Using Facebook Data to Predict 2016 US Presidential Election

Keng-Chi Chang   Chun-Fang Chiang   Ming-Jen Lin

Department of Economics
National Taiwan University

2018-05-29

Prepared for Innovations in Political Methodology
and China Study International Conference

In This Paper

- Previous social media ideology measures
  - Are mostly for elites, and uses “following” of a fan page

In This Paper

- Previous social media ideology measures
  - Are mostly for elites, and uses “following” of a fan page
  - But people also consume news and process info. through posts

In This Paper

- Previous social media ideology measures
  - Are mostly for elites, and uses “following” of a fan page
  - But people also consume news and process info. through posts
- We use 19B “likes” on posts of 2K US fan pages to scale ideology
  - Also account for media, interest groups, parties, etc, and users
  - Pages share similar ideology should share “likes” from similar users
  - Adds time, post content, and region (guessed states) dimensions

In This Paper

- Previous social media ideology measures
  - Are mostly for elites, and uses “following” of a fan page
  - But people also consume news and process info. through posts
- We use 19B “likes” on posts of 2K US fan pages to scale ideology
  - Also account for media, interest groups, parties, etc, and users
  - Pages share similar ideology should share “likes” from similar users
  - Adds time, post content, and region (guessed states) dimensions
- We predict 2016 US presidential election using this measure
  - Derive state level FB support rates based on spatial model
  - Compare with actual vote shares and state polls

In This Paper

- Previous social media ideology measures
  - Are mostly for elites, and uses “following” of a fan page
  - But people also consume news and process info. through posts
- We use 19B “likes” on posts of 2K US fan pages to scale ideology
  - Also account for media, interest groups, parties, etc, and users
  - Pages share similar ideology should share “likes” from similar users
  - Adds time, post content, and region (guessed states) dimensions
- We predict 2016 US presidential election using this measure
  - Derive state level FB support rates based on spatial model
  - Compare with actual vote shares and state polls
- We find under minimal assumptions, Facebook support rates:
  - Predicts election quite well and shares similar trends with polls
  - Overestimates winner’s vote share, but may enhance prediction

Facebook Data

- Facebook provides fan page data through Graph API

Facebook Data

- Facebook provides fan page data through Graph API
- Specify fan page ideological universe
  - 1475 fan pages of national politicians

    $\rightsquigarrow$ Members and candidates of Senate, House, and Governors

Facebook Data

- Facebook provides fan page data through Graph API
- Specify fan page ideological universe
  - 1475 fan pages of national politicians

    $\rightsquigarrow$ Members and candidates of Senate, House, and Governors
  - Top 1000 pages related to 2016 presidential election

    $\rightsquigarrow$ In Aug 2016, find all pages mentioned "Trump" and "Clinton"

    $\rightsquigarrow$ Weight by likes, comments, shares, find top 1000 pages

Facebook Data

- Facebook provides fan page data through Graph API
- Specify fan page ideological universe
  - 1475 fan pages of national politicians
    - $\rightsquigarrow$ Members and candidates of Senate, House, and Governors
  - Top 1000 pages related to 2016 presidential election
    - $\rightsquigarrow$ In Aug 2016, find all pages mentioned “Trump” and “Clinton”
    - $\rightsquigarrow$ Weight by likes, comments, shares, find top 1000 pages
    - $\rightsquigarrow$ Includes all major news outlets, interest groups, parties, etc
    - $\rightsquigarrow$ NYT, Fox News, NRA, RNC, Occupy Wall St, Tea Party, 9GAG, ...

Facebook Data

- Facebook provides fan page data through Graph API
- Specify fan page ideological universe
  - 1475 fan pages of national politicians

    $\rightsquigarrow$ Members and candidates of Senate, House, and Governors
  - Top 1000 pages related to 2016 presidential election

    $\rightsquigarrow$ In Aug 2016, find all pages mentioned "Trump" and "Clinton"

    $\rightsquigarrow$ Weight by likes, comments, shares, find top 1000 pages

    $\rightsquigarrow$ Includes all major news outlets, interest groups, parties, etc

    $\rightsquigarrow$ NYT, Fox News, NRA, RNC, Occupy Wall St, Tea Party, 9GAG, ...
- Collect all 24M posts in 2015 and 2016 on these pages

Facebook Data

- Facebook provides fan page data through Graph API
- Specify fan page ideological universe
  - 1475 fan pages of national politicians
    - ⇝ Members and candidates of Senate, House, and Governors
  - Top 1000 pages related to 2016 presidential election
    - ⇝ In Aug 2016, find all pages mentioned “Trump” and “Clinton”
    - ⇝ Weight by likes, comments, shares, find top 1000 pages
    - ⇝ Includes all major news outlets, interest groups, parties, etc
    - ⇝ NYT, Fox News, NRA, RNC, Occupy Wall St, Tea Party, 9GAG, ...
- Collect all 24M posts in 2015 and 2016 on these pages
- And user’s 19B reactions (mostly likes) to these posts

Data Summary

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

Estimation: Shared Users Matrix

- Measure ideology of pages, then measure those of users
  $\rightsquigarrow$ Similar to Bond and Messing (2015, APSR)

Estimation: Shared Users Matrix

- Measure ideology of pages, then measure those of users
  $\rightsquigarrow$ Similar to Bond and Messing (2015, APSR)
- First build the page by page affiliation matrix A
  $\rightsquigarrow$ Number of shared users (based on likes) 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: Transform to Ratios

- Transform A to matrix of ratios G, where $g_{ij} = a_{ij}/a_{ii}$

$$\rightsquigarrow 0.44 = \frac{\Pr(\text{Trump} \cap \text{FoxNews})}{\Pr(\text{FoxNews})} = \Pr(\text{Trump} \mid \text{FoxNews})$$

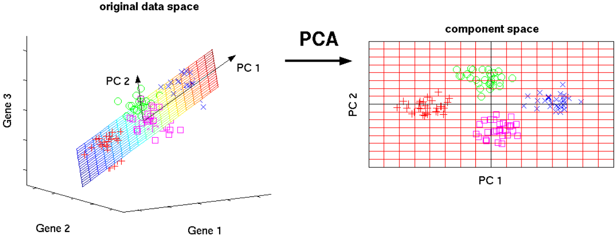
- Can interpret columns as features and rows as observations

  $\rightsquigarrow$ 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

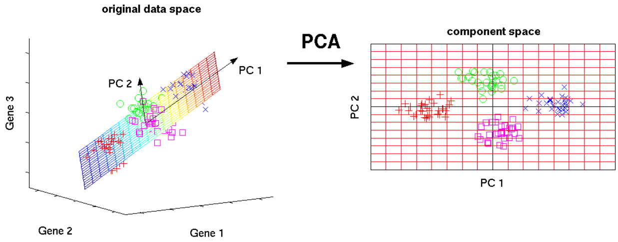
Estimation: Dimension Reduction

- Compute the principal components of G after standardizing



Estimation: Dimension Reduction

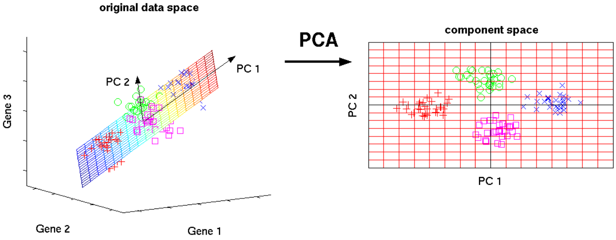
- Compute the principal components of G after standardizing



- PC1 is the dimension explains the largest variation
  $\rightsquigarrow$ Unsupervised $\Rightarrow$ Guess and verify PC1 is related to “ideology”

Estimation: Dimension Reduction

- Compute the principal components of G after standardizing

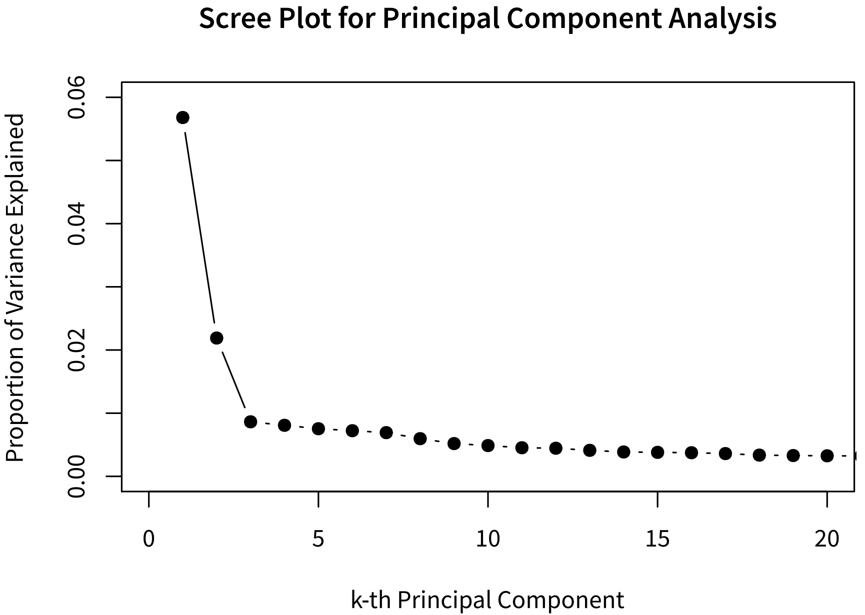


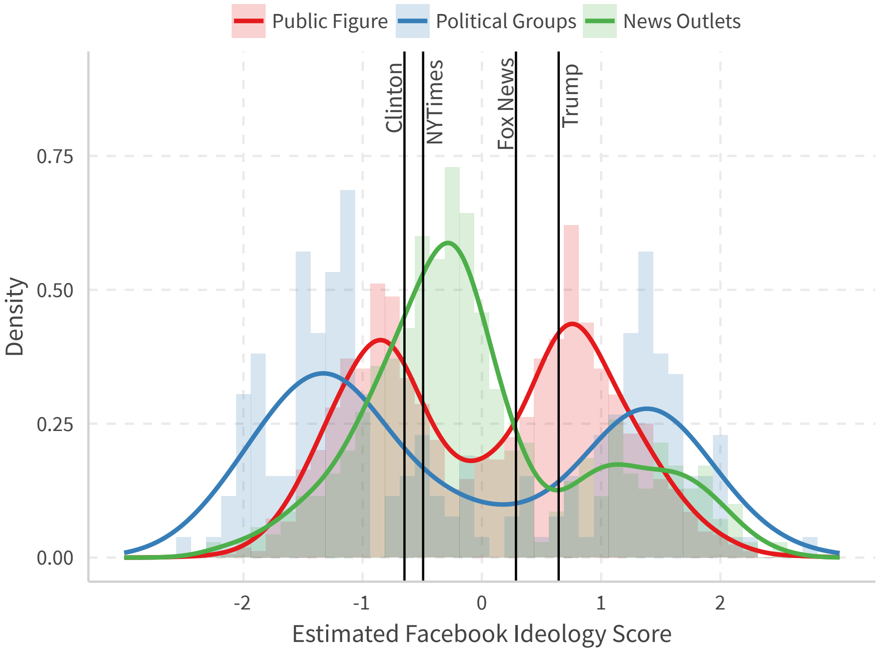
- PC1 is the dimension explains the largest variation

  $\rightsquigarrow$ Unsupervised $\Rightarrow$ Guess and verify PC1 is related to "ideology"

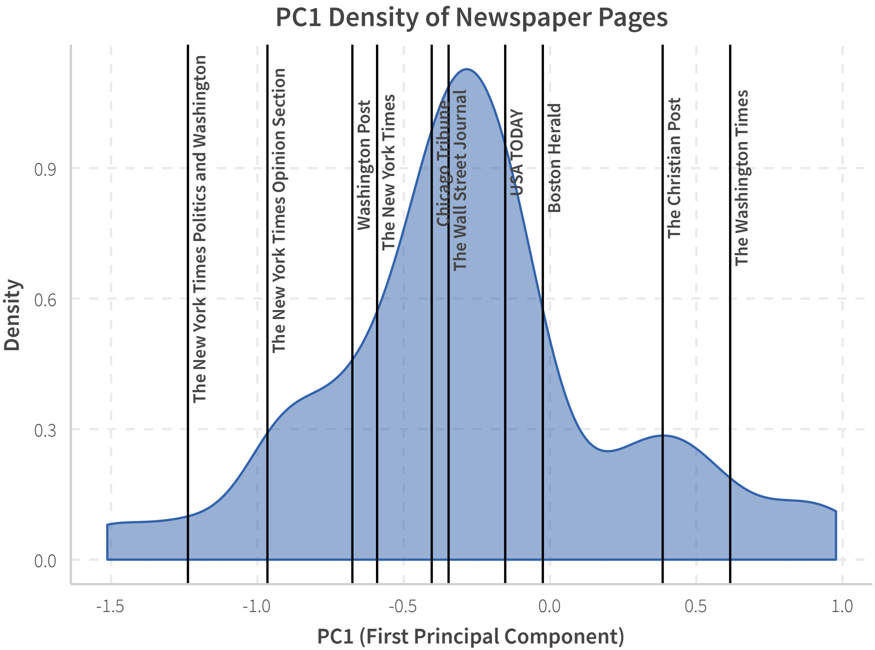
- User ideology = mean ideology of pages user liked
- Guess user's state residence by their likes on national politicians

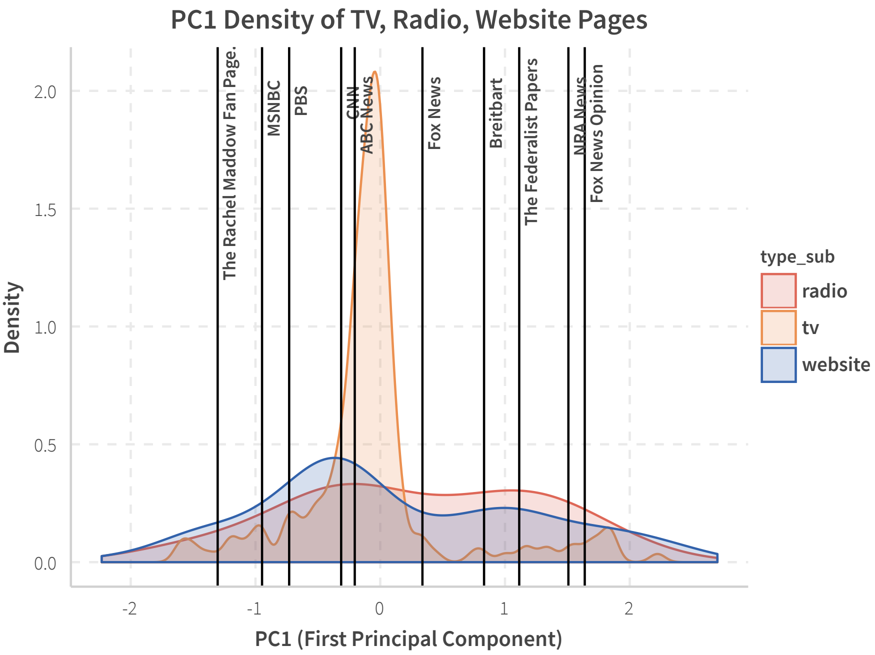
  $\rightsquigarrow$ Like more politicians from NY $\Rightarrow$ More likely from NY


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

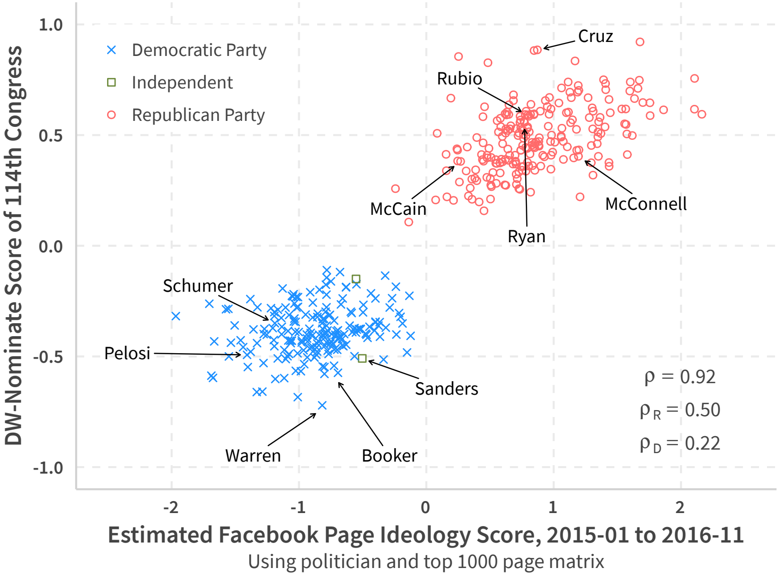


Public Figure
Political Groups
News Outlets
Clinton
NYTimes
Fox News
Trump
Density
0.75
0.50
0.25
0.00
-2
-1
0
1
2
Estimated Facebook Ideology Score

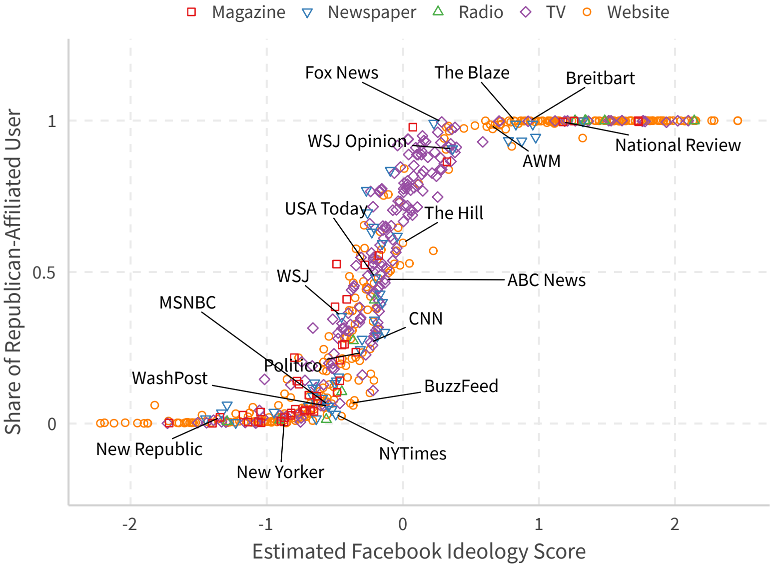

PC1 Density of Newspaper Pages
Density
PC1 (First Principal Component)
0.0
0.3
0.6
0.9
-1.5
-1.0
-0.5
0.0
0.5
1.0
The New York Times Politics and Washington
The New York Times Opinion Section
Washington Post
The New York Times
Chicago Tribune
The Wall Street Journal
USA TODAY
Boston Herald
The Christian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


PC1 Density of 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

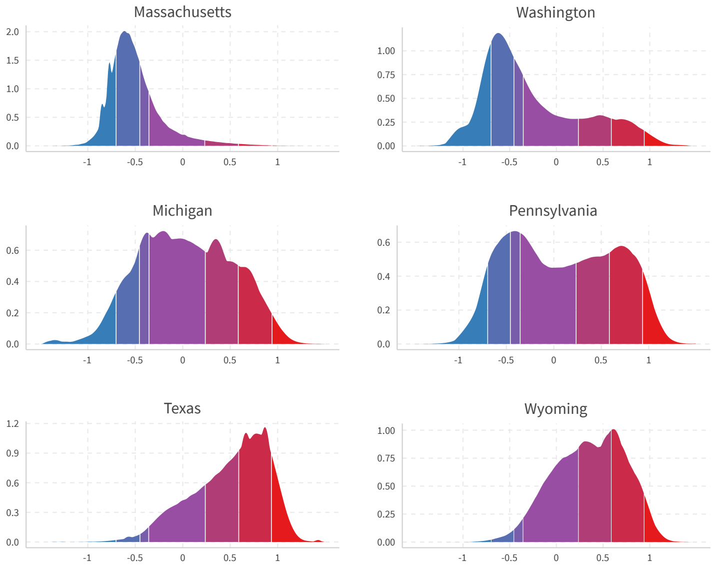
Validation for Congressional Politicians



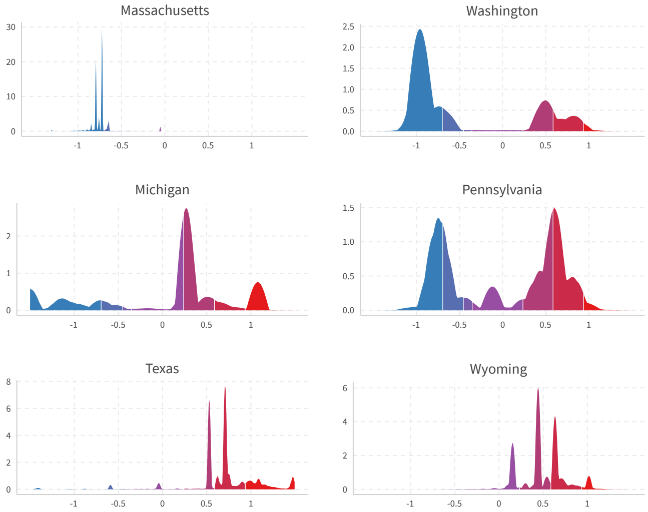
Validation for Media



User Ideology Density by States

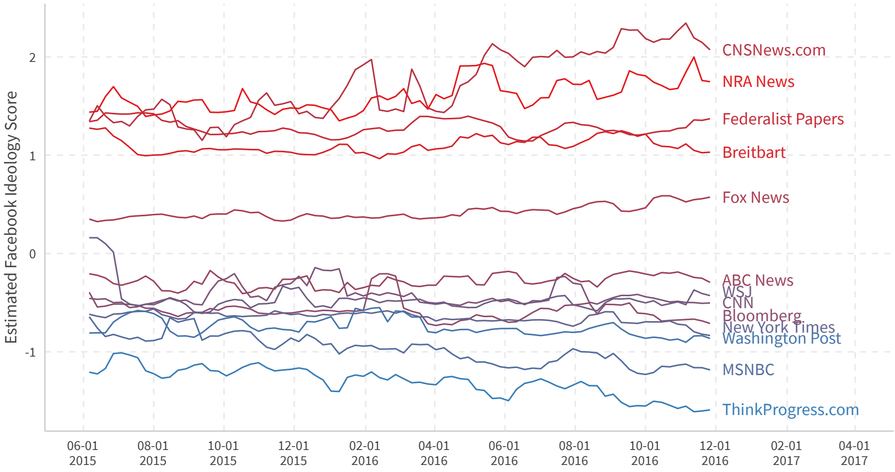


User Ideology Density by States

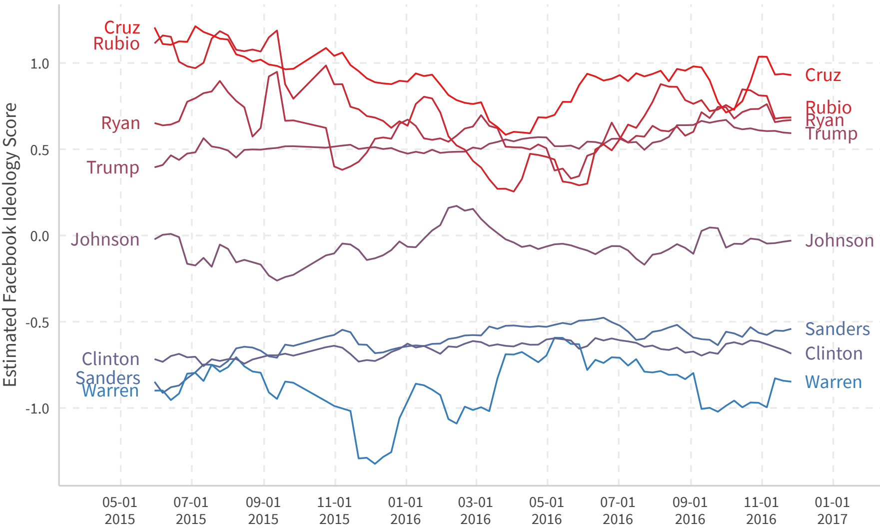


Politician-Only Method (Bond and Messing 2015)

Media Ideology Dynamics



Politician Ideology Dynamics



FB Support Rates, Polls, and Vote Shares

- Apply the Hotelling-Downs spatial model for voting: Voters support candidates closer to their own ideological location

FB Support Rates, Polls, and Vote Shares

- Apply the Hotelling-Downs spatial model for voting: Voters support candidates closer to their own ideological location
- In each state, we compare:
    - FB support rate: Share of user's ideology closer to Trump or Clinton

FB Support Rates, Polls, and Vote Shares

- Apply the Hotelling-Downs spatial model for voting: Voters support candidates closer to their own ideological location
- In each state, we compare:
  - FB support rate: Share of user's ideology closer to Trump or Clinton
    - $\rightsquigarrow$ May not be a precise estimator for vote shares
    - $\rightsquigarrow$ Since turnouts may not be the same across states
    - $\rightsquigarrow$ But adding assumptions may look like fitting the data

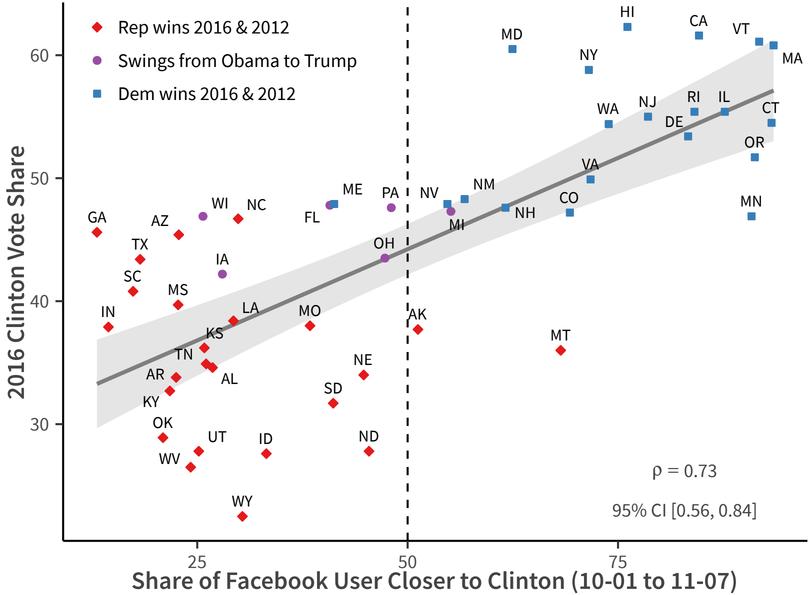
FB Support Rates, Polls, and Vote Shares

- Apply the Hotelling-Downs spatial model for voting: Voters support candidates closer to their own ideological location
- In each state, we compare:
  - FB support rate: Share of user's ideology closer to Trump or Clinton
    - $\rightsquigarrow$ May not be a precise estimator for vote shares
    - $\rightsquigarrow$ Since turnouts may not be the same across states
    - $\rightsquigarrow$ But adding assumptions may look like fitting the data
  - Polls: State polling averages calculated by FiveThirtyEight

FB Support Rates, Polls, and Vote Shares

- Apply the Hotelling-Downs spatial model for voting: Voters support candidates closer to their own ideological location
- In each state, we compare:
  - FB support rate: Share of user's ideology closer to Trump or Clinton
    - $\rightsquigarrow$ May not be a precise estimator for vote shares
    - $\rightsquigarrow$ Since turnouts may not be the same across states
    - $\rightsquigarrow$ But adding assumptions may look like fitting the data
  - Polls: State polling averages calculated by FiveThirtyEight
  - Actual vote shares in 2016 election

Predicting Vote Shares and Outcomes



Compare with Major Forecasters

Battleground States	E.V.[†]	Winner	FB	538	NYT	PEC[*]

Compare with Major Forecasters

Battleground States	E.V.[†]	Winner	FB	538	NYT	PEC[*]
Florida	29	Trump	○	×	×	×
Pennsylvania	20	Trump	○	×	×	×
Wisconsin	10	Trump	○	×	×	×

Compare with Major Forecasters

Battleground States	E.V.†	Winner	FB	538	NYT	PEC*
Florida	29	Trump	$\circ$	$\times$	$\times$	$\times$
Pennsylvania	20	Trump	$\circ$	$\times$	$\times$	$\times$
Wisconsin	10	Trump	$\circ$	$\times$	$\times$	$\times$
Michigan	16	Trump	$\times$	$\times$	$\times$	$\times$

Compare with Major Forecasters

Battleground States	E.V.[†]	Winner	FB	538	NYT	PEC[*]
Florida	29	Trump	$\circ$	$\times$	$\times$	$\times$
Pennsylvania	20	Trump	$\circ$	$\times$	$\times$	$\times$
Wisconsin	10	Trump	$\circ$	$\times$	$\times$	$\times$
Michigan	16	Trump	$\times$	$\times$	$\times$	$\times$
Ohio	18	Trump	$\circ$	$\circ$	$\circ$	$\circ$
Iowa	6	Trump	$\circ$	$\circ$	$\circ$	$\circ$

Compare with Major Forecasters

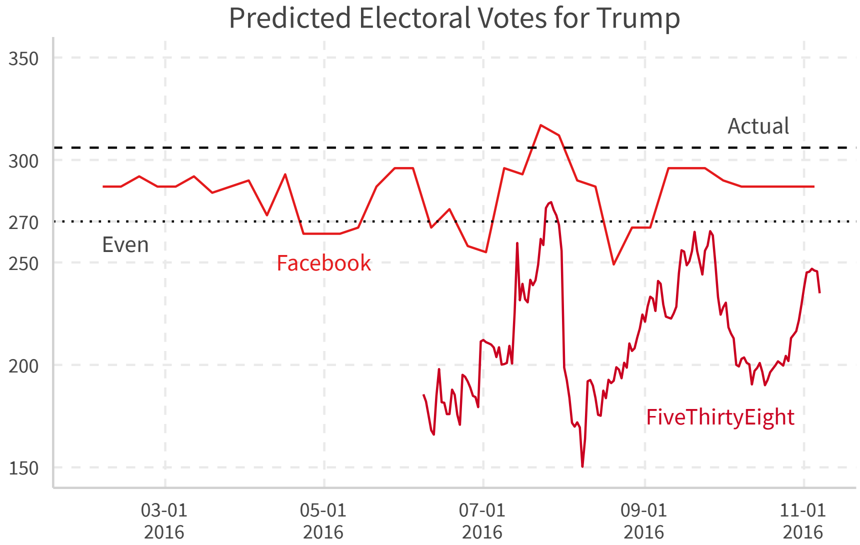
Battleground States	E.V.†	Winner	FB	538	NYT	PEC*
Florida	29	Trump	○	×	×	×
Pennsylvania	20	Trump	○	×	×	×
Wisconsin	10	Trump	○	×	×	×
Michigan	16	Trump	×	×	×	×
Ohio	18	Trump	○	○	○	○
Iowa	6	Trump	○	○	○	○
Montana	3	Trump	×	○	○	○
Alaska	3	Clinton	×	○	○	○
Maine	2	Clinton	×	○	○	○

Compare with Major Forecasters

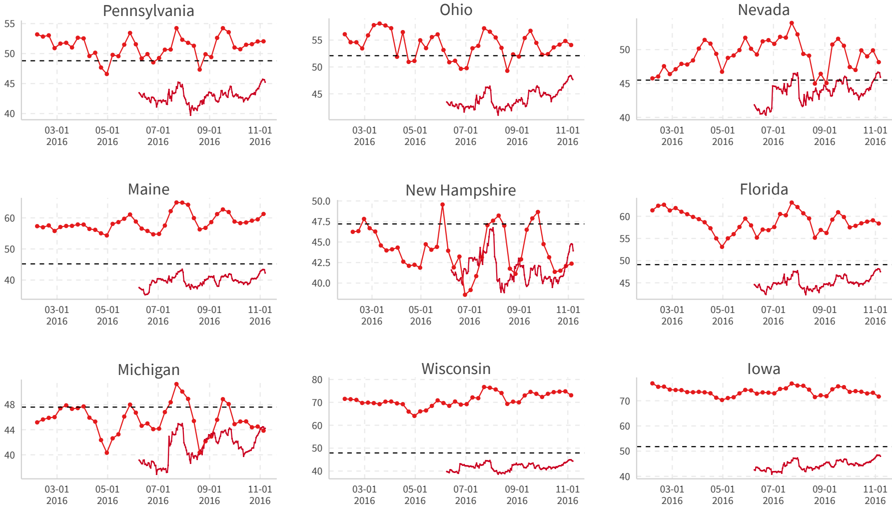
Battleground States	E.V.†	Winner	FB	538	NYT	PEC*
Florida	29	Trump	○	×	×	×
Pennsylvania	20	Trump	○	×	×	×
Wisconsin	10	Trump	○	×	×	×
Michigan	16	Trump	×	×	×	×
Ohio	18	Trump	○	○	○	○
Iowa	6	Trump	○	○	○	○
Montana	3	Trump	×	○	○	○
Alaska	3	Clinton	×	○	○	○
Maine	2	Clinton	×	○	○	○
Trump's Electoral Vote		306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

† Electoral Votes. * Princeton Election Consortium.

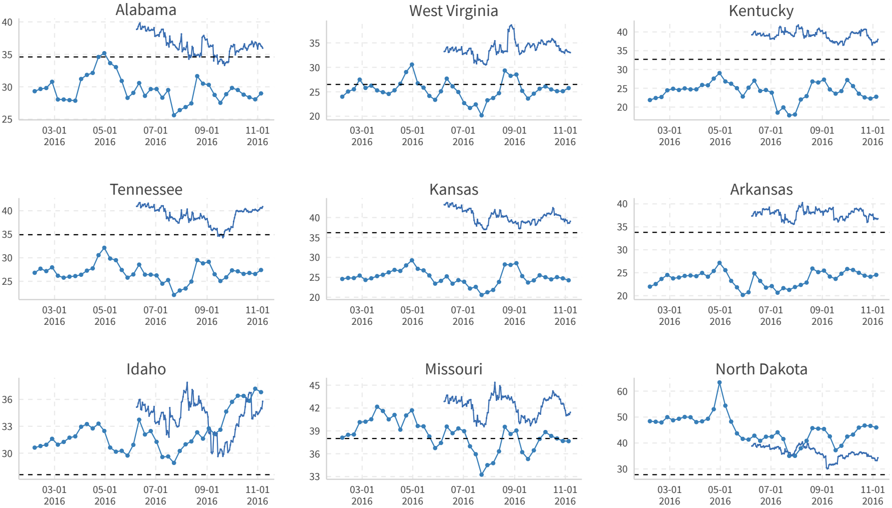
Predicting Electoral Votes



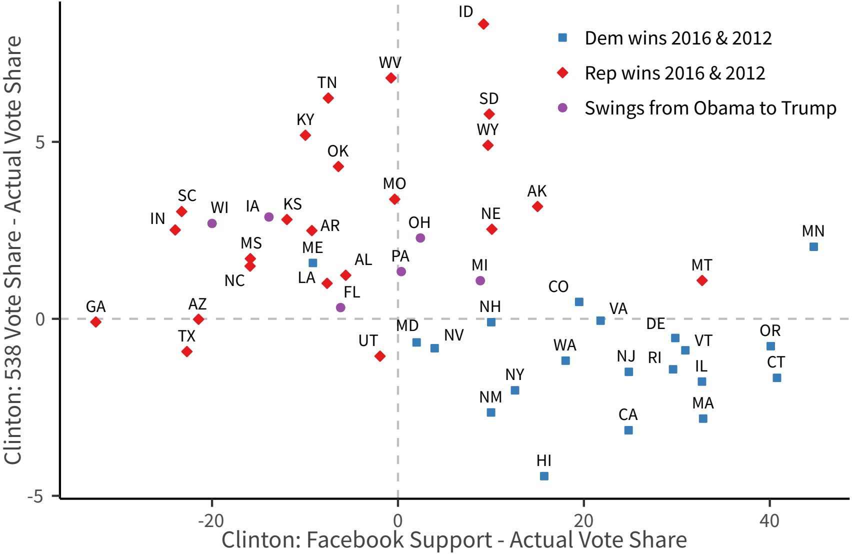
Trump: FB (Dotted), Polls, and Vote Shares



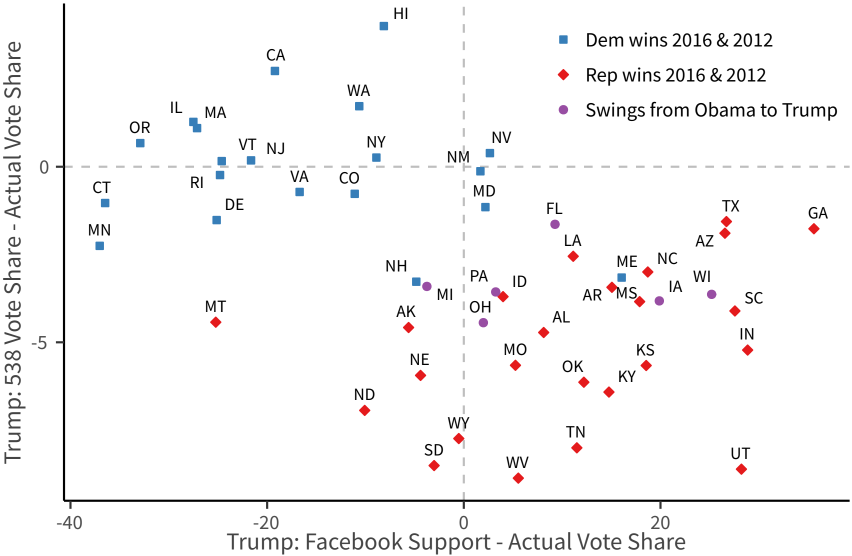
Clinton: FB (Dotted), Polls, and Vote Shares



Polls Overestimates Clinton in Red and Swing States



FB Overestimates Trump in Red and Swing States



Discussions

- Strengths of Facebook based prediction:
  - Revealed preference instead of self-report
  - Low cost and almost in real time
  - Trace individuals repeatedly over time
  - Overestimation for winners can help to make predictions

Discussions

- Strengths of Facebook based prediction:
  - Revealed preference instead of self-report
  - Low cost and almost in real time
  - Trace individuals repeatedly over time
  - Overestimation for winners can help to make predictions
- Weaknesses, compared to polls or surveys:
  - Not representative

    $\rightsquigarrow$ Can reweight if more social-demographic information is known
  - Hard to link with offline behaviors

    $\rightsquigarrow$ Ex. “Strong supporter” vs. “Likely voter”

Discussions

- Strengths of Facebook based prediction:
  - Revealed preference instead of self-report
  - Low cost and almost in real time
  - Trace individuals repeatedly over time
  - Overestimation for winners can help to make predictions
- Weaknesses, compared to polls or surveys:
  - Not representative

    $\rightsquigarrow$ Can reweight if more social-demographic information is known
  - Hard to link with offline behaviors

    $\rightsquigarrow$ Ex. "Strong supporter" vs. "Likely voter"
- Can complement each other if more research try to link the two

Working on: Effect of Fake News

- Joint with Chun-Fang Chiang, Brian Knight, and Ming-Jen Lin
- Would consuming fake news change people's ideology or information consumption?

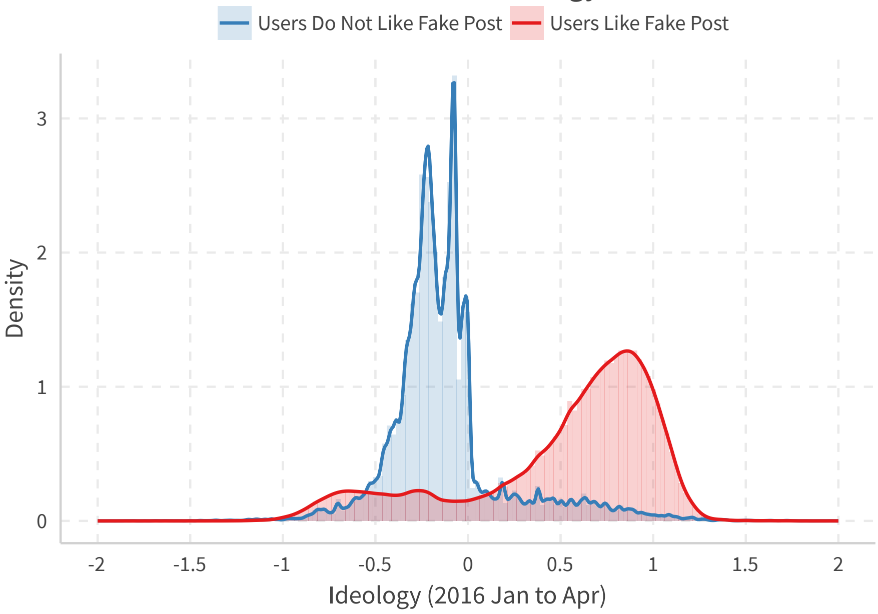
Working on: Effect of Fake News

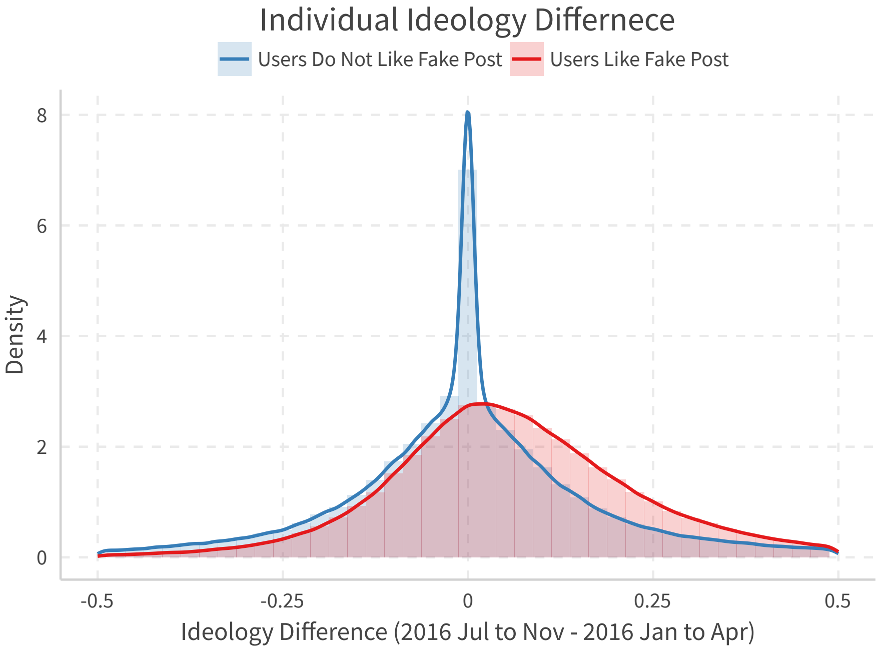
- Joint with Chun-Fang Chiang, Brian Knight, and Ming-Jen Lin
- Would consuming fake news change people's ideology or information consumption?
- If so, what kind of fake stories have larger effect, and why?

Working on: Effect of Fake News

- Joint with Chun-Fang Chiang, Brian Knight, and Ming-Jen Lin
- Would consuming fake news change people's ideology or information consumption?
- If so, what kind of fake stories have larger effect, and why?
- Fake news pool on Facebook:
  - Top 40 fake stories, 536 posts, 130 pages
  - Posts link to fake domains, 139,074 posts, 177 pages

Individual Ideology




Individual Ideology Differnece
Users Do Not Like Fake Post
Users Like Fake Post
Density
8
6
4
2
0
-0.5
-0.25
0
0.25
0.5
Ideology Difference (2016 Jul to Nov - 2016 Jan to Apr)

Strategies for Identification

- Challenges:
  - People "like" fake post may be very different
  - Pages posting fake posts may attract very different users
  - Some stories may be "too fake" for people to believe, even backfire

Strategies for Identification

- Challenges:
  - People “like” fake post may be very different
  - Pages posting fake posts may attract very different users
  - Some stories may be “too fake” for people to believe, even backfire
- For each fake post, we:
  - Find nonfake pages very similar to fake page through different matching methods as control

Strategies for Identification

- Challenges:
  - People “like” fake post may be very different
  - Pages posting fake posts may attract very different users
  - Some stories may be “too fake” for people to believe, even backfire
- For each fake post, we:
  - Find nonfake pages very similar to fake page through different matching methods as control
  - Find potential *followers* of these pages, instead of “likes”

Strategies for Identification

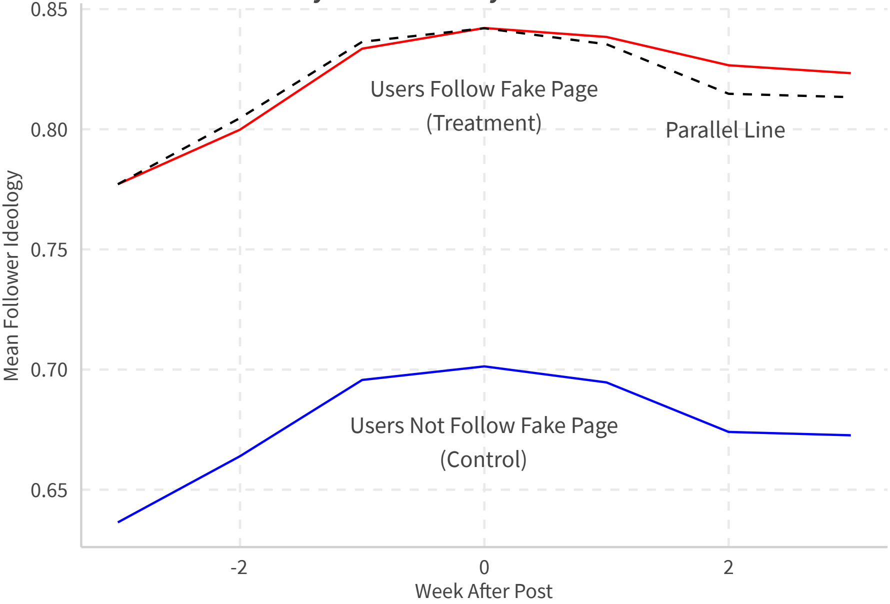
- Challenges:
  - People "like" fake post may be very different
  - Pages posting fake posts may attract very different users
  - Some stories may be "too fake" for people to believe, even backfire
- For each fake post, we:
  - Find nonfake pages very similar to fake page through different matching methods as control
  - Find potential *followers* of these pages, instead of "likes"
  - Compare the ideology of these fake and nonfake followers before and after fake page *unexpectedly started posting fake story*

Strategies for Identification

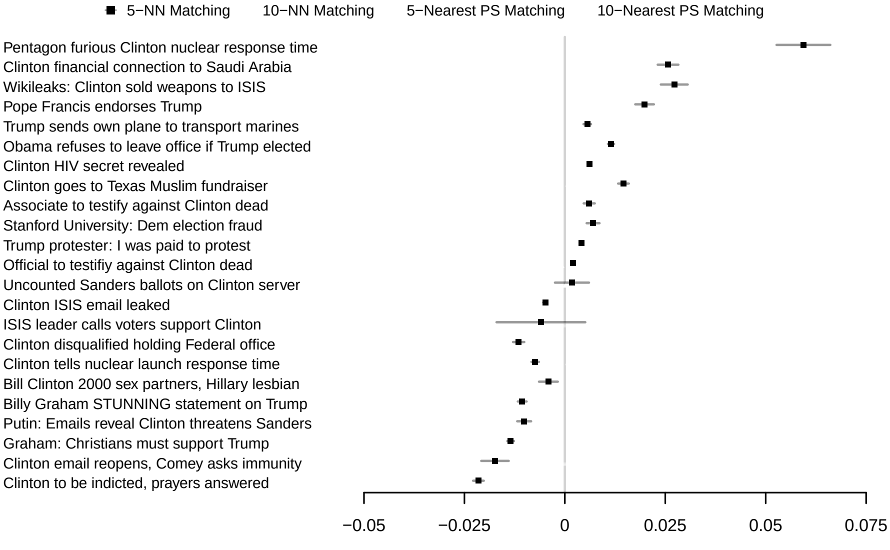
- Challenges:
  - People "like" fake post may be very different
  - Pages posting fake posts may attract very different users
  - Some stories may be "too fake" for people to believe, even backfire
- For each fake post, we:
  - Find nonfake pages very similar to fake page through different matching methods as control
  - Find potential *followers* of these pages, instead of "likes"
  - Compare the ideology of these fake and nonfake followers before and after fake page *unexpectedly started posting fake story*

$$\begin{aligned}\text{Ideology}_{it} = \alpha\,\mathbb{1}(\text{After}_t) + \gamma\,\mathbb{1}(\text{FollowFake}_i) \\ + \beta\,\mathbb{1}(\text{FollowFake}_i)\mathbb{1}(\text{After}_t) + \varepsilon_{it}\end{aligned}$$

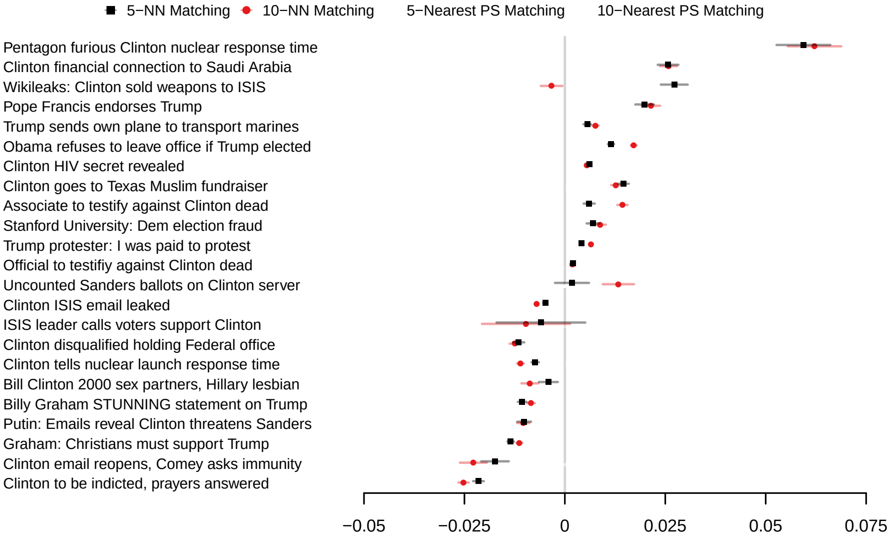
"BREAKING: Official Set to Testify Against Hillary Found Dead" by Western Journalism



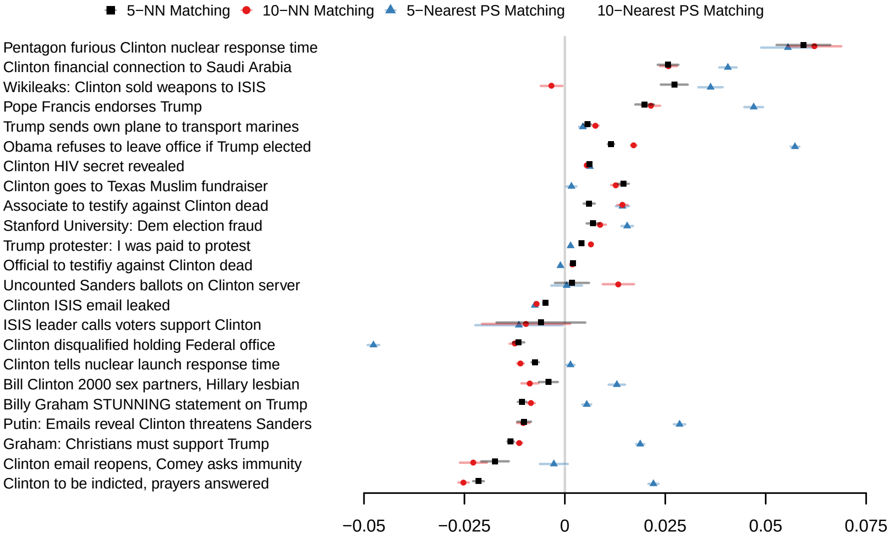
Story Level Ideology Change for Following Pages Sharing Pro−Trump Fake News Week +1 to −1, DiD Estimates with Individual Fixed Effects and 99.9% CI



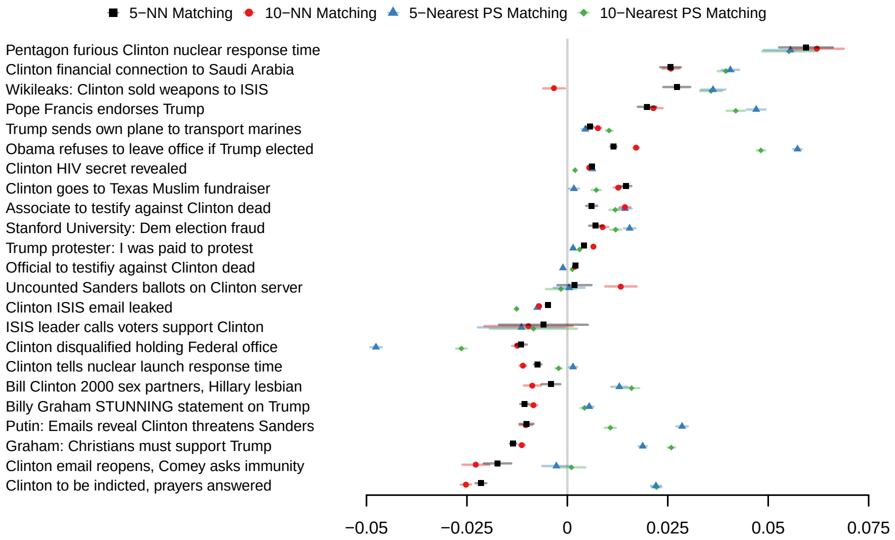
Story Level Ideology Change for Following Pages Sharing Pro−Trump Fake News Week +1 to −1, DiD Estimates with Individual Fixed Effects and 99.9% CI



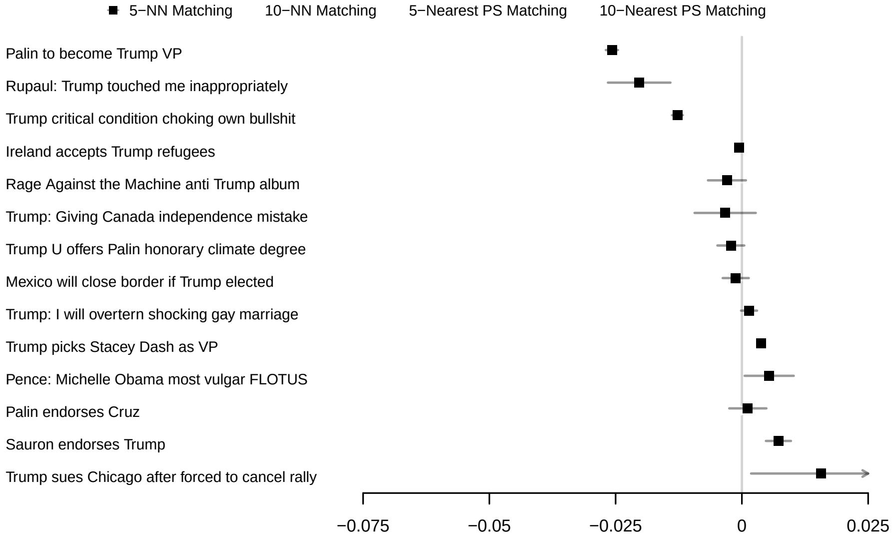
Story Level Ideology Change for Following Pages Sharing Pro−Trump Fake News Week +1 to −1, DiD Estimates with Individual Fixed Effects and 99.9% CI



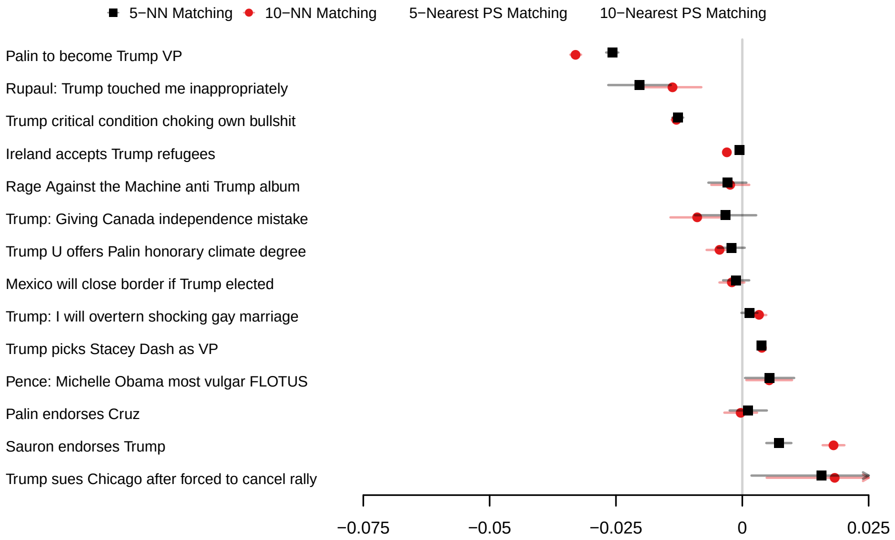
Story Level Ideology Change for Following Pages Sharing Pro−Trump Fake News
Week +1 to −1, DiD Estimates with Individual Fixed Effects and 99.9% CI



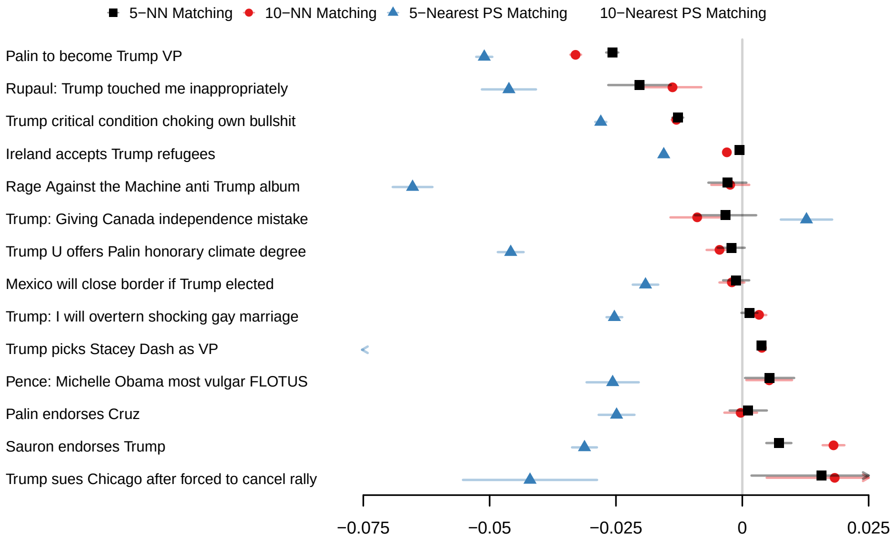
Story Level Ideology Change for Following Pages Sharing Pro−Clinton Fake News
Week +1 to −1, DiD Estimates with Individual Fixed Effects and 99.9% CI



Story Level Ideology Change for Following Pages Sharing Pro−Clinton Fake News Week +1 to −1, DiD Estimates with Individual Fixed Effects and 99.9% CI



Story Level Ideology Change for Following Pages Sharing Pro−Clinton Fake News
Week +1 to −1, DiD Estimates with Individual Fixed Effects and 99.9% CI



Story Level Ideology Change for Following Pages Sharing Pro−Clinton Fake News Week +1 to −1, DiD Estimates with Individual Fixed Effects and 99.9% CI