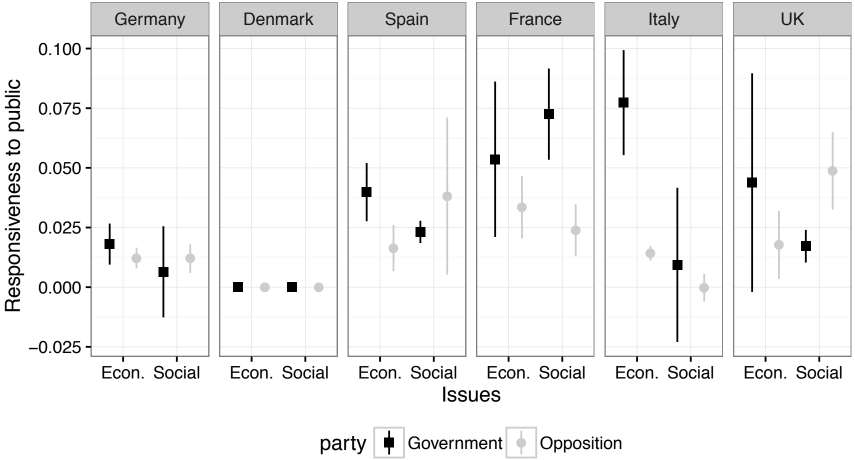
Note: The entries are bivariate correlations between the (logged) average issue priorities of the respective actors and the public for the year of 2014.

Responsiveness

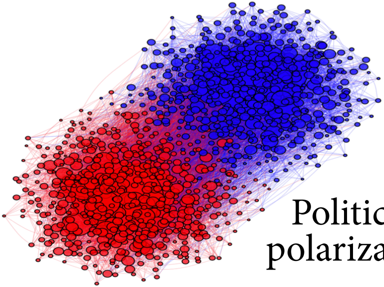


Issue Ownership and Responsiveness

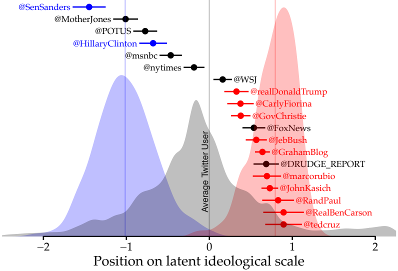
COIRFs: Govt. vs Opp. Responsiveness to the Public, by Issue Category



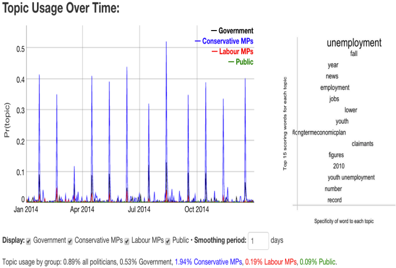
What social media data reveals about...



Political polarization



Latent individual traits



Issue salience

What social media reveals about...

1. Political polarization
   - Higher exposure to disagreement $\rightarrow$ moderation
2. Collective action
   - "Slacktivists" play a critical role in the success of protest, by increasing activity and reach of protest networks
3. Latent individual traits
   - Digital footprints from social media can be used to accurately predict ideology and other sociodemographic traits
4. Issue salience
   - Social media posts by elites and citizens reflect attention to issues, and can help us understand agenda-setting dynamics

Open questions

1. Political polarization
   - Exposure to disagreement $\rightarrow$ moderation $\rightarrow$ disinterest?
   - Algorithms: should Facebook try to stop Trump?
2. Collective action
   - Causal effect of social media on protest
   - Governments' response: disruption, censorship, engagement.
3. Latent individual traits
   - Online vs offline segregation, inequality in exposure to information
   - Combining online and offline data, surveys, multiple web sources
4. Issue salience
   - Who sets the public agenda? Citizens, media, government, parties, interest groups...?
   - Who can influence parties' agendas? Co-partisans, high- vs low-income, constituents vs general population... Inequality in political representation.

Open-source Software

Collecting social media data with R:

- streamR: Twitter streaming API
- smappR: Twitter REST API and DB management
- Rfacebook: Facebook Graph API

Analyzing social media data:

- Methods: github.com/pablobarbera/twitter_ideology
- Applications: github.com/SMAPPNYU/echo_chambers
- Teaching materials: github.com/pablobarbera/data-science-workshop

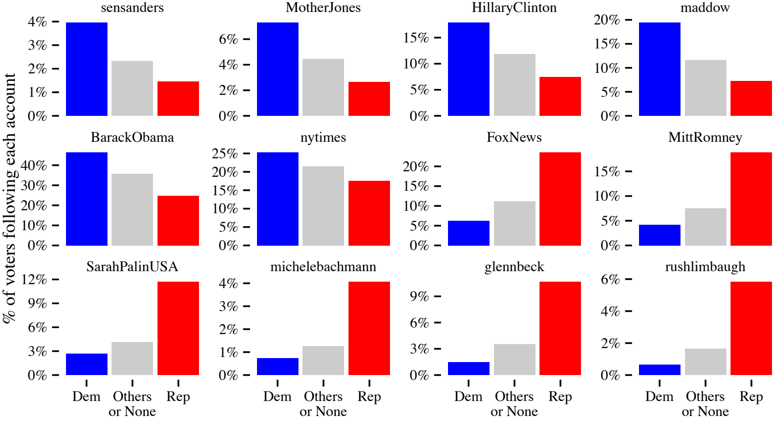
Birds of the same feather tweet together?
What social media data reveals about political behavior.

Pablo Barberá
Center for Data Science
New York University
www.pablobarbera.com

#iwsgrcp16

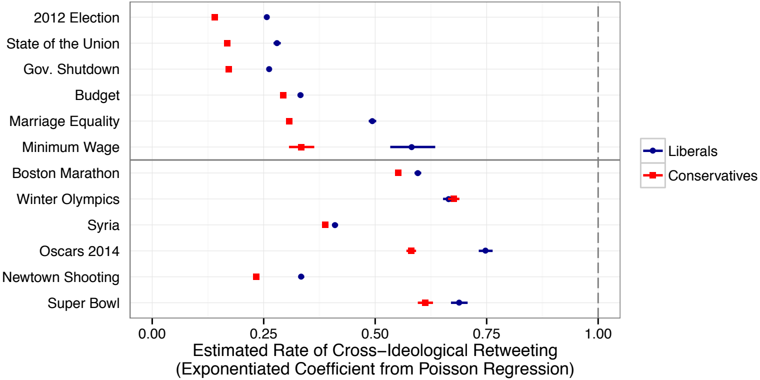
Ideological distance and following decisions

Observed proportion of users following a sample of political accounts (U.S. sample of voters matched with Twitter profiles)



(Accounts are ordered from most liberal to most conservative)

Application: Ideological Asymmetries in Pol. Comm.



Barberá, Jost, Nagler, Tucker, & Bonneau (2015) "Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?" *Psychological Science*

Evidence from Survey Data

Do respondents with social media accounts moderate their positions during election campaigns?

Data:

- 2012 ANES Time Series Study

Panel design:

- $E_{i,t=pre}$ and $E_{i,t=post}$: political extremism
  - Absolute difference between self-reported ideological position and position of average voter
- D_i: social media usage during campaign (dummy)

Regression model:

$$-(E_{i,t=post} - E_{i,t=pre}) = \beta_0 + \beta_1 D_i + \mathbf{X}\xi + \epsilon_i$$

$\mathbf{X}$: demographic, political, and media controls

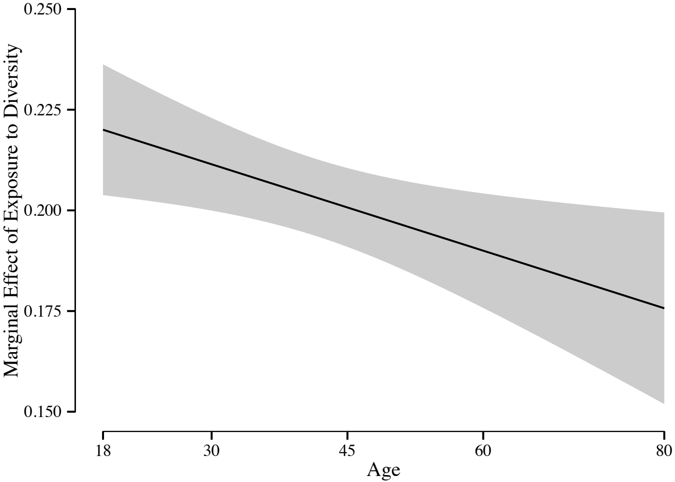
Table: OLS Regressions of Change in Political Moderation on Social Media Usage

	U.S.
Social Media Use	0.046*
	(0.022)
Demographic controls	✓
Political controls	✓
Media controls	✓
District fixed effects	✓
N	4,486
R^2	0.20
Resid. sd	0.63

Note: Robust standard errors, clustered by state, in parentheses.
* significant at $p < 0.10$

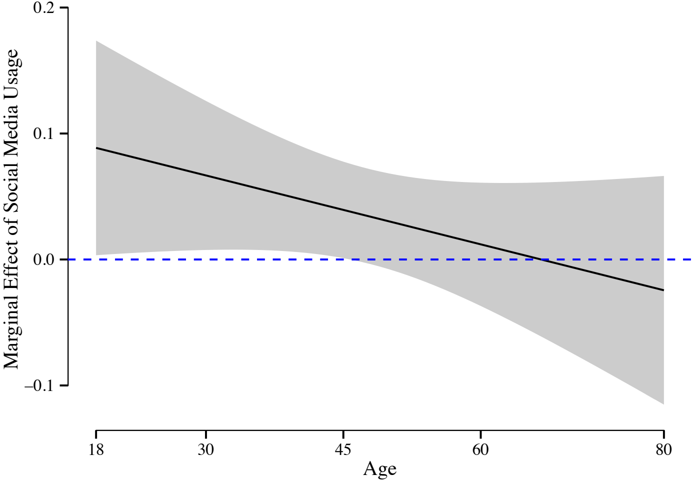
Findings: Age

Effect of exposure to diversity on political moderation is slightly larger for young voters (U.S. sample; social media data)



Findings: Age

Effect of exposure to diversity on political moderation is slightly larger for young respondents (U.S. sample; survey)



Alternative Operationalization of Diversity

$$\text{Diversity}_i = \frac{1}{|N_i|} \sum_{j \in N_i} ||\theta_i - \theta_j||$$

where N_i is set of social ties for user i

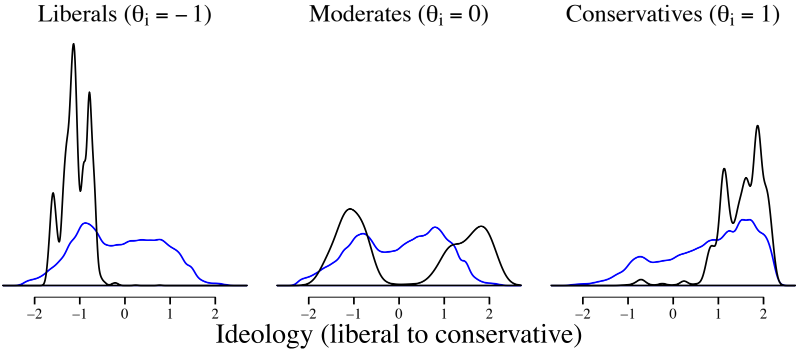
Table: OLS Regressions of Change in Political Moderation on Alternative Measure of Exposure to Disagreement in 2013

	United States		Germany	Spain
Relative Network Diversity	0.13*	0.13*	0.11*	0.13*
	(0.00)	(0.00)	(0.00)	(0.00)
Intercept	-0.05*	-0.05*	-0.11*	-0.14*
	(0.01)	(0.01)	(0.01)	(0.01)
Network controls	✓	✓	✓	✓
Offline controls		✓		
N	72,461	72,461	23,220	32,608
R^2	0.05	0.05	0.03	0.044
Resid. sd	0.23	0.23	0.31	0.28

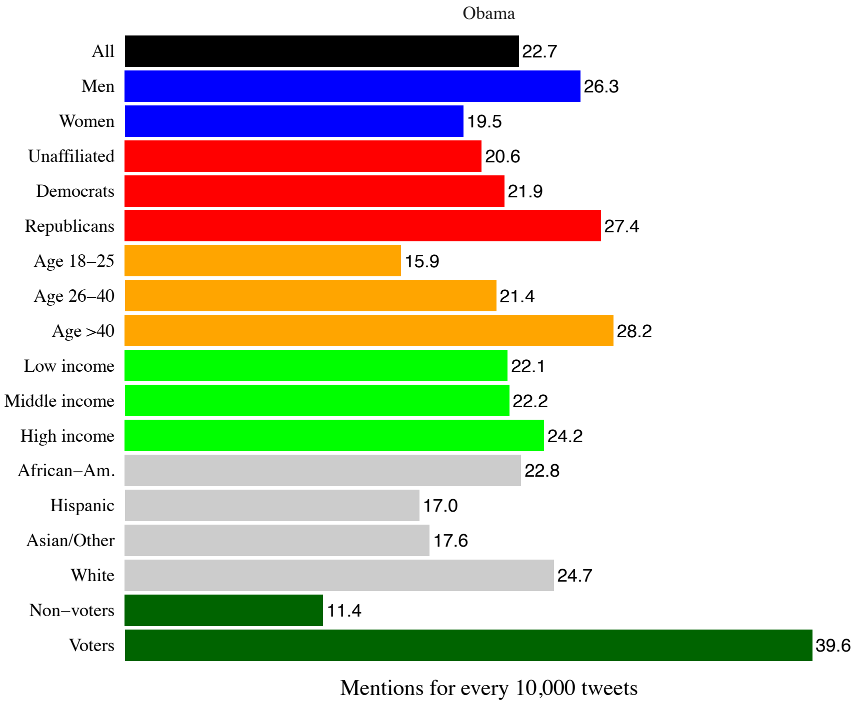
Note: * significant at $p < 0.05$. Standard errors in parentheses.

Exposure to Dissonant Opinions and Political Moderation

Ideological distribution of **political accounts** and social ties in users' Twitter networks



How many people are tweeting about politics?


Obama
All
22.7
Men
26.3
Women
19.5
Unaffiliated
20.6
Democrats
21.9
Republicans
27.4
Age 18–25
15.9
Age 26–40
21.4
Age >40
28.2
Low income
22.1
Middle income
22.2
High income
24.2
African–Am.
22.8
Hispanic
17.0
Asian/Other
17.6
White
24.7
Non–voters
11.4
Voters
39.6
Mentions for every 10,000 tweets