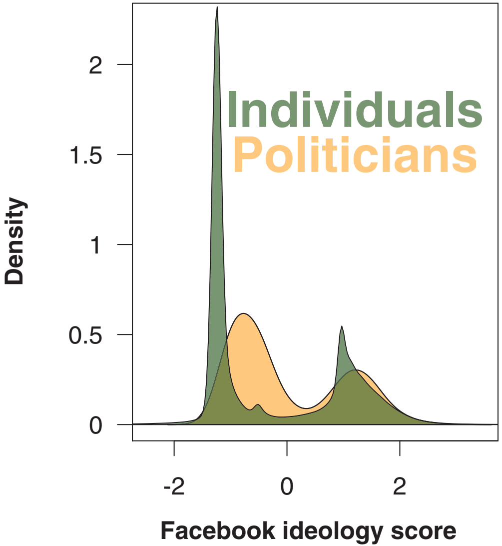
Ideology Estimation, Media Slant, and Opinion Segregation: Facebook as a Social Barometer

Keng-Chi Chang

National Taiwan University

2017-09-15

Motivation: Bond and Messing (2015)



Highlights

- Specify potential ideological universe
- Select possible US users
- Place different political actors on the same ideological spectrum (politicians, figures, news outlets, parties, and interest groups)
- Replicate mass ideology distribution at national and state level
- Allow time and text dimensions to explore
- All using publicly available open data

Introduction

- Best place to get interactions between differnt actors
- Extract information from the major action: "like"
- e.g. What can we infer if two pages share many users?
- Joint measure of ideologies of political elites, news outlets, interest groups, and ordinary citizens
- cf. surveys: Revealed preference, low cost, real time
- Past papers only look at *following* of pages, we look at *like* of posts (adds time and post content dimension)

Literature: Ideology Measures

- Need “bridges” to connect different actors
  - Politicians: Poole and Rosenthal (1985), Clinton et al. (2004)
  - Media-Politicians: Groseclose and Milyo (2005)
  - Media-Citizens: Gentzkow and Shapiro (2011)
  - Politician-Citizens: Bonica (2014)
- Lack of joint ideological measures across all these actors
- Social media acts as brigdes for different actors
- Both Bond and Messing (2015) (Facebook) and Barberá (2015) (Twitter) only consider political elites

Model and Traditional Estimation

- Assume that user i's ideological position is θ_i and politician/media j's position ϕ_j
- Assume also that the probability i likes j's post is proportional to the negative distance between θ_i and ϕ_j

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

- Traditionally this is solved by Markov-Chain Monte Carlo (MCMC) to maximize posterior density (MLE)

$$\left\{\hat{\theta}_i, \hat{\phi}_j\right\} = \arg\max_{\theta_i, \phi_j} \prod_{i \in \text{user}} \prod_{j \in \text{page}} \text{logit}^{-1}\left(\pi_{ij}\right)^{y_{ij}} \left(1 - \text{logit}^{-1}\left(\pi_{ij}\right)\right)^{1 - y_{ij}}$$

- This is slow for large numbers of parameters

Methodology: Dimension Reduction

- Similar to Heckman and Snyder (1997) (Congress), Bond and Messing (2015) (Facebook), and Barberá (2015) (Twitter), we use dimension reduction to recover the latent ideological space
- Barberá et al. (2015) also shows that simulations and dimension reduction (Correspondence Analysis) generate very similar results ($\rho = 0.98$)
- We show that Correspondence Analysis and Principal Component Analysis (2 stage) generate similar results ($\rho > 0.94$)
- Drawbacks:
  - What are the dimensions? $\rightsquigarrow$ Guess and verify
  - How many dimensions to consider? $\rightsquigarrow$ Scree plot

Select Meaningful Data

- Select fan pages mentioned two major presidential candidates
- Calculate likes, comments, shares and select top 1000 pages
- Also include past and present national politicians (Sen, Rep, Gov)
- Facebook open data do not give any personal information
- Select users ever liked national politicians in 2015 and 2016

Data Summary (Main Sample)

Time Period	2015-01-01 to 2016-11-30
Total Reactions	19,085,783,534
US User Likes	16,180,488,916
Total Users	366,840,068
US Users	29,412,610
Total Posts	24,788,093
Total Pages	2132
Politician	1225
News Outlets	560
Political Groups	211
Other Public Figures	93
Others	43

Estimation: Affiliation Matrix

- First we build the affiliation matrix A, which contains number of shared users between pages

	Trump	FoxNews	TeaParty	Clinton	CNN	NYTimes
Trump	2243216	1078513	128225	32731	120963	25842
FoxNews	1078513	2449174	148016	87084	186850	63401
TeaParty	128225	148016	242089	1528	10738	2162
Clinton	32731	87084	1528	1768980	351210	367021
CNN	120963	186850	10738	351210	1201156	216163
NYTimes	25842	63401	2162	367021	216163	986613

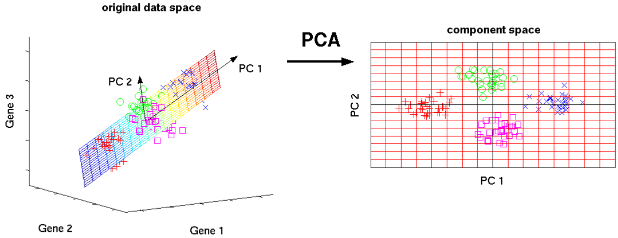
Estimation: Agreement Matrix

- Agreement matrix G is computed by $g_{ij} = a_{ij}/a_{ii}$
- Ex. 0.44 is (Trump & Fox) / Fox
- Can interpret each column as feature and row as observaton
- Interpretation: Col 1 is how each row similar to "Trump" feature

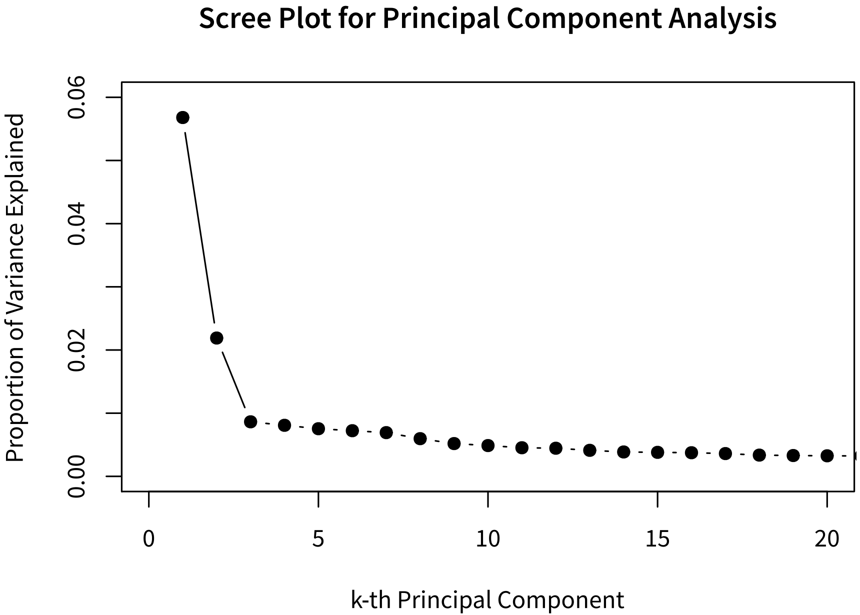
	Trump	FoxNews	TeaParty	Clinton	CNN	NYTimes
Trump	1.00	0.48	0.06	0.01	0.05	0.01
FoxNews	0.44	1.00	0.06	0.04	0.08	0.03
TeaParty	0.53	0.61	1.00	0.01	0.04	0.01
Clinton	0.02	0.05	0.00	1.00	0.20	0.21
CNN	0.10	0.16	0.01	0.29	1.00	0.18
NYTimes	0.03	0.06	0.00	0.37	0.22	1.00

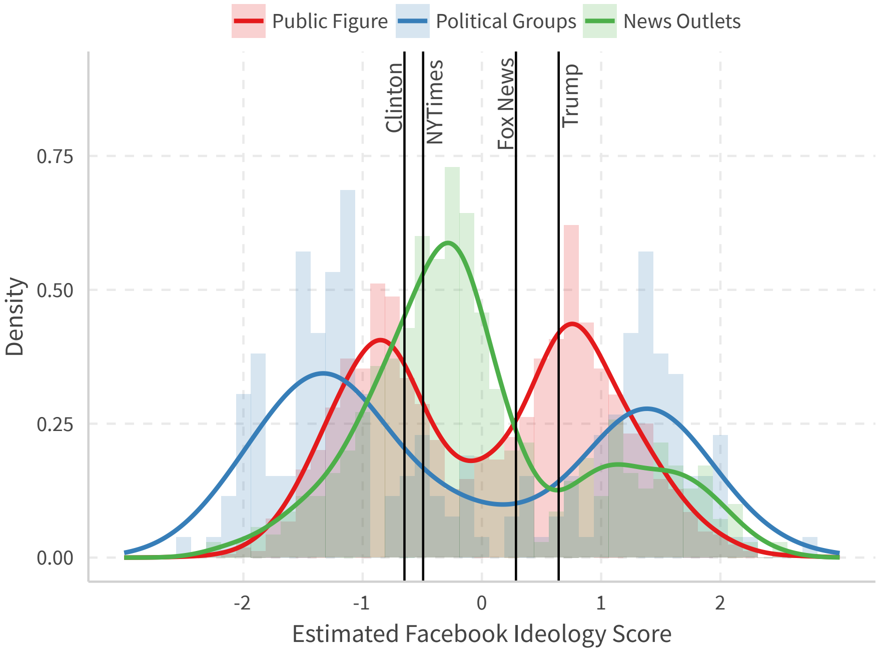
Estimation: Compute PCA

- Compute the principal components of G after standardizing
- The first principal component is the dimension that can explain the laregest variance, guess that positions on this dimension can represent "ideology" of pages

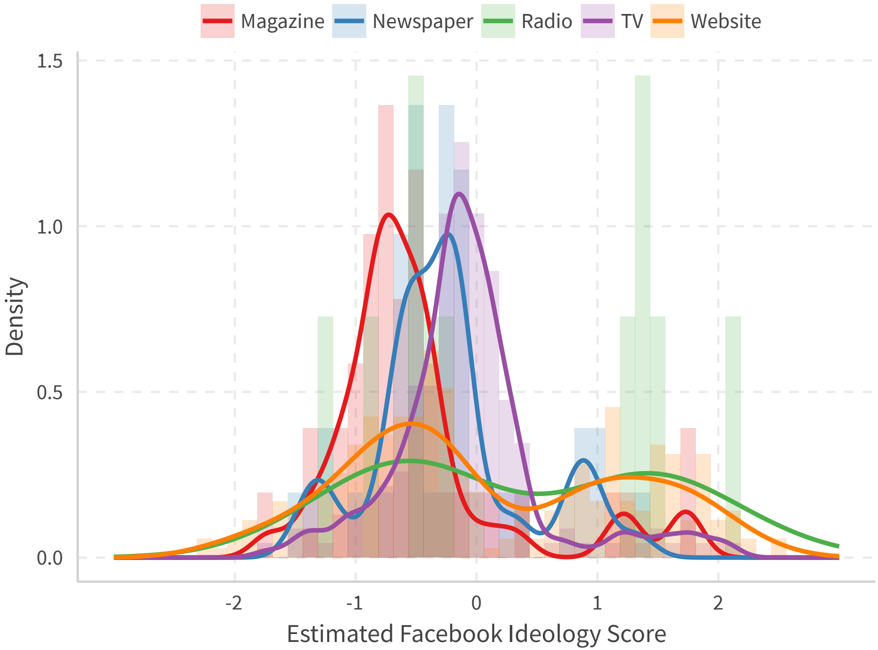


- Calculate the ideological position of users by computing the means of the ideologies of the pages they like (minimizer under Euclidean norm)

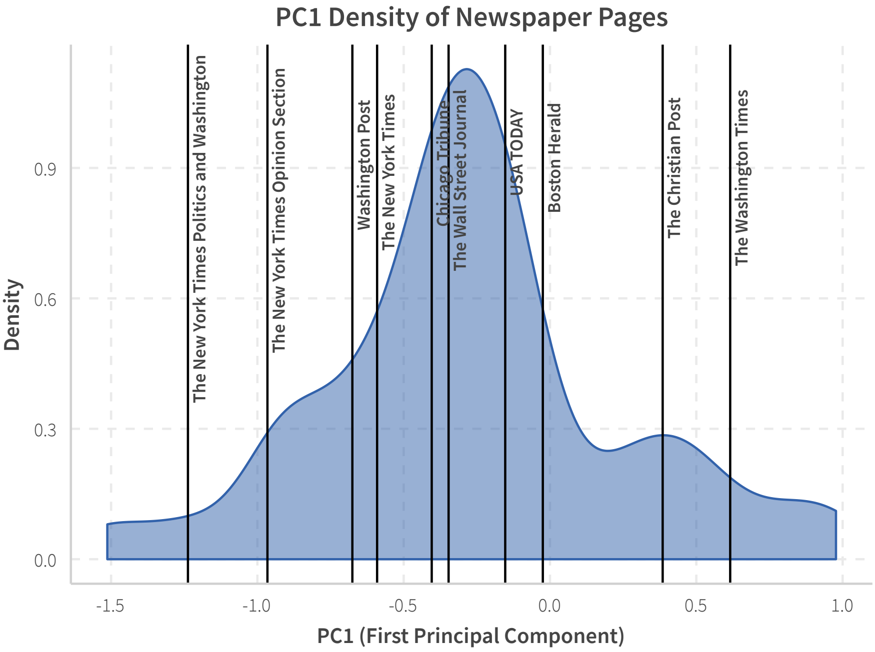

Scree Plot for Principal Component Analysis
Proportion of Variance Explained
0.00
0.02
0.04
0.06
0
5
10
15
20
k-th Principal Component

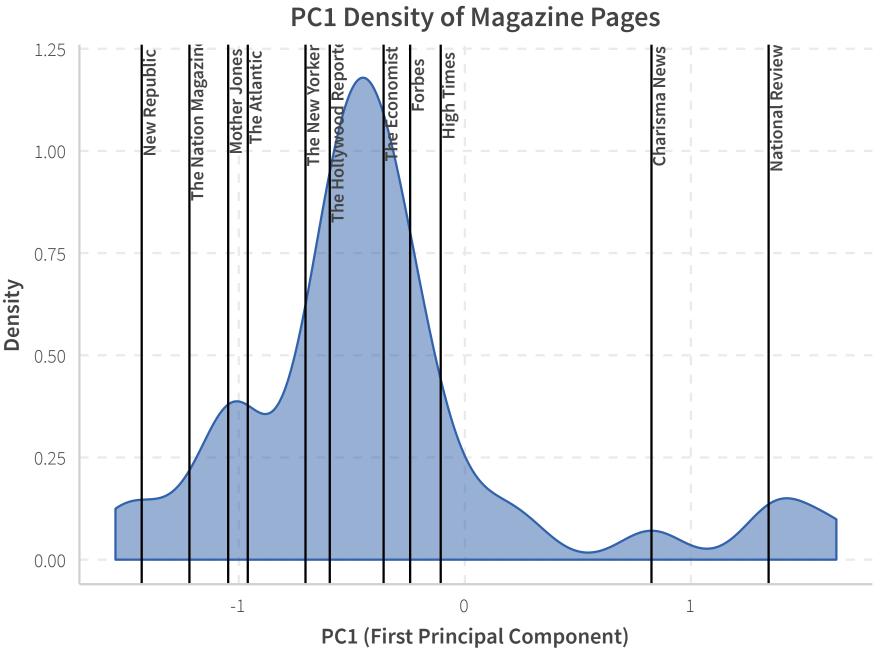


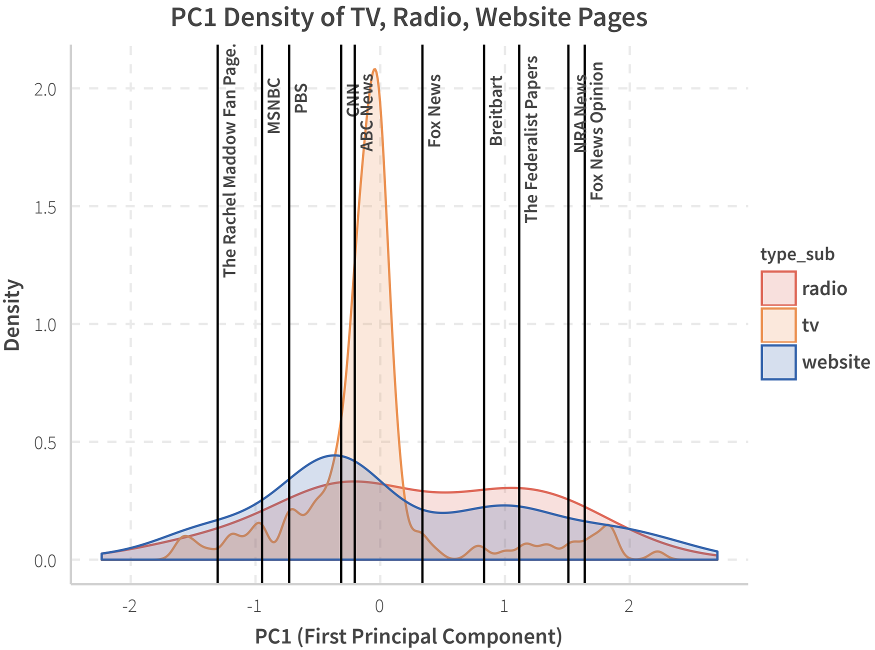
Public Figure
Political Groups
News Outlets
Clinton
NYTimes
Fox News
Trump
Density
0.75
0.50
0.25
0.00
-2
-1
0
1
2
Estimated Facebook Ideology Score

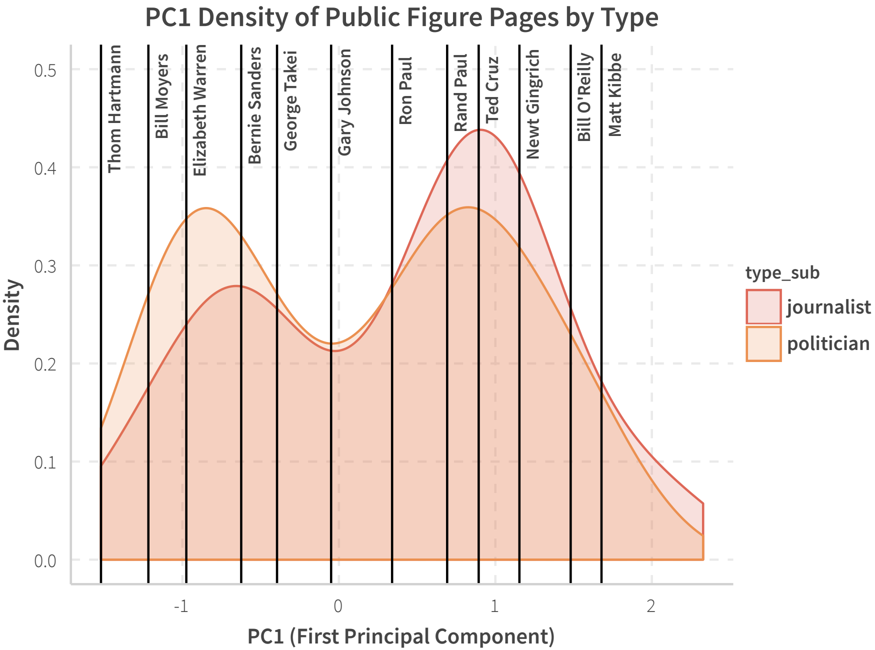


Magazine
Newspaper
Radio
TV
Website
Density
1.5
1.0
0.5
0.0
-2
-1
0
1
2
Estimated Facebook Ideology Score

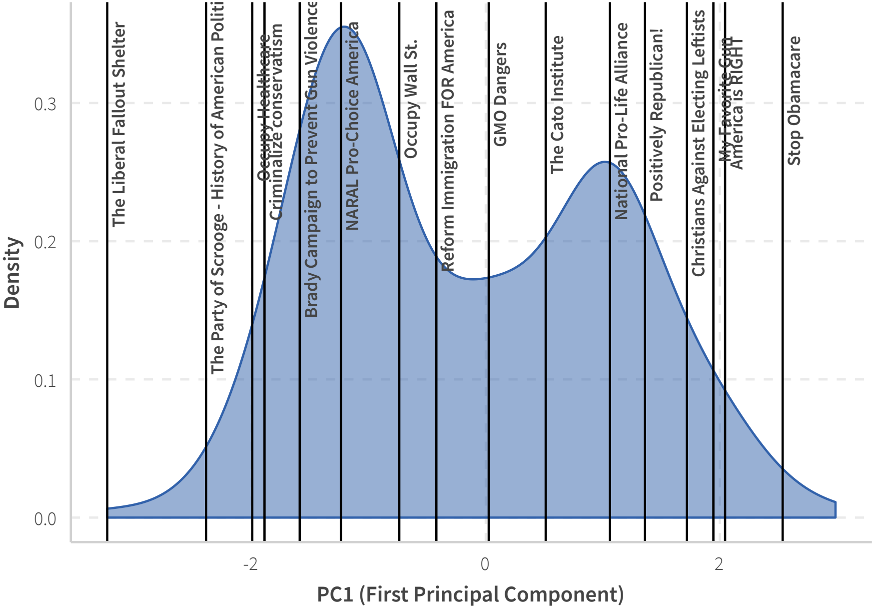

PC1 Density of Newspaper Pages
Density
PC1 (First Principal Component)
0.0
0.3
0.6
0.9
-1.5
-1.0
-0.5
0.0
0.5
1.0
The New York Times Politics and Washington
The New York Times Opinion Section
Washington Post
The New York Times
Chicago Tribune
The Wall Street Journal
USA TODAY
Boston Herald
The Christian Post
The Washington Times

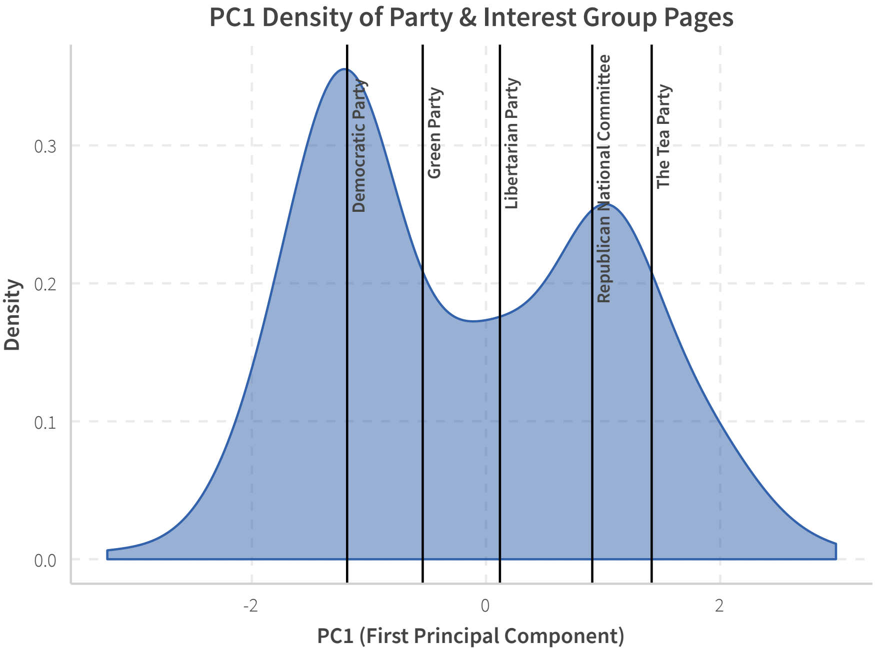

PC1 Density of Magazine Pages
Density
PC1 (First Principal Component)
1.25
1.00
0.75
0.50
0.25
0.00
-1
0
1
New Republic
The Nation Magazine
Mother Jones
The Atlantic
The New Yorker
The Hollywood Reporter
The Economist
Forbes
High Times
Charisma News
National Review


PC1 Density of TV, Radio, Website Pages
Density
PC1 (First Principal Component)
2.0
1.5
1.0
0.5
0.0
-2
-1
0
1
2
The Rachel Maddow Fan Page.
MSNBC
PBS
CNN
ABC News
Fox News
Breitbart
The Federalist Papers
NRA News
Fox News Opinion
type_sub
radio
tv
website

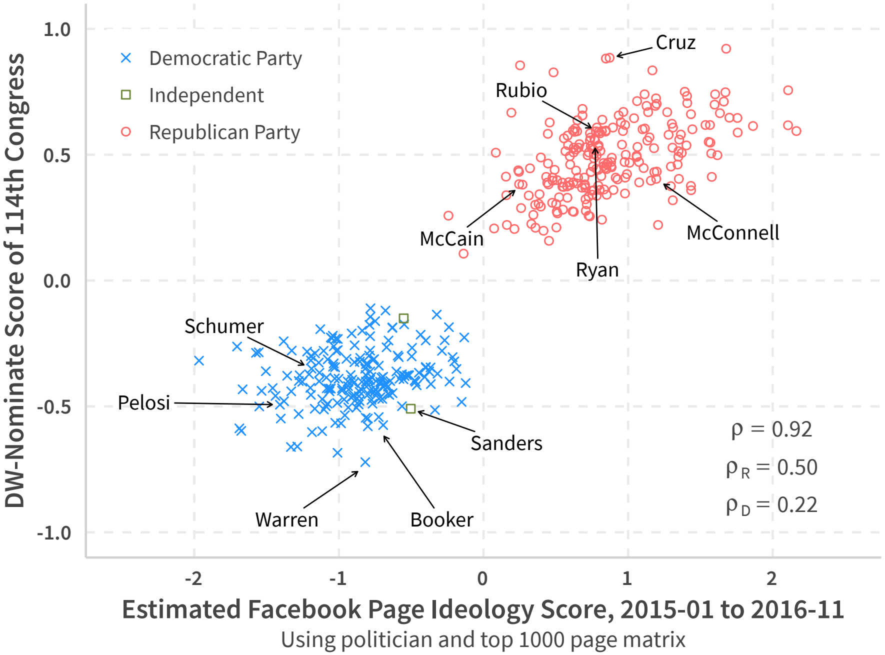

PC1 Density of Public Figure Pages by Type
Thom Hartmann
Bill Moyers
Elizabeth Warren
Bernie Sanders
George Takei
Gary Johnson
Ron Paul
Rand Paul
Ted Cruz
Newt Gingrich
Bill O'Reilly
Matt Kibbe
Density
0.0
0.1
0.2
0.3
0.4
0.5
-1
0
1
2
PC1 (First Principal Component)
type_sub
journalist
politician

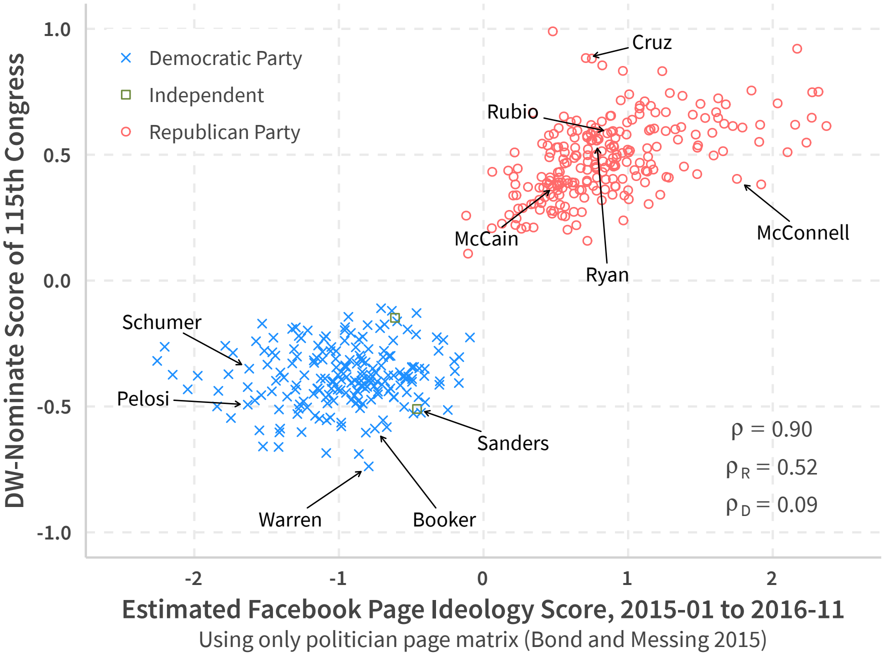
PC1 Density of Party & Interest Group Pages



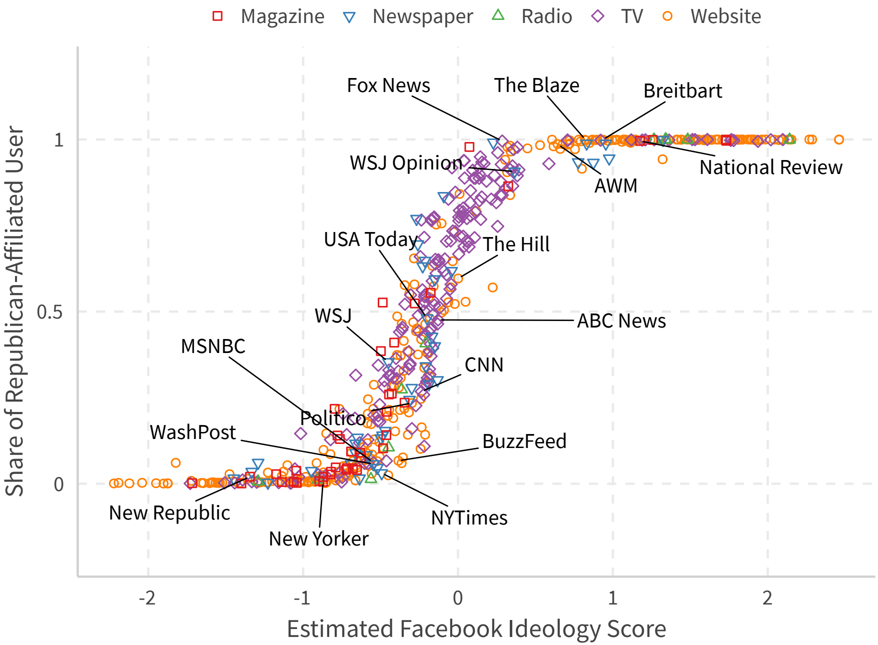


PC1 Density of Party & Interest Group Pages
Democratic Party
Green Party
Libertarian Party
Republican National Committee
The Tea Party
Density
0.3
0.2
0.1
0.0
-2
0
2
PC1 (First Principal Component)

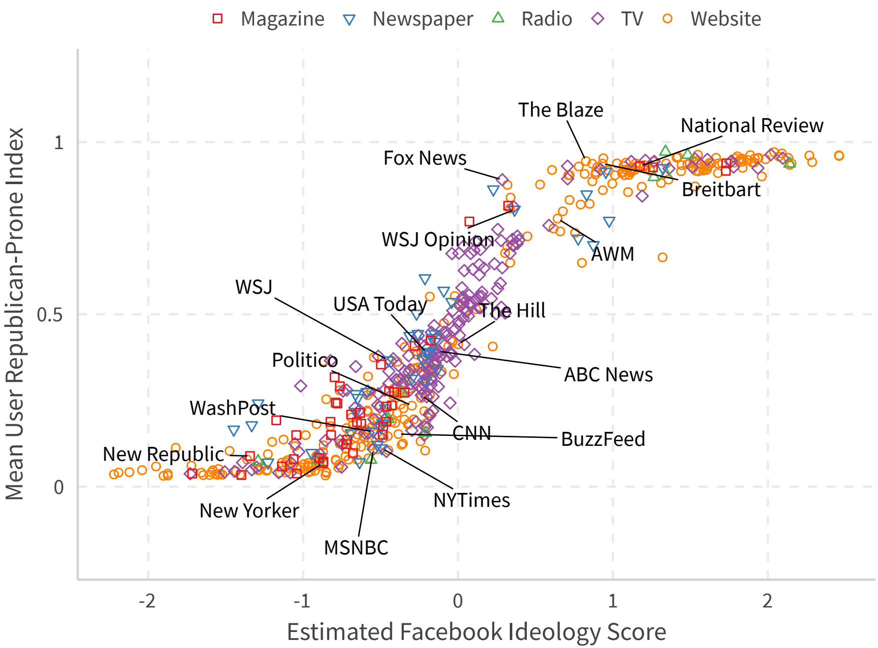

Democratic Party
Independent
Republican Party
Cruz
Rubio
McCain
Ryan
McConnell
Schumer
Pelosi
Sanders
Warren
Booker
ρ = 0.92
ρ_R = 0.50
ρ_D = 0.22
DW-Nominate Score of 114th Congress
1.0
0.5
0.0
-0.5
-1.0
-2
-1
0
1
2
Estimated Facebook Page Ideology Score, 2015-01 to 2016-11
Using politician and top 1000 page matrix



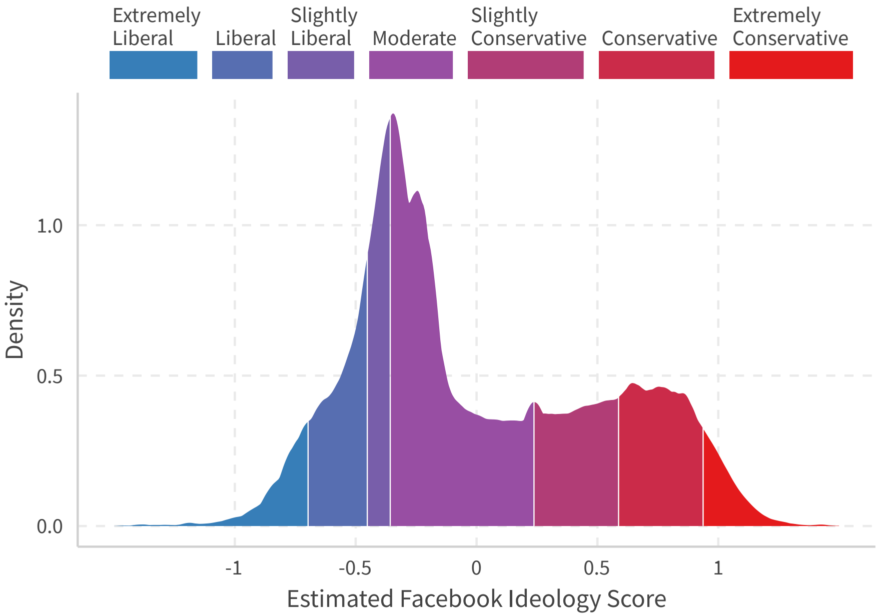
DW-Nominate Score of 115th Congress
1.0
0.5
0.0
-0.5
-1.0
-2
-1
0
1
2
Democratic Party
Independent
Republican Party
Cruz
Rubio
McCain
Ryan
McConnell
Schumer
Pelosi
Sanders
Warren
Booker
$\rho = 0.90$
$\rho_R = 0.52$
$\rho_D = 0.09$
Estimated Facebook Page Ideology Score, 2015-01 to 2016-11
Using only politician page matrix (Bond and Messing 2015)



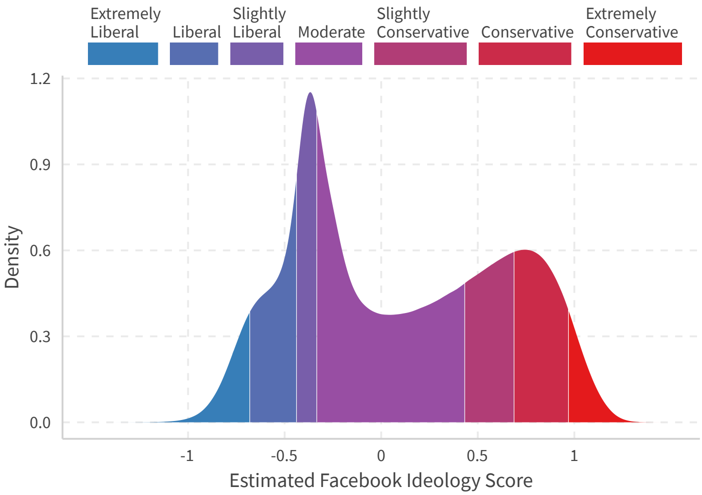
Magazine
Newspaper
Radio
TV
Website
Fox News
The Blaze
Breitbart
WSJ Opinion
AWM
National Review
USA Today
The Hill
WSJ
ABC News
MSNBC
CNN
Politico
WashPost
BuzzFeed
New Republic
New Yorker
NYTimes
Share of Republican-Affiliated User
Estimated Facebook Ideology Score
1
0.5
0
-2
-1
0
1
2



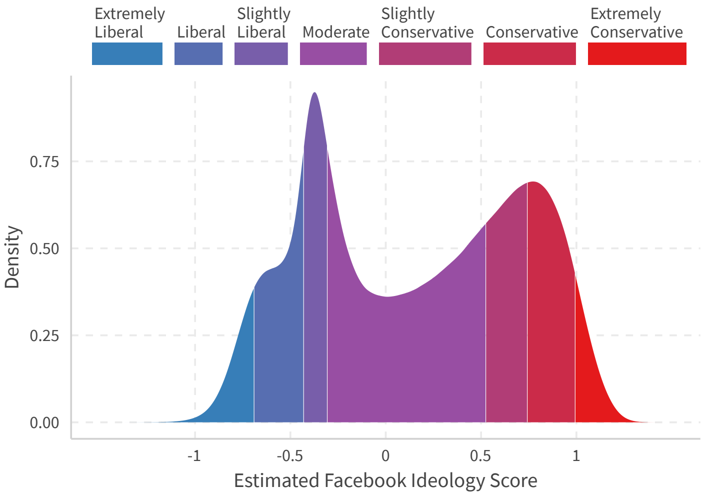
Magazine
Newspaper
Radio
TV
Website
Mean User Republican-Prone Index
1
0.5
0
Estimated Facebook Ideology Score
-2
-1
0
1
2
The Blaze
National Review
Fox News
Breitbart
WSJ Opinion
AWM
WSJ
USA Today
The Hill
Politico
ABC News
WashPost
CNN
BuzzFeed
New Republic
NYTimes
New Yorker
MSNBC

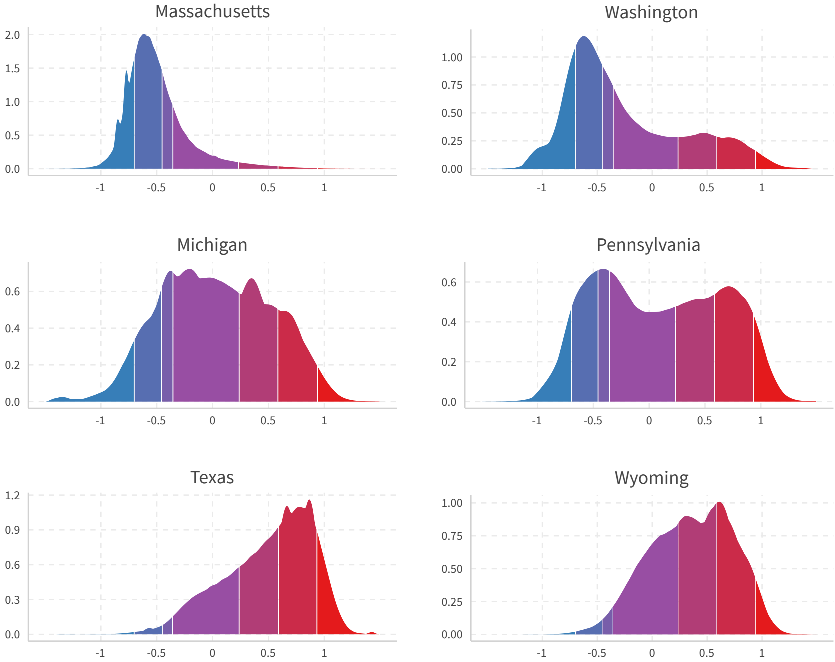

Extremely Liberal
Liberal
Slightly Liberal
Moderate
Slightly Conservative
Conservative
Extremely Conservative
Density
1.0
0.5
0.0
-1
-0.5
0
0.5
1
Estimated Facebook Ideology Score

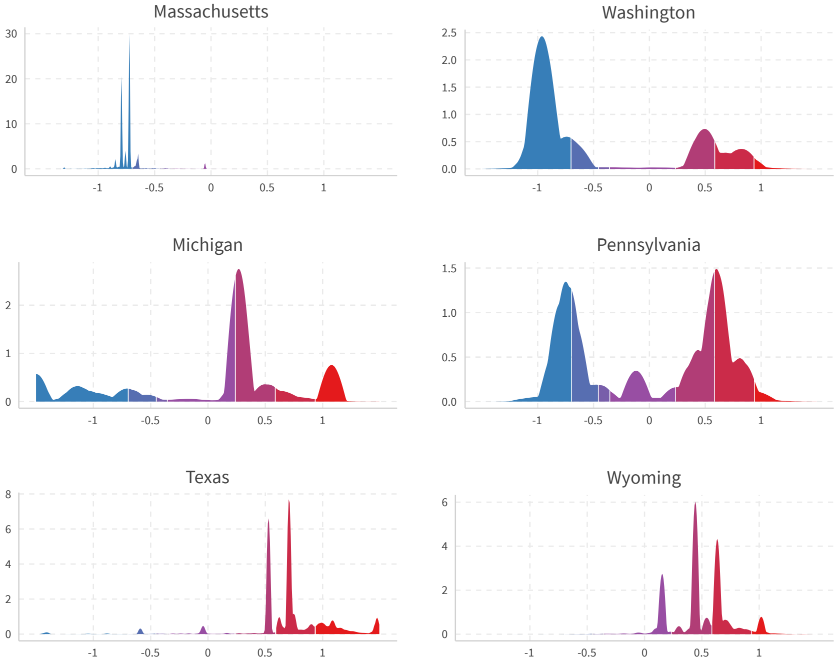
Users Like More than 10 Pages and Posts



Users Like More than 20 Pages and Posts

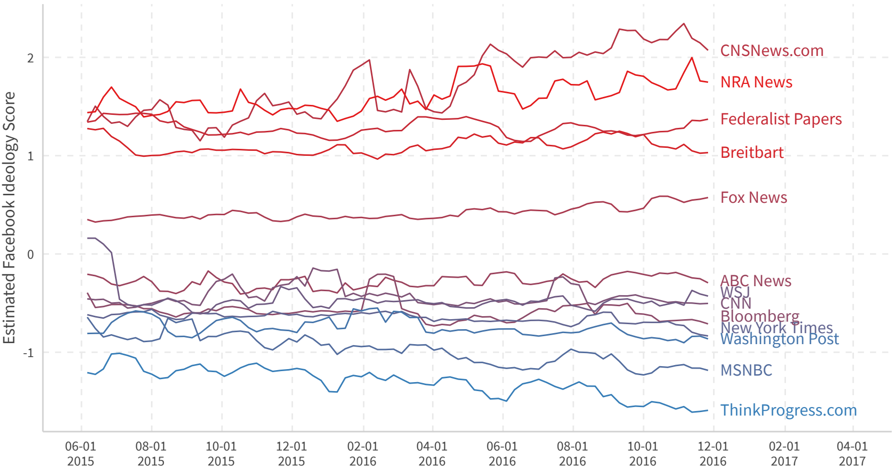



Massachusetts
2.0
1.5
1.0
0.5
0.0
-1
-0.5
0
0.5
1
Washington
1.00
0.75
0.50
0.25
0.00
-1
-0.5
0
0.5
1
Michigan
0.6
0.4
0.2
0.0
-1
-0.5
0
0.5
1
Pennsylvania
0.6
0.4
0.2
0.0
-1
-0.5
0
0.5
1
Texas
1.2
0.9
0.6
0.3
0.0
-1
-0.5
0
0.5
1
Wyoming
1.00
0.75
0.50
0.25
0.00
-1
-0.5
0
0.5
1

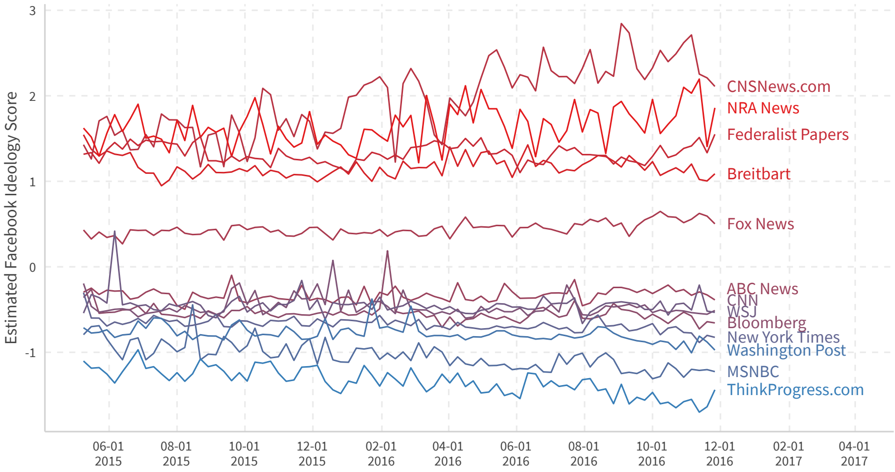

Massachusetts
30
20
10
0
-1
-0.5
0
0.5
1
Washington
2.5
2.0
1.5
1.0
0.5
0.0
-1
-0.5
0
0.5
1
Michigan
2
1
0
-1
-0.5
0
0.5
1
Pennsylvania
1.5
1.0
0.5
0.0
-1
-0.5
0
0.5
1
Texas
8
6
4
2
0
-1
-0.5
0
0.5
1
Wyoming
6
4
2
0
-1
-0.5
0
0.5
1

Applications

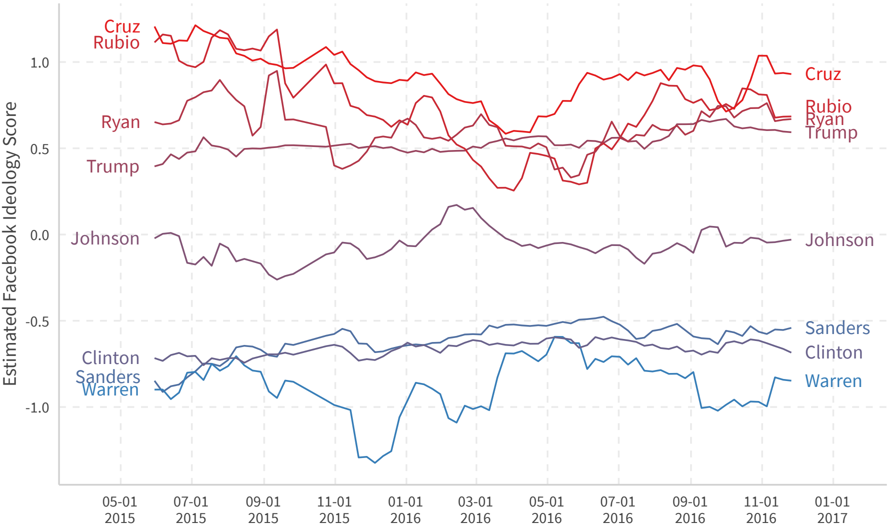
- Time Dimension: Polarization and Spatial Voting
- Text Dimension: Echo Chambers
- Election Forecasting
- Opinion or Ideological Segregation



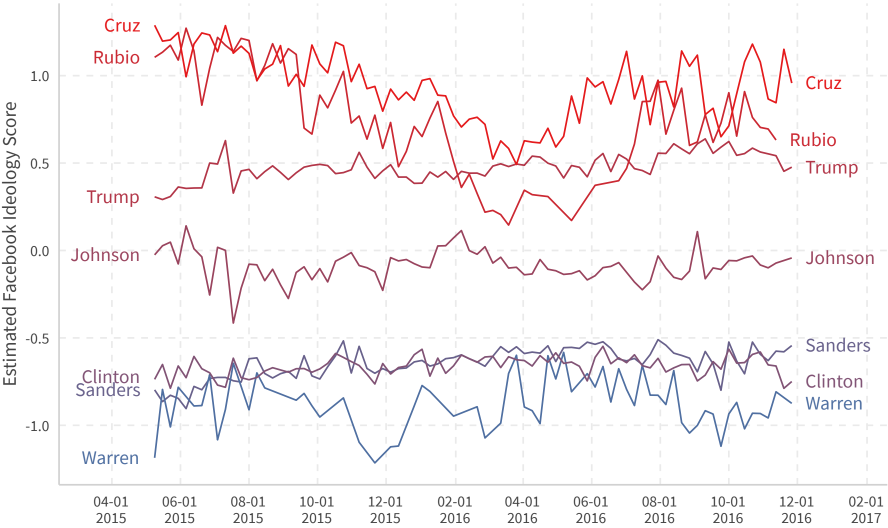
4-week window



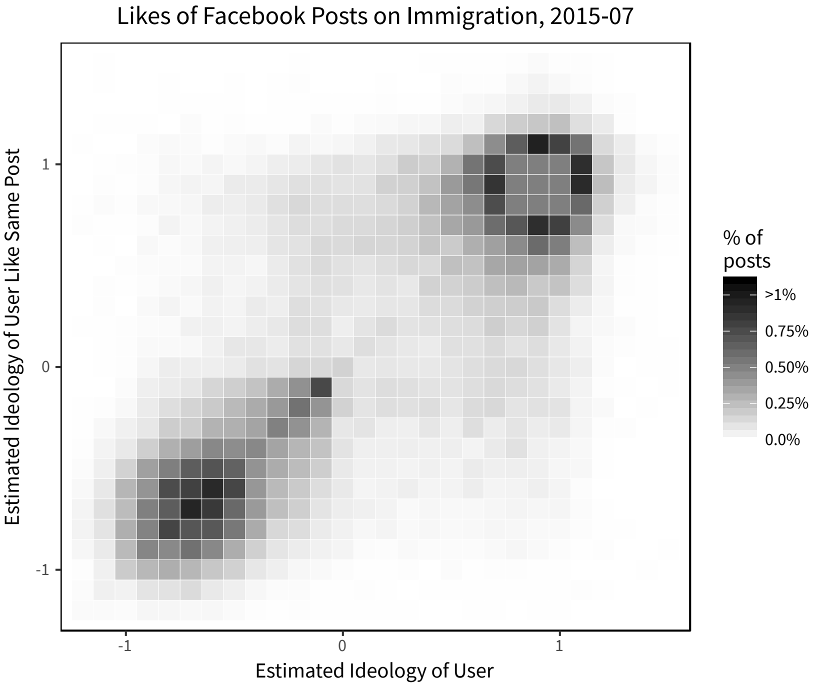
1-week window

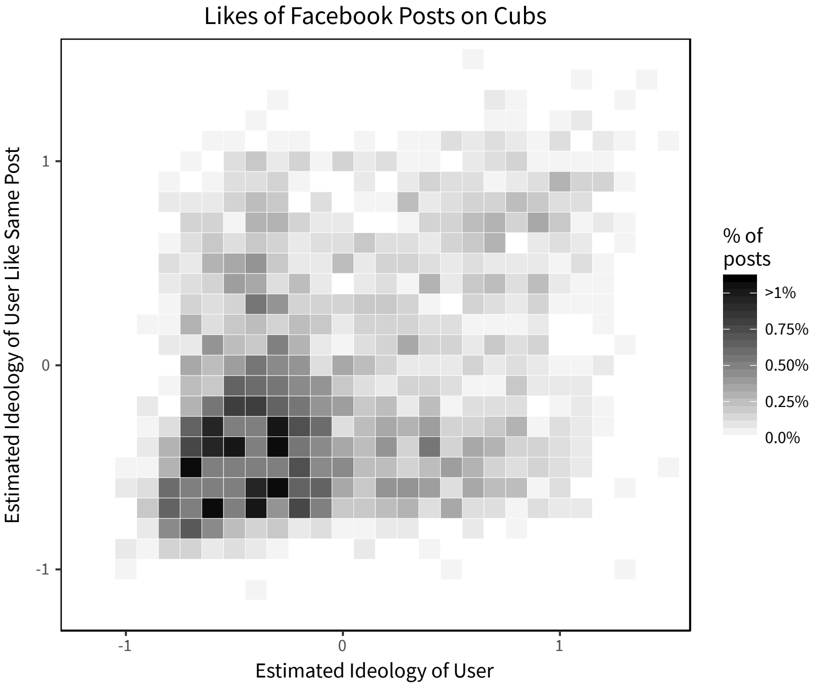


4-week window

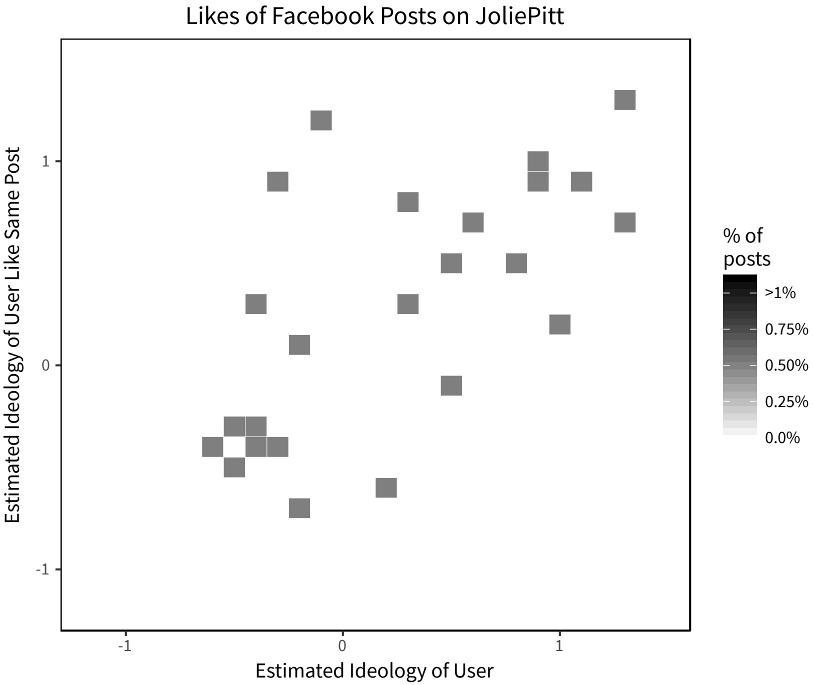


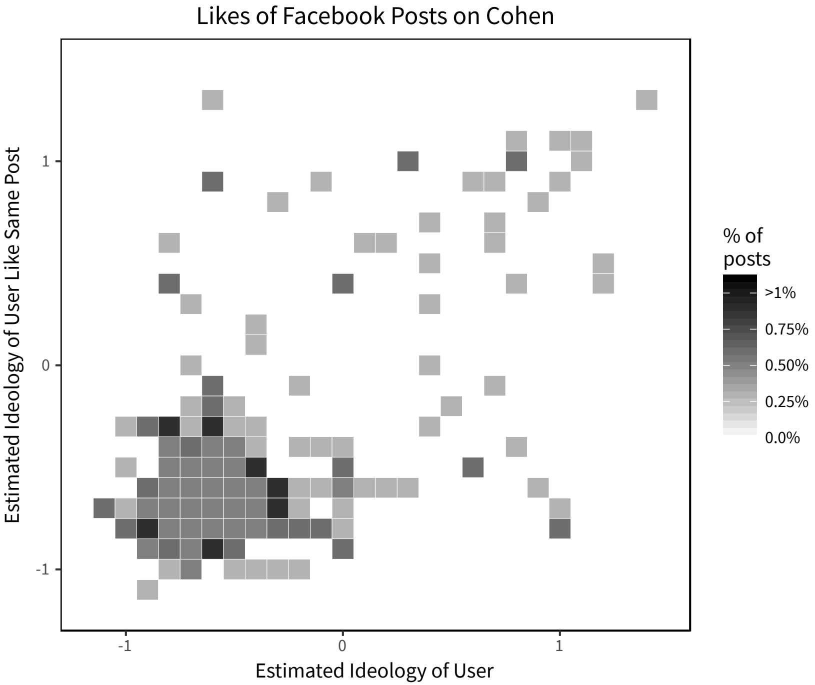
1-week window


Likes of Facebook Posts on Immigration, 2015-07
Estimated Ideology of User Like Same Post
1
0
-1
-1
0
1
Estimated Ideology of User
% of posts
>1%
0.75%
0.50%
0.25%
0.0%

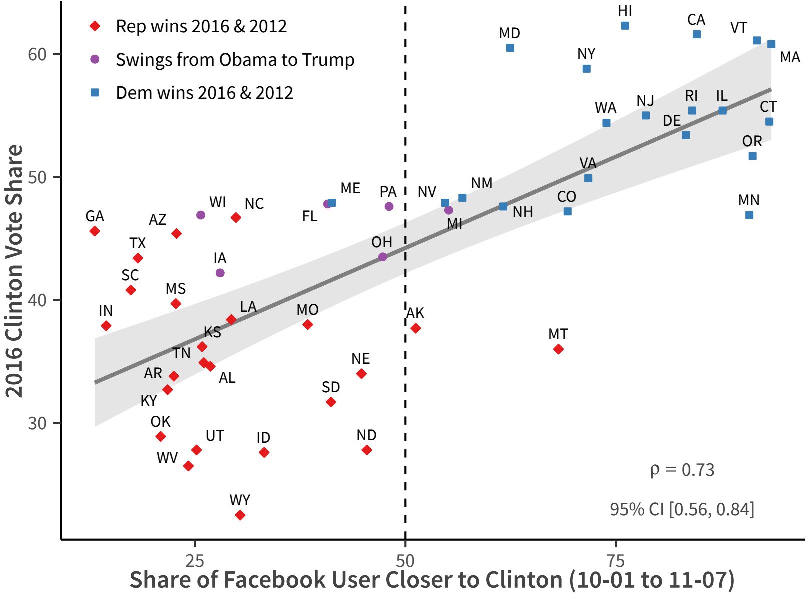


Likes of Facebook Posts on Cubs
Estimated Ideology of User Like Same Post
1
0
-1
Estimated Ideology of User
-1
0
1
% of posts
>1%
0.75%
0.50%
0.25%
0.0%


Likes of Facebook Posts on JoliePitt
Estimated Ideology of User Like Same Post
1
0
-1
-1
0
1
Estimated Ideology of User
% of posts
>1%
0.75%
0.50%
0.25%
0.0%


Likes of Facebook Posts on Cohen
Estimated Ideology of User Like Same Post
1
0
-1
-1
0
1
Estimated Ideology of User
% of posts
>1%
0.75%
0.50%
0.25%
0.0%

Predicting Election Outcomes



Compare with Major Election Forecasts

State	E.V.†	Winner	FB	538	NYT	PEC*
Wisconsin	10	Trump	○	×	×	×
Iowa	6	Trump	○	○	○	○
Florida	29	Trump	○	×	×	×
Pennsylvania	20	Trump	○	×	×	×
Ohio	18	Trump	○	○	○	○
Michigan	16	Trump	×	×	×	×
Maine	2	Clinton	×	○	○	○
Alaska	3	Clinton	×	○	○	○
Montana	3	Trump	×	○	○	○
Trump's Electoral Vote		306	292	235	216	215

† Electoral Votes. * Princeton Election Consortium.

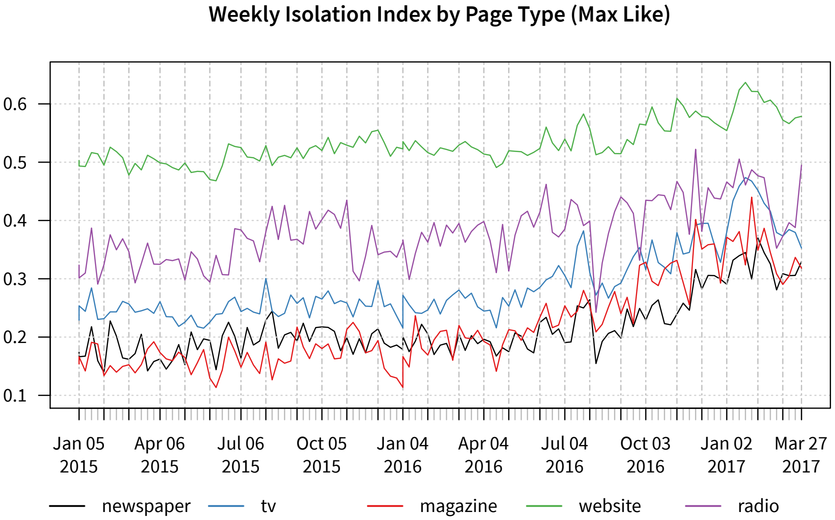
Ideological Segregation

- Parallel indexes as in Gentzkow and Shapiro (2011)

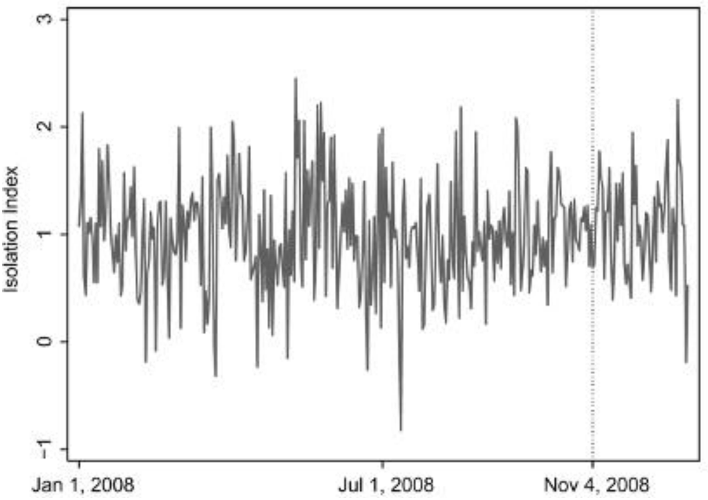
$$S_m = \sum_{j \in J_m} \left(\frac{\text{cons}_j}{\text{cons}_m} \cdot \frac{\text{cons}_j}{\text{visits}_j} \right) - \sum_{j \in J_m} \left(\frac{\text{lib}_j}{\text{lib}_m} \cdot \frac{\text{cons}_j}{\text{visits}_j} \right)$$

- For each news outlet j of type m (news outlets, politicians; tv, newspapers; etc), we can calculate the share of conservative daily visitors and weight by the relative importance of that page inside the conservative or liberal campaign
- 0: All conservatives and liberals visits the same page
- 1: Conservatives only visits all conservative pages

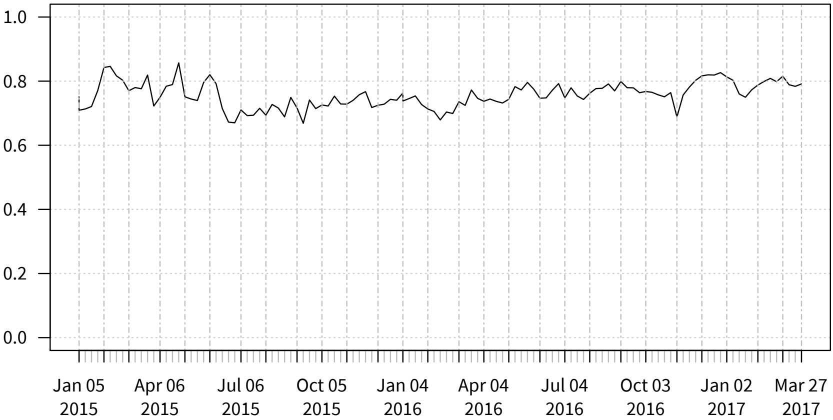
Ideological Segregation

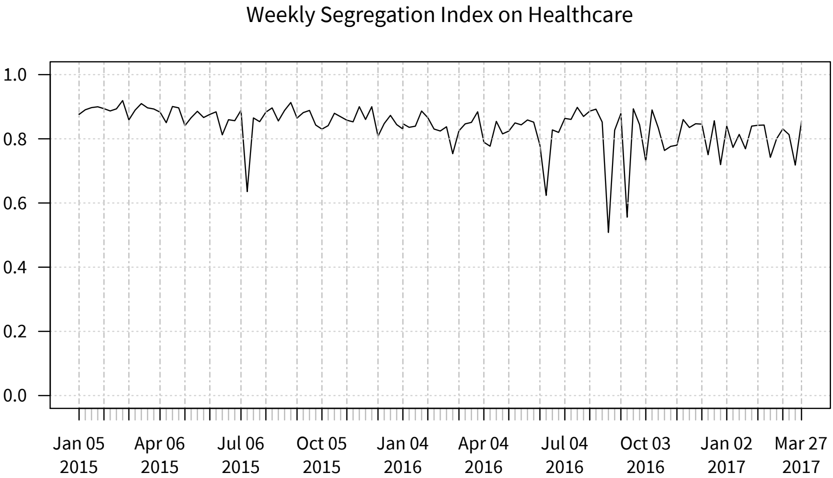


Gentzkow and Shapiro (2015)

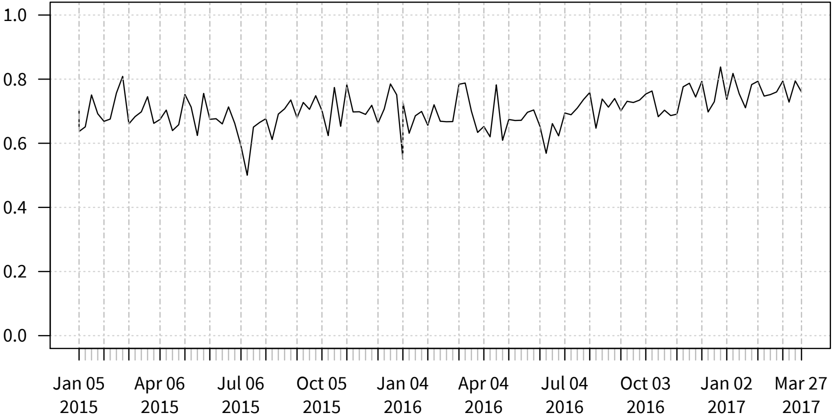


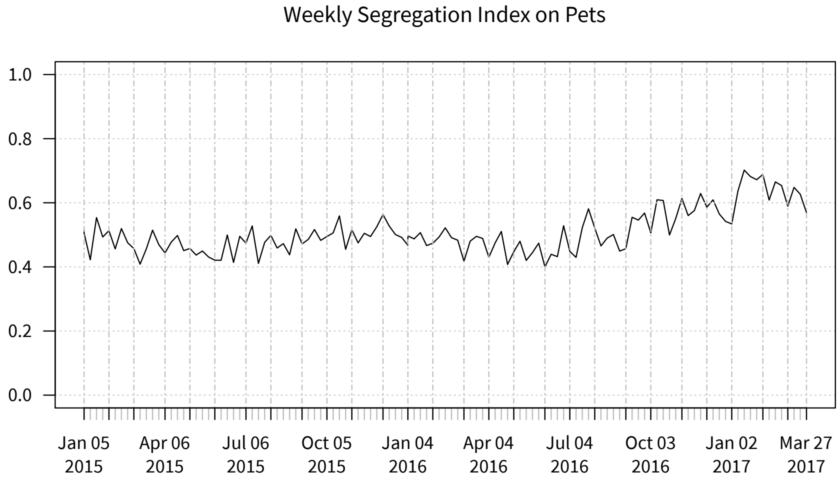
Weekly Segregation Index on Election



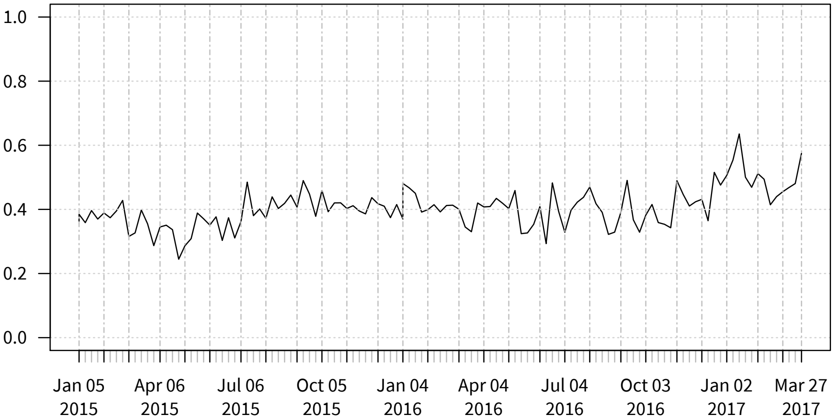

Weekly Segregation Index on Healthcare
1.0
0.8
0.6
0.4
0.2
0.0
Jan 05 2015
Apr 06 2015
Jul 06 2015
Oct 05 2015
Jan 04 2016
Apr 04 2016
Jul 04 2016
Oct 03 2016
Jan 02 2017
Mar 27 2017

Weekly Segregation Index on Immigration

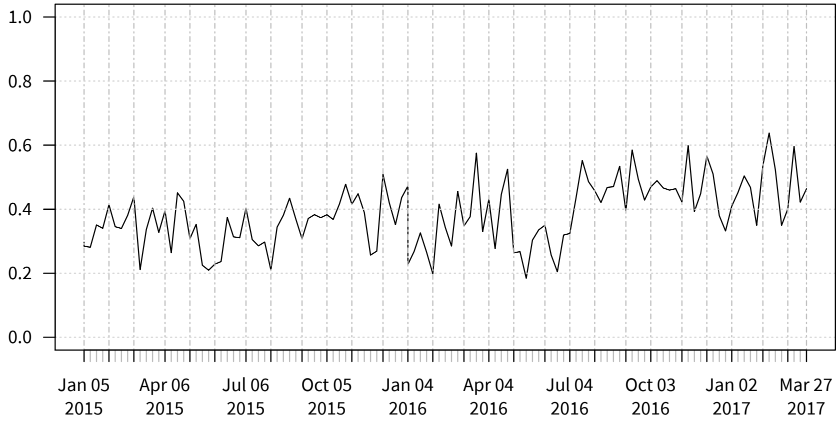



Weekly Segregation Index on Pets
1.0
0.8
0.6
0.4
0.2
0.0
Jan 05 2015
Apr 06 2015
Jul 06 2015
Oct 05 2015
Jan 04 2016
Apr 04 2016
Jul 04 2016
Oct 03 2016
Jan 02 2017
Mar 27 2017

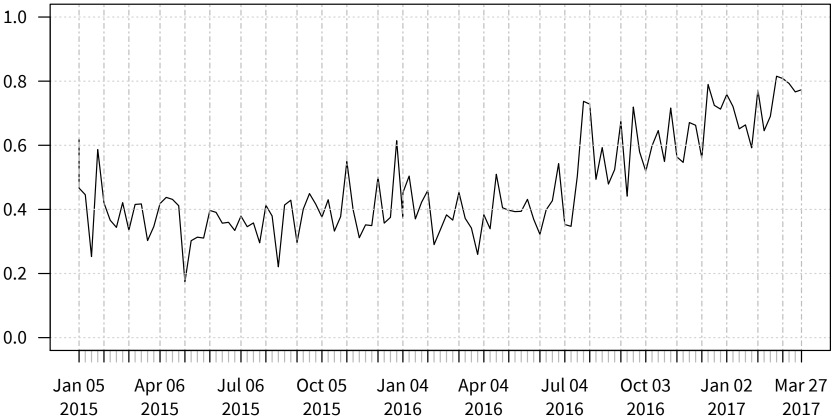
Weekly Segregation Index on Kids



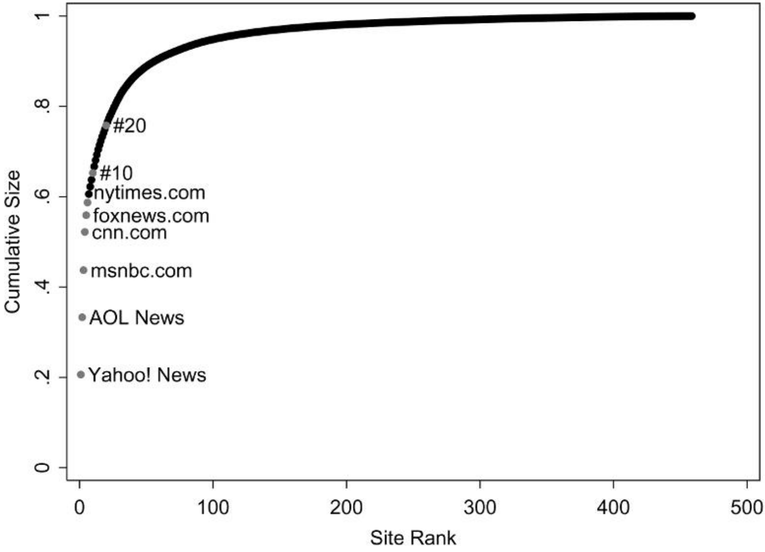
Weekly Segregation Index on Sports

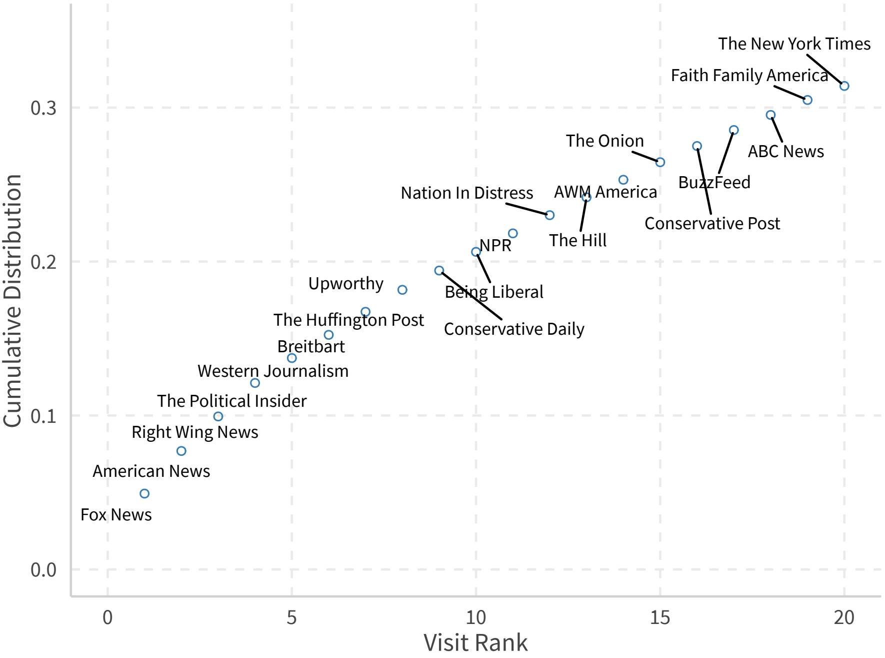


Weekly Segregation Index on Russia

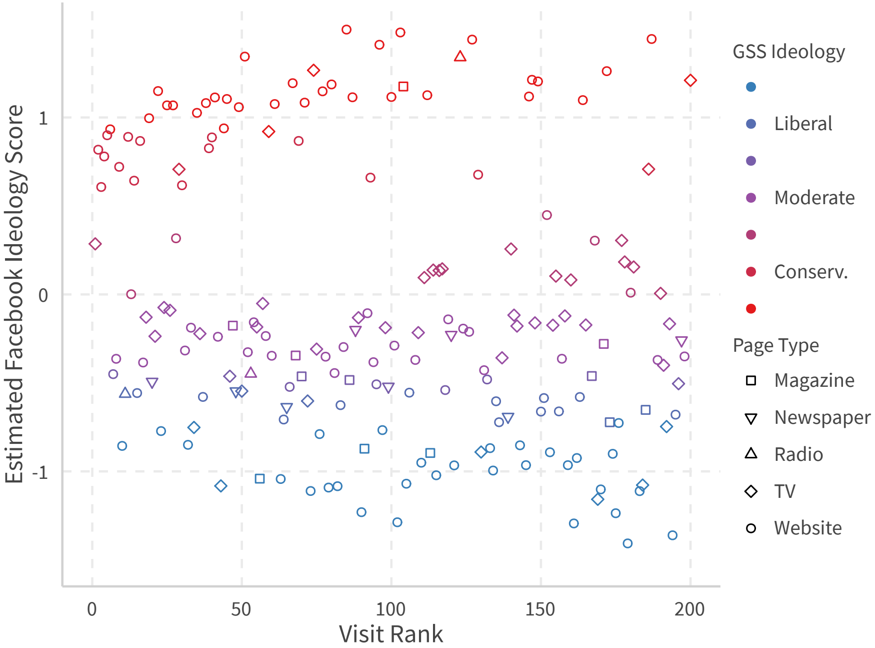


Gentzkow and Shapiro (2015)





The New York Times
Faith Family America
ABC News
BuzzFeed
Conservative Post
The Onion
AWM America
The Hill
Nation In Distress
NPR
Being Liberal
Conservative Daily
Upworthy
The Huffington Post
Breitbart
Western Journalism
The Political Insider
Right Wing News
American News
Fox News
Cumulative Distribution
0.0
0.1
0.2
0.3
0
5
10
15
20
Visit Rank


GSS Ideology
Liberal
Moderate
Conserv.
Page Type
Magazine
Newspaper
Radio
TV
Website
Estimated Facebook Ideology Score
1
0
-1
0
50
100
150
200
Visit Rank

References I

Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23 (1): 76–91. http://dx.doi.org/10.1093/pan/mpu011.

Barberá, Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. 2015. "Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber?" *Psychological Science* 26 (10): 1531–1542. http://dx.doi.org/10.1177/0956797615594620.

Bond, Robert M., and Solomon Messing. 2015. "Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook." *American Political Science Review* 109 (1): 62–78. http://dx.doi.org/10.1017/s0003055414000525.

References II

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58 (2): 367–386. http://dx.doi.org/10.1111/ajps.12062.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98 (2): 355–370. http://dx.doi.org/10.1017/S0003055404001194.

Gentzkow, Matthew, and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126 (4): 1799–1839. http://dx.doi.org/10.1093/qje/qjr044.

Gentzkow, Matthew, and Jesse M. Shapiro. 2015. "Ideology and Online News." In *Economic Analysis of the Digital Economy*.: University of Chicago.

References III

Groseclose, Tim, and Jeffrey Milyo. 2005. “A Measure of Media Bias.” *Quarterly Journal of Economics* 120 (4): 1191–1237. http://dx.doi.org/10.1162/003355305775097542.

Heckman, James J., and James M. Snyder, Jr. 1997. “Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators.” *RAND Journal of Economics* 28 : S142–S189. http://www.jstor.org/stable/3087459.

Poole, Keith T., and Howard L. Rosenthal. 1985. “A Spatial Model for Legislative Roll Call Analysis.” *American Journal of Political Science* 29 (2): 357–384. http://www.jstor.org/stable/2111172.