# Political Analysis of Social Media Data

## Ideology

Instructor: Gregory Eady
Office: 18.2.10
Office hours: Fridays 13-15

## Today

- Ideological scaling
- Video lectures & exercises

## Political ideology plays a massive role in our understanding of politics

- Polarization
- Vote choice
- Voting
- Misinformation
- Radicalization

## In the offline world, there are a wide variety of ways to measure political ideology:

- Ideological self-placement
- Surveys of political attitudes
- Votes by politicians
- Judge's court decisions
- Word use in parliamentary speeches
- Word use in news media editorials
- Campaign donations
- Expert surveys

## Ideology is difficult to measure because it cannot be directly observed:

- Self-placement scales have perceptual biases
- The meaning of ideology differs across people
- Ideology $\neq$ partisanship

## We thus measure ideology indirectly:

- Attitudes are the best known indicators of political ideology
- But are many behavioral indicators
- Measuring ideology is a cottage industry in political science

## The best known behavioral measure of political ideology is nominate scores

*A Spatial Model for Legislative Roll Call Analysis**

Keith T. Poole, *Carnegie-Mellon University*
Howard Rosenthal, *Carnegie-Mellon University*

A general nonlinear logit model is used to analyze political choice data. The model assumes probabilistic voting based on a spatial utility function. The parameters of the utility function and the spatial coordinates of the choices and the choosers can all be estimated on the basis of observed choices. Ordinary Guttman scaling is a degenerate case of this model. Estimation of the model is implemented in the NOMINATE program for one dimensional analysis of two alternative choices with no nonvoting. The robustness and face validity of the program outputs are evaluated on the basis of roll call voting data for the U.S. House and Senate.

## nominate scores

- Developed 40 years ago
- Extremely well-known
- Widely used
- Cited frequently in the media
- Used to assess convergent validity for basically every measure of ideology that captures politicians' ideology

## What are nominate scores?

- Authors use "roll-call" votes (Yea/Nay) in the US House and Senate
- Applicable anywhere roll-call votes are available
- Challenging to use in parliamentary systems, however, because of party discipline
- Scores themselves can be found here: `https://voteview.com`

## Data to calculate nominate scores look like this:

| | v1 | v2 | v3 | v4 | v5 | v6 | |
|---|---|---|---|---|---|---|---|
| legislator 1 | 0 | 0 | 1 | 0 | 1 | 0 | ... |
| legislator 2 | 1 | 1 | 1 | 1 | 1 | 0 | ... |
| legislator 3 | 0 | 0 | 0 | 1 | 1 | 1 | ... |
| legislator 4 | 1 | 0 | 1 | 0 | 1 | 1 | ... |
| legislator 5 | 1 | 1 | 1 | 0 | 0 | 0 | ... |

## Nominate scores are calculated from a model based on these data

- Are *unsupervised* models
  - Typically input no information about individuals except their votes
  - If we were trying to predict a known measure of ideology, it would be a *supervised* model
- We use the structure of the data to infer ideology, without using information about each person (e.g. party ID, gender, age, etc.)

## Basic intuition behind these models

- A legislator who votes primarily in favor of right-wing bills signals that his or her "latent" (unobserved) ideology is right-wing
- A legislator who votes primarily in favor of left-wing bills signals that his or her "latent" ideology is left-wing
- Ideology is latent here simply because we cannot directly observe it
- "Latent variable models" sound fancy, but they're just models where we infer the value of a variable that we observe only indirectly

## Let's say we actually somehow know the ideology of each legislator, in addition to how they voted on each bill

| | ideology | v1 | v2 | v3 | v4 | v5 | v6 | |
|---|---|---|---|---|---|---|---|---|
| legislator 1 | 1.53 | 0 | 0 | 1 | 0 | 1 | 0 | ... |
| legislator 2 | 0.71 | 1 | 1 | 1 | 1 | 1 | 0 | ... |
| legislator 3 | -1.44 | 0 | 0 | 0 | 1 | 1 | 1 | ... |
| legislator 4 | 0.02 | 1 | 0 | 1 | 0 | 1 | 1 | ... |
| legislator 5 | -0.89 | 1 | 1 | 1 | 0 | 0 | 0 | ... |

The variable "ideology" is the ideology of each legislator $i$, and the variables "v1", "v2", "v3", ..., indicate whether a legislator voted Nay (0) or Yea (1) on some bill $j$

**Then... we could estimate the relationship between ideology and voting for a specific bill $j$, for example, maybe $y_{j=1}$ (i.e. variable v1)**

$$\Pr(y_{i,j=1} = 1) = \text{logit}^{-1}(\alpha + \beta \text{ideology}_i), \tag{1}$$

where $y_{i,j=1} \in \{0, 1\}$ denotes a legislator $i$'s vote on one specific bill, $\alpha$ is the intercept; and $\beta$ captures the strength of the relationship between ideology and voting Yea or Nay on a given bill

## But what if we don't know the value of the variable "ideology"? (i.e. it is "latent"):

| | ideology | v1 | v2 | v3 | v4 | v5 | v6 | |
|---|---|---|---|---|---|---|---|---|
| legislator 1 | ? | 0 | 0 | 1 | 0 | 1 | 0 | ... |
| legislator 2 | ? | 1 | 1 | 1 | 1 | 1 | 0 | ... |
| legislator 3 | ? | 0 | 0 | 0 | 1 | 1 | 1 | ... |
| legislator 4 | ? | 1 | 0 | 1 | 0 | 1 | 1 | ... |
| legislator 5 | ? | 1 | 1 | 1 | 0 | 0 | 0 | ... |

## We turn the variable "ideology" into a parameter to be estimated from the data...

Thus:

$$\Pr(y_{i,j=1} = 1) = \text{logit}^{-1}(\alpha + \beta \text{ideology}_i) \tag{2}$$

Becomes:

$$\Pr(y_{ij} = 1) = \text{logit}^{-1}(\alpha_j + \beta_j \theta_i), \tag{3}$$

where $y_{ij} \in \{0, 1\}$ denotes a legislator $i$'s vote on bill $j$, $\alpha_j$ is the intercept for each specific bill; $\beta_j$ captures the strength of the relationship between ideology and voting Nay ($y = 0$) or Yea ($y = 1$) on a given bill; and $\theta_i$ denotes the ideology of legislator $i$

## After we fit this model, we can essentially fill in the latent/missing ideology variable:

| | theta | v1 | v2 | v3 | v4 | v5 | v6 | |
|---|---|---|---|---|---|---|---|---|
| legislator 1 | 1.53 | 0 | 0 | 1 | 0 | 1 | 0 | ... |
| legislator 2 | 0.71 | 1 | 1 | 1 | 1 | 1 | 0 | ... |
| legislator 3 | -1.44 | 0 | 0 | 0 | 1 | 1 | 1 | ... |
| legislator 4 | 0.02 | 1 | 0 | 1 | 0 | 1 | 1 | ... |
| legislator 5 | -0.89 | 1 | 1 | 1 | 0 | 0 | 0 | ... |

## This basic model has many possible extensions

- Changes in ideology over time
- Differences across domains (state versus federal legislatures)
- Different variable types (e.g. binary, continuous, count)
- Different types of data (votes + text)
- Many applications to many domains...

# Supreme Court decisions (with only 9 justices)

## Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999

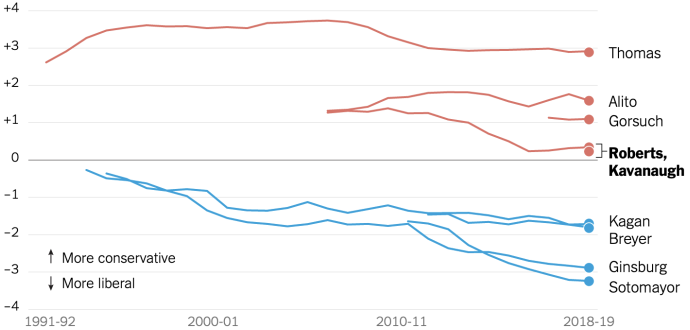
**Andrew D. Martin**
*Department of Political Science, Washington University,*
*Campus Box 1063, One Brookings Drive,*
*St. Louis, MO 63130-4899*
*e-mail: admartin@artsci.wustl.edu*

**Kevin M. Quinn**
*Department of Political Science and*
*Center for Statistics in the Social Sciences, Box 354320,*
*University of Washington, Seattle, WA 98195-4322*
*e-mail: quinn@stat.washington.edu*

At the heart of attitudinal and strategic explanations of judicial behavior is the assumption that justices have policy preferences. In this paper we employ Markov chain Monte Carlo methods to fit a Bayesian measurement model of ideal points for all justices serving on the U.S. Supreme Court from 1953 through 1999. We are particularly interested in determining to what extent ideal points of justices change throughout their tenure on the Court. This is important because judicial politics scholars oftentimes invoke preference measures that are time invariant. To investigate preference change, we posit a dynamic item response model that allows ideal points to change systematically over time. Additionally, we introduce Bayesian methods for fitting multivariate dynamic linear models to political scientists. Our results suggest that many justices do not have temporally constant ideal points. Moreover, our ideal point estimates outperform existing measures and explain judicial behavior quite well across civil rights, civil liberties, economics, and federalism cases.

# Martin and Quinn scores in the New York Times:

**Justice Ideology Based on Martin-Quinn Scores**




Note: Red lines indicate justices appointed by a Republican, and blue lines by a Democrat.

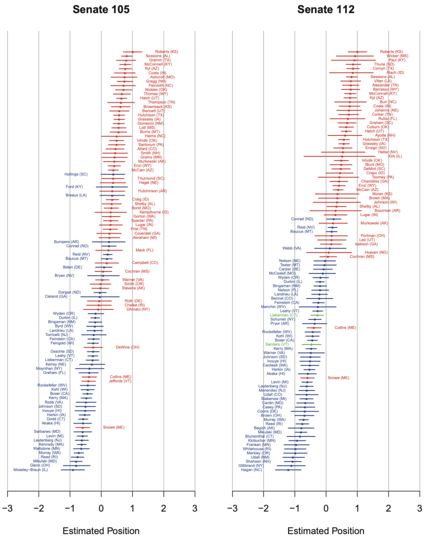
Source: Ideology scores are based on voting patterns and developed from the Supreme Court Database by Lee Epstein and Andrew D. Martin, Washington University in St. Louis, and Kevin Quinn, University of Michigan.

## Text-based ideology ("Wordfish")

# A Scaling Model for Estimating Time-Series Party Positions from Texts

**Jonathan B. Slapin** Trinity College, Dublin
**Sven-Oliver Proksch** University of California, Los Angeles

*Recent advances in computational content analysis have provided scholars promising new ways for estimating party positions. However, existing text-based methods face challenges in producing valid and reliable time-series data. This article proposes a scaling algorithm called WORDFISH to estimate policy positions based on word frequencies in texts. The technique allows researchers to locate parties in one or multiple elections. We demonstrate the algorithm by estimating the positions of German political parties from 1990 to 2005 using word frequencies in party manifestos. The extracted positions reflect changes in the party system more accurately than existing time-series estimates. In addition, the method allows researchers to examine which words are important for placing parties on the left and on the right. We find that words with strong political connotations are the best discriminators between parties. Finally, a series of robustness checks demonstrate that the estimated positions are insensitive to distributional assumptions and document selection.*


Senate 105
Senate 112
−3 −2 −1 0 1 2 3
Estimated Position
−3 −2 −1 0 1 2 3
Estimated Position

## Scaling democracy

# Democracy as a Latent Variable

**Shawn Treier** University of Minnesota
**Simon Jackman** Stanford University

*We apply formal, statistical measurement models to the Polity indicators, used widely in studies of international relations to measure democracy. In so doing, we make explicit the hitherto implicit assumptions underlying scales built using the Polity indicators. Modeling democracy as a latent variable allows us to assess the "noise" (measurement error) in the resulting measure. We show that this measurement error is considerable and has substantive consequences when using a measure of democracy as an independent variable in cross-national statistical analyses. Our analysis suggests that skepticism as to the precision of the Polity democracy scale is well founded and that many researchers have been overly sanguine about the properties of the Polity democracy scale in applied statistical work.*

## Campaign finance scores

# Mapping the Ideological Marketplace

**Adam Bonica** Stanford University

*I develop a method to measure the ideology of candidates and contributors using campaign finance data. Combined with a data set of over 100 million contribution records from state and federal elections, the method estimates ideal points for an expansive range of political actors. The common pool of contributors who give across institutions and levels of politics makes it possible to recover a unified set of ideological measures for members of Congress, the president and executive branch, state legislators, governors, and other state officials, as well as the interest groups and individuals who make political donations. Since candidates fundraise regardless of incumbency status, the method estimates ideal points for both incumbents and nonincumbents. After establishing measure validity and addressing issues concerning strategic behavior, I present results for a variety of political actors and discuss several promising avenues of research made possible by the new measures.*

# Finally, social media...

**Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data**

**Pablo Barberá**

*Wilf Family Department of Politics, New York University, 19 W 4th Street, 2nd Floor, New York, NY 10012.*
*e-mail: pablo.barbera@nyu.edu*

**Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook**

ROBERT BOND *Ohio State University*
SOLOMON MESSING *Facebook Data Science*

# Birds of the Same Feather Tweet Together (Barberá, 2015)

That this paper isn't really framed around social media itself says a lot about where we have come since the earlier days of social media:

- Need measures of ideology for understanding electoral behavior, government formation, and party competition
- But most measures only are for citizens, or legislators alone
- Often difficult to get measures of both citizens and legislators simultaneously

## Why are social media data useful for examining political ideology?

- Both ordinary members of the public *and* politicians use it
- Almost all politicians in the US (and many in other countries) use social media for political communication
- The public and politicians frequently interact

## Twitter data are beneficial for examining ideology for three reasons:

1. Massive number of users interact in various ways with politicians
2. Social media are highly dynamic, potentially allowing fine-grained estimates in real-time
3. Twitter profiles that have real names that can be linked to offline data like voting records

The drawback

- Not a representative sample
- Interactions on social media that we signal ideology might be signal of other things too

## The goal:

- Estimate the ideology of both Twitter users and politicians simultaneously on the same ideology scale
- Use the structure of network data based on the politicians that ordinary users *follow*
    - This network structure setup is analogous to Bond & Messing's (2015) article, which uses which politicians users "endorse" as data

## The intuition:

- Twitter users follow politicians whose (latent) ideology is similar to their own:
  - A user will be more likely to follow a politician if that politician is perceived as ideologically "close"
  - A user will be less likely to follow a politician if that politician is perceived as ideologically "distant"
- The challenge is to develop a statistical model to use this assumption to measure ideology

## Basic data setup:

- Collect all of the followers of politicians on Twitter
- Clean the data so that we know which users follow which politicians
- The rows in the data are the users
- The columns in the data are the politicians (maybe media accounts too, e.g. @CNN, @FoxNews, @BreitbartNews, etc.)

## Example data:

| | pol1 | pol2 | pol3 | pol4 | pol5 | pol6 | |
|---|---|---|---|---|---|---|---|
| user 1 | 0 | 0 | 1 | 0 | 0 | 0 | ... |
| user 2 | 1 | 1 | 1 | 1 | 1 | 0 | ... |
| user 3 | 0 | 0 | 1 | 1 | 1 | 1 | ... |
| user 4 | 1 | 1 | 1 | 0 | 1 | 1 | ... |
| user 5 | 1 | 1 | 1 | 1 | 0 | 0 | ... |

## Statistical model:

$$\Pr(y_{ij} = 1) = \text{logit}^{-1}(\alpha_j + \beta_i - \gamma||\theta_i - \phi_j||^2) \tag{4}$$

$y_{ij} \in \{0, 1\}$: whether user $i$ follows politician $j$

$\alpha_j$: the extent that politician $j$ is followed in general
$\beta_i$: the extent that a user follows many politicians generally
$\theta_i$: ideology of user $i$
$\phi_j$: ideology of politician $j$
$\gamma$: extent that the ideological distance between user $i$ and politician $j$ affects the probability of the user following that politician

# Do not be intimidated by this kind of thing:

$$P(y_{ij} = 1|\alpha_j, \beta_i, \gamma, \theta_i, \phi_j) = \text{logit}^{-1}\left(\alpha_j + \beta_i - \gamma||\theta_i - \phi_j||^2\right). \quad (1)$$

Given that none of these parameters is directly observed, the statistical problem here is the inference of $\boldsymbol{\theta} = (\theta_\mathrm{i}, \ldots, \theta_\mathrm{n})'$, $\boldsymbol{\phi} = (\phi_\mathrm{j}, \ldots, \phi_\mathrm{m})'$, $\boldsymbol{\alpha} = (\alpha_\mathrm{j}, \ldots, \alpha_\mathrm{m})'$, $\boldsymbol{\beta} = (\beta_\mathrm{i}, \ldots, \beta_\mathrm{n})'$, and $\gamma$. Assuming local independence (individual decisions to follow are independent across users $n$ and $m$, conditional on the estimated parameters), the likelihood function to maximize this model is as follows:

$$p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) = \prod_{\mathrm{i}=1}^{\mathrm{n}}\prod_{\mathrm{j}=1}^{\mathrm{m}} \text{logit}^{-1}(\pi_{\mathrm{ij}})^{\mathrm{y_{ij}}}(1 - \text{logit}^{-1}(\pi_{\mathrm{ij}}))^{1-\mathrm{y_{ij}}}, \quad (2)$$

where $\pi_{ij} = \alpha_j + \beta_i - \gamma||\theta_i - \phi_j||^2$.

Estimation and inference for this type of model is not trivial. Maximum-likelihood estimation methods are usually intractable given the large number of parameters involved. However, samples from the posterior density of each parameter in the model can be obtained using Markov Chain Monte Carlo methods. To improve the efficiency of this procedure, I use a Hamiltonian Monte Carlo algorithm (Gelman et al. 2013) and employ a hierarchical setup that considers each of the four sets of parameters as draws from four common population distributions: $\alpha_j \sim \mathrm{N}(\mu_\alpha, \sigma_\alpha)$, $\beta_j \sim \mathrm{N}(\mu_\beta, \sigma_\beta)$, $\theta_i \sim \mathrm{N}(\mu_\theta, \sigma_\theta)$, and $\phi_j s \sim \mathrm{N}(\mu_\phi, \sigma_\phi)$. The full joint posterior distribution is thus

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma|\mathbf{y}) \propto \mathrm{p}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{\mu}, \boldsymbol{\sigma}) & \\ \prod_{i=1}^{n}\prod_{j=1}^{m} \text{logit}^{-1}(\pi_{ij})^{y_{ij}}(1 - \text{logit}^{-1}(\pi_{ij}))^{1-y_{ij}} & \\ \prod_{j=1}^{m}\mathrm{N}(\alpha_j|\mu_\alpha, \sigma_\alpha)\prod_{i=1}^{n}\mathrm{N}(\beta_i|\mu_\beta, \sigma_\beta)\prod_{i=1}^{n}\mathrm{N}(\theta_i|\mu_\theta, \sigma_\theta)\prod_{j=1}^{m}\mathrm{N}(\phi_j|\mu_\phi, \sigma_\phi). & \end{aligned} \quad (3)$$

## What data to use?

- Model is generalizable to any type of actors that users follow
- Could fit this model for users who follow footballers, musicians, or actors
- *But* the meaning of the scale would heavily change
- You thus want actors that are informative about ideology specifically
- Barberá (2015) includes other actors with clear ideological positions: politicians, news media, and think tanks

## Collecting and pre-processing the data:

1. Create a list of the political accounts that are informative about ideology
   - e.g. collect the user IDs of Danish politicians, media, & commentators
2. Collect the lists of users who follow each of these accounts
3. Remove inactive users/bots/non-country residents:
   - Collect the user profiles of all users who follow at least one account. Then remove users who:
     - Sent fewer than one hundred tweets
     - Have not sent a tweet in last 6 months
     - Have fewer than 25 followers
     - Are outside of the country of interest
     - Follow fewer than three accounts of politicians

## Steps:

4. Clean the data to generate a user-politician matrix
   - Rows as the users
   - Columns as politicians, news media, and commentators
5. Fit the model (use the R library `emIRT`)

## Remember, in practice, you just need a dataset like this:

| user_id | p432032 | p904390 | p439234 | p9726 | p726 | |
|---|---|---|---|---|---|---|
| 54325970 | 0 | 0 | 1 | 0 | 0 | ... |
| 1203213 | 1 | 1 | 1 | 1 | 1 | ... |
| 5454930 | 0 | 0 | 1 | 1 | 1 | ... |
| 443 | 1 | 1 | 1 | 0 | 1 | ... |
| 90784532 | 1 | 1 | 1 | 1 | 0 | ... |

Where the users (rows) are those who follow politicians, news media, and commentators (columns). A cell with a 1 simply indicates that a user follows a given politician, and a 0 indicates that he or she does not. That's it.

## Barberá (2015) fits his model to the following:

- US (n = 301,537)
- UK (n = 135,015)
- Spain (n = 123,846)
- Italy (n = 150,143)
- Germany (n = 49,142)
- Netherlands (n = 96,625)

## Recall that the model returns estimates for:

1. Users, $\theta_i$
2. Politicians, media, & commentators, $\phi_i$

We can thus validate the model for both sets of actors. How might we do this?

# Remember “nominate” from earlier?

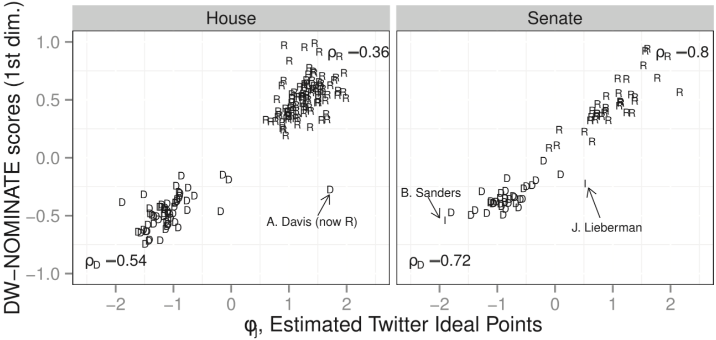
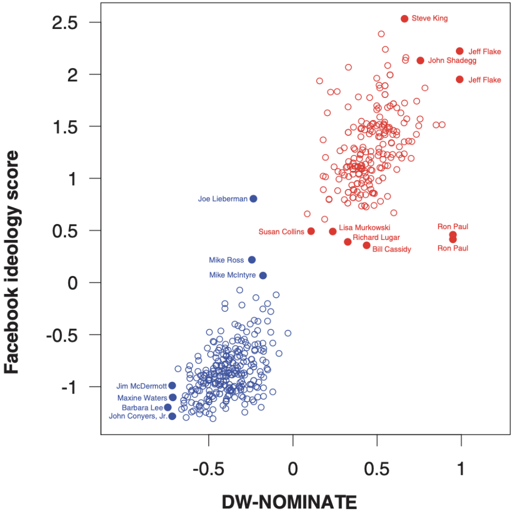



**Fig. 1** Ideal point estimates for members of US Congress.

*Within*-party validation is critical

# For Bond & Messing's (2015) Facebook model



# European party positions from an expert survey

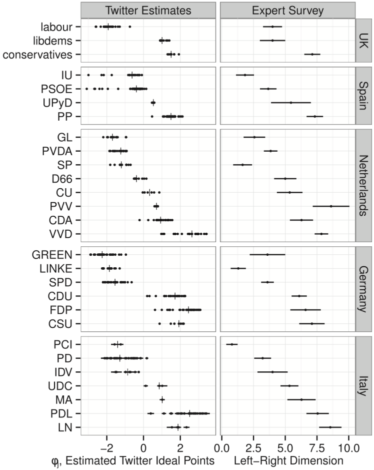



**Fig. 3** Ideological location of parties in five European countries.

# Comparing ordinary users and elites

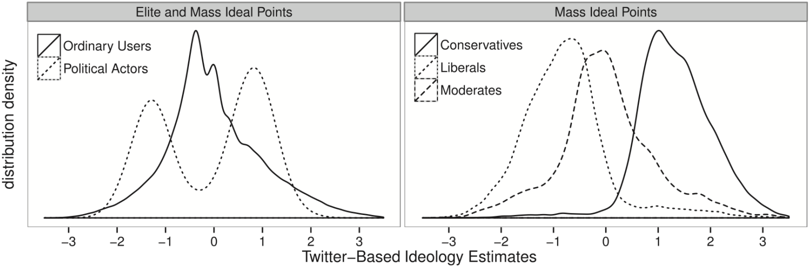



**Fig. 4** Distribution of political actors and ordinary Twitter users' ideal points.

Liberals: {"liberal," "progressive," "Democrat"}
Moderates: {"independent," "moderate"}
Conservatives: {"conservative," "GOP," "Republican"}

## Also compares to voting records and campaign donations

- Twitter ideology scores correlate at 0.8 with campaign finance scores for those who have donated to political candidates
- Twitter ideology scores correlate with party registration as Democrat or Republican (in Ohio)

# Simple application: Echo chambers (sort of)

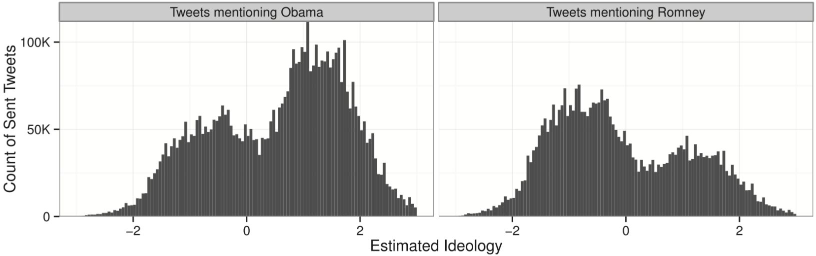



**Fig. 6** Number of tweets mentioning presidential candidates, by ideal point bin.

Those discussing Obama and Romney are from the ideological poles (bi-modal distribution compared to uni-modal distribution of all users)

## Simple application: Echo chambers (sort of)

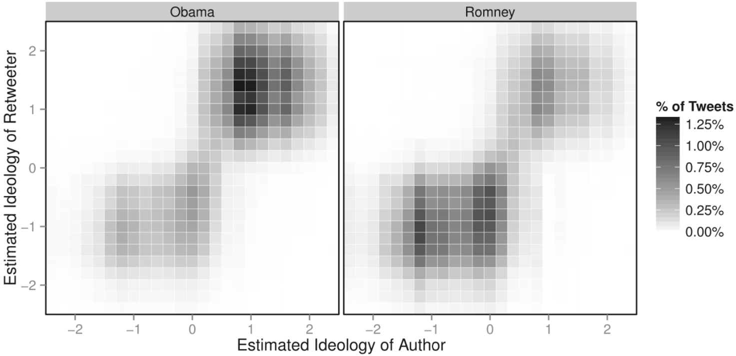



**Fig. 7** Political polarization in retweets mentioning presidential candidates.

Users retweet others with similar ideology

## These methods have many potential uses. To give one recent example:

# Exposure to opposing views on social media can increase political polarization

**Christopher A. Bail[a,1], Lisa P. Argyle[b], Taylor W. Brown[a], John P. Bumpus[a], Haohan Chen[c], M. B. Fallin Hunzaker[d], Jaemin Lee[a], Marcus Mann[a], Friedolin Merhout[a], and Alexander Volfovsky[e]**

[a]Department of Sociology, Duke University, Durham, NC 27708; [b]Department of Political Science, Brigham Young University, Provo, UT 84602; [c]Department of Political Science, Duke University, Durham, NC 27708; [d]Department of Sociology, New York University, New York, NY 10012; and [e]Department of Statistical Science, Duke University, Durham, NC 27708

## A few things to keep in mind:

- How you select the politicians, news media, and commentators matters
- Ideology estimates of users who follow very few actors will be very noisy
- Ideology estimates of politicians, news media, and commentators will typically be precise
- Follower data can take a *long* time to collect from the Twitter API, especially with actors like Donald Trump, Barack Obama, Joe Biden (tens of millions of followers)

# Political Information Sharing and Ideological Polarization

Why do some politicians share ideologically extreme content on social media, while others share content that is ideologically moderate?

# News Sharing on Social Media: Mapping the Ideology of News Media, Politicians, and the Mass Public

Gregory Eady[*] Richard Bonneau[†] Joshua A. Tucker[§] Jonathan Nagler[*]

This article examines the information sharing behavior of US politicians and the mass public by mapping the ideological sharing space of political news on social media. As data, we use the near-universal currency of online information exchange: web links. We introduce a methodological approach and statistical software to unify the measurement of ideology on social media platforms by using sharing data to jointly estimate the ideology of news media organizations, politicians, and the mass public. Empirically, we investigate the electoral incentives that members of Congress have to share ideologically polarizing information online. We show that the more competitive an election is, the less likely politicians are to share ideologically polarizing information. This finding has important implications for our understanding of the role of election pressures as constraints on sharing behavior in our highly polarized information ecosystem.

## Introduces a method to measure the social media ideological sharing space of:

1. Politicians
2. Users
3. News media content

## Ideology measures are not unified for these actors

- Users
  - Who users follow or endorse (Barberá 2015, Bond & Messing 2015)
- Politicians
  - Roll-call votes (Poole & Rosenthal 1985), campaign donations (Bonica 2013, 2014), expert placements (Bakker et al. 2015)
- News media
  - Editorials (Ho & Quinn 2008), text comparisons to politicians (Gentzkow & Shapiro 2010, Martin & Yurukoglu 2017), crowd-sourcing (Budak et al. 2016), sharing by liberals & conservatives (Bakshy et al. 2015)

## Approach

- Use the universal currency of online information exchange: web links
    - e.g. https://nytimes.com/2018/02/13/upshot/...

## The many benefits of web link data

1. Platform-agnostic
2. Behavioral measure for politicians
3. Captures how people use news media
4. URLs are frequently shared
5. Changes quickly over time
6. Works with relatively small datasets
7. Do not need partisan classifications of users for measures of news media ideology

## Data from Twitter timelines

- Members of Congress (535)
- Governors (50)
- President, Vice President, presidential candidates
- Ordinary users (5,000 users geo-located to US)

# Example of news sharing

# Exclude quote tweets

Walter Russell Mead
@wrmead

Follow

For all its flaws, the Great Potato Famine taught Irish farmers to launch new careers. #NYTHotTakes

NYT Opinion @nytopinion
For all its flaws, the Communist revolution taught Chinese women to dream big
nyti.ms/2wNBFqo

8:47 PM - 25 Sep 2017

## Define the universe of national news media domains ($n = 223$):

- breitbart.com
- foxnews.com
- nytimes.com
- cnn.com
- huffingtonpost.com
- etc.

# Example count matrix of tweeted news URLs:

| | thenation.com | huffingtonpost.com | washingtonpost.com | wsj.com | foxnews.com | breitbart.com | ⋯ |
|---|---|---|---|---|---|---|---|
| Ted Cruz (R) | 0 | **1** | **156** | **204** | **464** | **195** | ⋯ |
| Mitch McConnell (R) | 0 | **2** | **67** | **53** | **37** | 0 | ⋯ |
| Susan Collins (R) | 0 | **1** | **8** | **4** | 0 | 0 | ⋯ |
| Joe Manchin (D) | 0 | **4** | **13** | **2** | **3** | 0 | ⋯ |
| Alexandria Ocasio-Cortez (D) | **27** | **6** | **65** | **5** | **2** | 0 | ⋯ |
| Bernie Sanders (I) | **71** | **110** | **373** | **40** | **1** | 0 | ⋯ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |

## Homophily assumption

Social media users & politicians are more likely to tweet and retweet URLs of news media stories that are 'close' to themselves ideologically

## Model:

$$y_{img} \sim \text{NegBin}(\pi_{img}, \omega_m)$$

$$\pi_{img} = \exp(\alpha_i + \gamma_m - \underbrace{|\theta_i - \zeta_m|}_{\substack{\text{user-media} \\ \text{ideological distance}}}{}^2)$$

$y_{img}$ is the count of links from domain $m$ that have been tweeted by user $i$, who is affiliated with group (party) $g$

$\theta_{ig}$ latent ideology of user $i$ affiliated with group (party) $g$
$\zeta_m$ latent ideology of domain $m$
$\alpha_i$ user-specific intercept
$\gamma_m$ domain-specific intercept
$\omega_m$ domain-specific dispersion parameter

# Open-source software (`mediascores`):

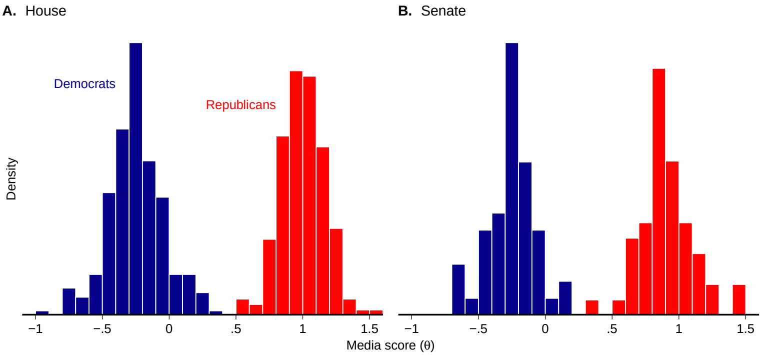
`github.com/smappnyu/mediascores`

Software vignette:

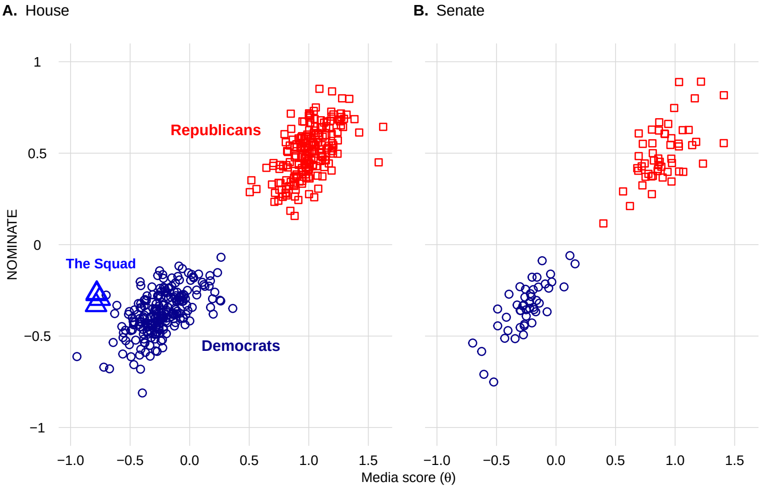
## Vignette

A vignette describing the library in greater detail—the model, model-fitting, convergence, and extracting quantities of interest—is available **here**.

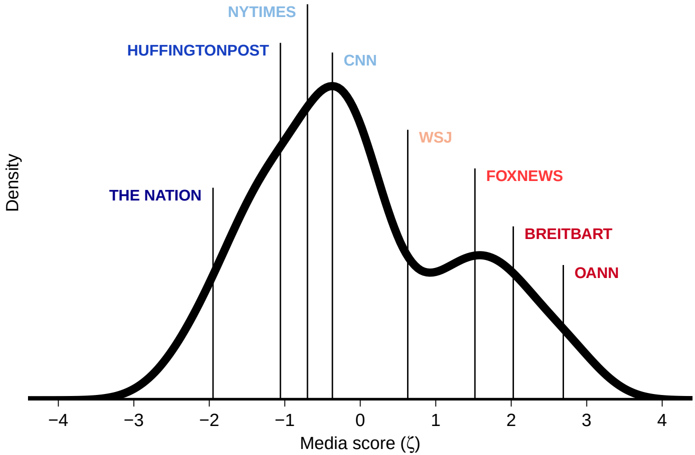
# News-sharing ideology of Members of Congress

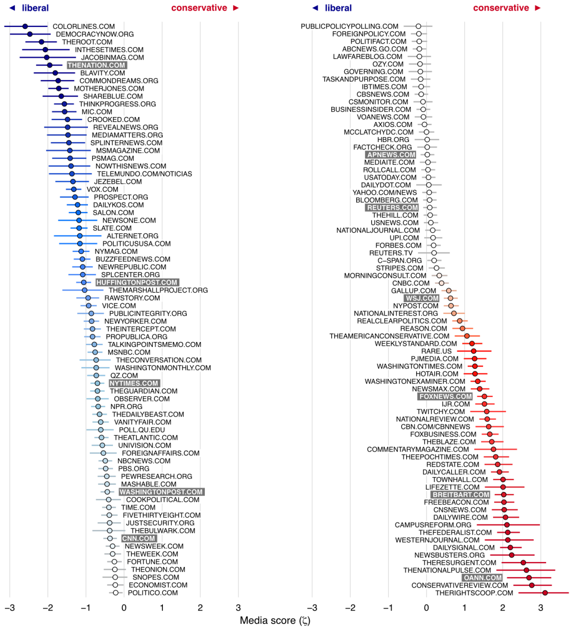


# News-sharing ideology & nominate

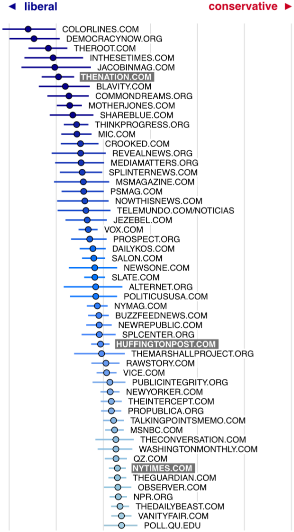


# News media ideology

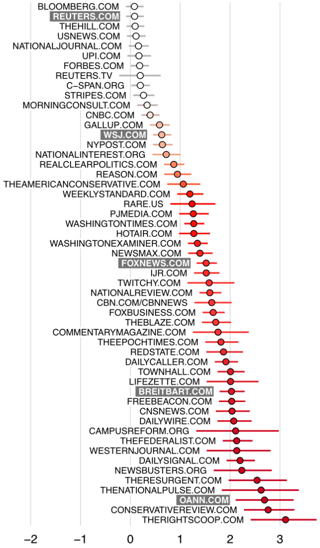



◄ liberal
conservative ►
COLORLINES.COM
DEMOCRACYNOW.ORG
THEROOT.COM
INTHESETIMES.COM
JACOBINMAG.COM
THENATION.COM
BLAVITY.COM
COMMONDREAMS.ORG
MOTHERJONES.COM
SHAREBLUE.COM
THINKPROGRESS.ORG
MIC.COM
CROOKED.COM
REVEALNEWS.ORG
MEDIAMATTERS.ORG
SPLINTERNEWS.COM
MSMAGAZINE.COM
PSMAG.COM
NOWTHISNEWS.COM
TELEMUNDO.COM/NOTICIAS
JEZEBEL.COM
VOX.COM
PROSPECT.ORG
DAILYKOS.COM
SALON.COM
NEWSONE.COM
SLATE.COM
ALTERNET.ORG
POLITICUSUSA.COM
NYMAG.COM
BUZZFEEDNEWS.COM
NEWREPUBLIC.COM
SPLCENTER.ORG
HUFFINGTONPOST.COM
THEMARSHALLPROJECT.ORG
RAWSTORY.COM
VICE.COM
PUBLICINTEGRITY.ORG
NEWYORKER.COM
THEINTERCEPT.COM
PROPUBLICA.ORG
TALKINGPOINTSMEMO.COM
MSNBC.COM
THECONVERSATION.COM
WASHINGTONMONTHLY.COM
QZ.COM
NYTIMES.COM
THEGUARDIAN.COM
OBSERVER.COM
NPR.ORG
THEDAILYBEAST.COM
VANITYFAIR.COM
POLL.QU.EDU
THEATLANTIC.COM
UNIVISION.COM
FOREIGNAFFAIRS.COM
NBCNEWS.COM
PBS.ORG
PEWRESEARCH.ORG
MASHABLE.COM
WASHINGTONPOST.COM
COOKPOLITICAL.COM
TIME.COM
FIVETHIRTYEIGHT.COM
JUSTSECURITY.ORG
THEBULWARK.COM
CNN.COM
NEWSWEEK.COM
THEWEEK.COM
FORTUNE.COM
THEONION.COM
SNOPES.COM
ECONOMIST.COM
POLITICO.COM
PUBLICPOLICYPOLLING.COM
FOREIGNPOLICY.COM
POLITIFACT.COM
ABCNEWS.GO.COM
LAWFAREBLOG.COM
OZY.COM
GOVERNING.COM
TASKANDPURPOSE.COM
IBTIMES.COM
CBSNEWS.COM
CSMONITOR.COM
BUSINESSINSIDER.COM
VOANEWS.COM
AXIOS.COM
MCCLATCHYDC.COM
HBR.ORG
FACTCHECK.ORG
APNEWS.COM
MEDIAITE.COM
ROLLCALL.COM
USATODAY.COM
DAILYDOT.COM
YAHOO.COM/NEWS
BLOOMBERG.COM
REUTERS.COM
THEHILL.COM
USNEWS.COM
NATIONALJOURNAL.COM
UPI.COM
FORBES.COM
REUTERS.TV
C–SPAN.ORG
STRIPES.COM
MORNINGCONSULT.COM
CNBC.COM
GALLUP.COM
WSJ.COM
NYPOST.COM
NATIONALINTEREST.ORG
REALCLEARPOLITICS.COM
REASON.COM
THEAMERICANCONSERVATIVE.COM
WEEKLYSTANDARD.COM
RARE.US
PJMEDIA.COM
WASHINGTONTIMES.COM
HOTAIR.COM
WASHINGTONEXAMINER.COM
NEWSMAX.COM
FOXNEWS.COM
IJR.COM
TWITCHY.COM
NATIONALREVIEW.COM
CBN.COM/CBNNEWS
FOXBUSINESS.COM
THEBLAZE.COM
COMMENTARYMAGAZINE.COM
THEEPOCHTIMES.COM
REDSTATE.COM
DAILYCALLER.COM
TOWNHALL.COM
LIFEZETTE.COM
BREITBART.COM
FREEBEACON.COM
CNSNEWS.COM
DAILYWIRE.COM
CAMPUSREFORM.ORG
THEFEDERALIST.COM
WESTERNJOURNAL.COM
DAILYSIGNAL.COM
NEWSBUSTERS.ORG
THERESURGENT.COM
THENATIONALPULSE.COM
OANN.COM
CONSERVATIVEREVIEW.COM
THERIGHTSCOOP.COM
−3 −2 −1 0 1 2 3
Media score (ζ)

# Liberal media



# Conservative media



## What explains the news-sharing behavior of politicians on social media?

- Legislators are strategic in how they present themselves to their constituents (Mayhew 1974, Fenno 1978)
- Legislators in more competitive districts are wary of alienating moderate voters
- Legislators in less competitive districts are less constrained, but also wary of alienating primary election voters

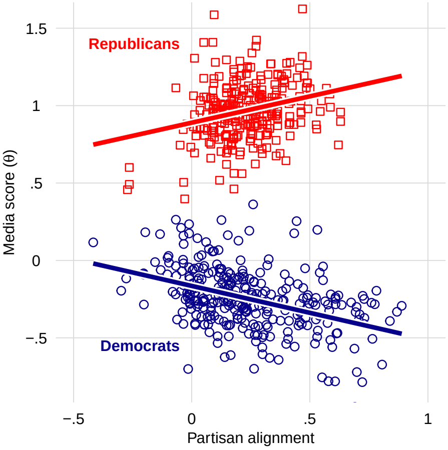
## Hypothesis:

**H**$_1$: Politicians in marginal districts will share less ideologically extreme news on social media than those in partisan-aligned districts

## Partisan alignment in district $i$ is measured by the 2016 presidential vote share margin:

$$partisanAlignment_i = voteShare_i^{(\text{Trump})} - voteShare_i^{(\text{Clinton})}$$

# Relationship between district partisan alignment and news-sharing ideology

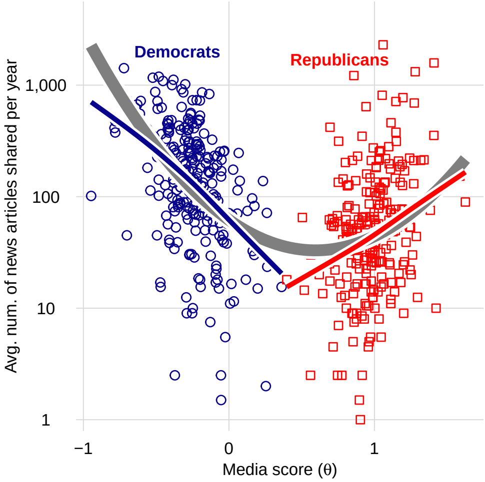


| | DV: Ideological extremity of news sharing | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| District alignment | 0.317 | 0.309 | 0.108 | 0.130 |
| | (0.041) | (0.042) | (0.041) | (0.044) |
| Republican | | 0.009 | 0.009 | 0.008 |
| | | (0.016) | (0.015) | (0.015) |
| Senator | | −0.017 | −0.038 | −0.036 |
| | | (0.021) | (0.019) | (0.019) |
| Nominate score | | | 0.691 | 0.591 |
| | | | (0.058) | (0.092) |
| Nominate score × Republican | | | | 0.161 |
| | | | | (0.114) |
| Intercept | −0.078 | −0.077 | −0.025 | −0.030 |
| | (0.013) | (0.016) | (0.015) | (0.016) |
| N | 527 | 527 | 527 | 527 |

Standard errors in parentheses.

**Table 2: Relationship between the ideological extremity of news sharing and district/state alignment.** Standard errors in parentheses. All estimates of the coefficient “District alignment” are statistically significant at the 99% level.

## Lastly, who dominates news and information sharing on social media?



## Conclusions

- Who controls policy debates matters, both for agenda-setting and position-taking
- Moderate politicians are heavily constrained to the extent that setting the agenda and position-taking can be detrimental if they are in a marginal district
- Politicians in marginal districts thus remain quiet on social media, even though the content they share is moderate
- Those without such constraints, however, can freely share ideologically extreme content without an electoral penalty, and may benefit in primary elections

## How to apply the method in practice

1. Collect or create a list of news media domain names (e.g. nytimes.com, wsj.com)
2. Collect or create the user IDs of users and/or politicians who you want to estimate media-sharing ideology for
3. Collect the timelines for all of these users
4. Unshorten all URLs if necessary (e.g. https://bit.ly/fd312kj)
5. Use regular expressions to create a user-domain count matrix based on the links each user has shared
    - (Potentially) remove quote tweets, and links to non-political stories
6. Fit the model