Beyond Online Echo Chambers: Measuring Levels of Online and Offline Partisan Segregation on Twitter

Megan Brown[1,3], Tiago Ventura [2,3] , Jonathan Nagler[3], and Joshua A. Tucker[3]

[1] University of Michigan, [2] Georgetown University, [3] Center for Social Media and Politics, NYU

CSMaP Meeting
09/18/2023

Motivation

OPINION > CAMPAIGN

THE VIEWS EXPRESSED BY CONTRIBUTORS ARE THEIR OWN AND NOT THE VIEW OF THE HILL

How social media fuels U.S. political polarization — what to do about it

BY PAUL BARRETT, JUSTIN HENDRIX AND GRANT SIMS, OPINION CONTRIBUTORS - 09/13/21 4:00 PM ET

SHARE TWEET

BOOKS

LISTEN & FOLLOW

How the polarizing effect of social media is speeding up

September 9, 2022 · 5:01 AM ET

By Ari Shapiro, Michael Levitt, Christopher Intagliata

KEYWORDS: CHRISTOPHER MIMS

Why Social Media Is So Good at Polarizing Us

Mathematicians are teaming up with political scientists to create models of how social media divides us, and results suggest at least one popular solution might actually make the problem worse

Social media

↓

Echo Chambers

↓

Reduce cross-cutting exposure

↓

Polarization

Social media

⊕ ?

Echo Chambers ?

⊕ ?

Reduce cross-cutting exposure ?

⊕ ?

Polarization

Are social media echo-chambers real?

- Online Media Consumption is similar to offline consumption (Gentzkow and Shapiro, 2011; Wojcieszak and Mutz, 2009; Bisbee and Larsson, 2017)
- Users' friendship networks are heterogeneous outside of politics (Bakshy et. al. 2012; Barbéra e.t al., 015)
- Users' digital media diets are balanced, and strongly influenced by big reputable outlets (Guess 2021; Cardenal et. al., 2019)
- Substantial amount of overlap in the ideological distributions of accounts followed by Twitter users (Eady et. al. 2019)

These studies:

Social media

Echo Chambers

Reduce cross-cutting exposure

Polarization

Our focus

Social media

Echo Chambers

Reduce cross-cutting exposure

Polarization

Our contributtion

➜ **To solve the causal chain**, we need to measure online segregation relative to other channels through which voters **consume information or interact** with ingroup and outgroup voters

➜ **Previous related studies:**

- Self-reported online vs offline networks (Gentzkow and Shapiro 2011)
- TV news Consumption (Muise et. al. 2022)
- **No studies looking at partisan geographical segregation and online behavior.**

➜ We provide meaningful comparisons for the same social media users between levels of **offline and online segregation**

Why it matters?

- American voters are highly sorted offline with respect to partisanship (Brown and Enos, 2021)
- Partisan geographical segregation affects mass and elite polarization (Bonica, 2014), willigness to cooperated across groups (Enos and Gidron, 2016), trust in government and anti-system attitudes (Cramer, 2016), and health behavior (Baxter-King et al., 2022).
- **No scholarly work focusing on the effects of partisan geographical segregation and online behavior.**

Research Question 1 (Today)

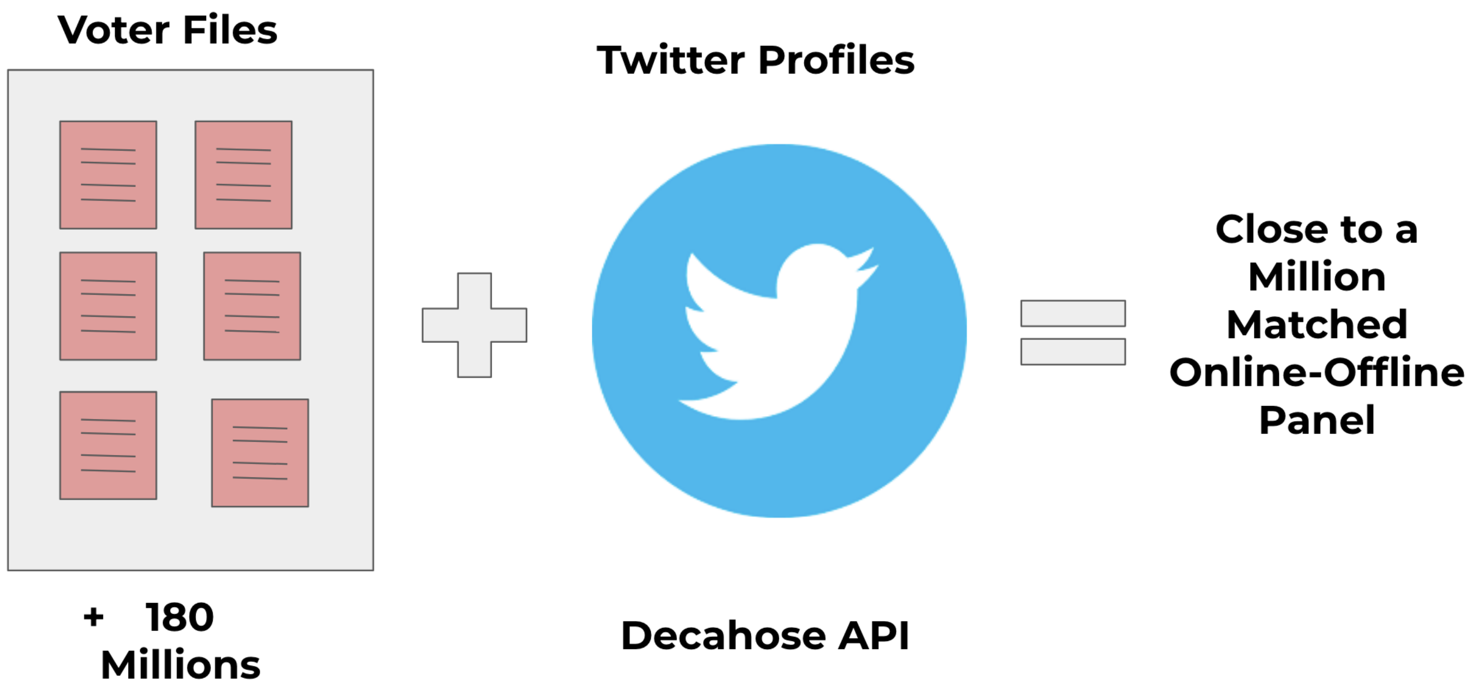
1) What is the relationship between offline partisan sorting and online partisan sorting?

Research Question 2 (TBD)

2) How does offline political segregation influence online behavior?

Materials and Methods

Data Infrastructure



Simple Matching Procedure

Step 1: Parse Voter File Data

- For every voter in each US City, we collect unique:
  - First name
  - Last name

Step 2: Find Candidates on Twitter

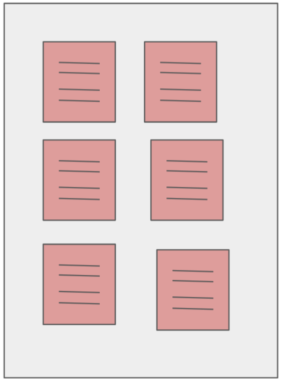
- For every month on API Decahose data:
  - Find matches with the three parsed data
  - Keep all matches

Step 3: Discard Repeated Matches

- Multiple Jonh Does living in New York

Step 4: Keep unique matches

Offline Information: Voter Files



Data Collection for every matched voters:

- Voter file demographics (gender, race, partisanship, religion)
- Residential location (9 digits lat and long)
- Closest 1.000 neighboors + their partisanship.

Online Information: Twitter Data

Data Collection for every matched voters:

- Collect their full network (people they follow and follow them) ~ 57M
- Collect their most recent timelines (3200 tweets) + 900k * 3,2k
- Parse their timelines.

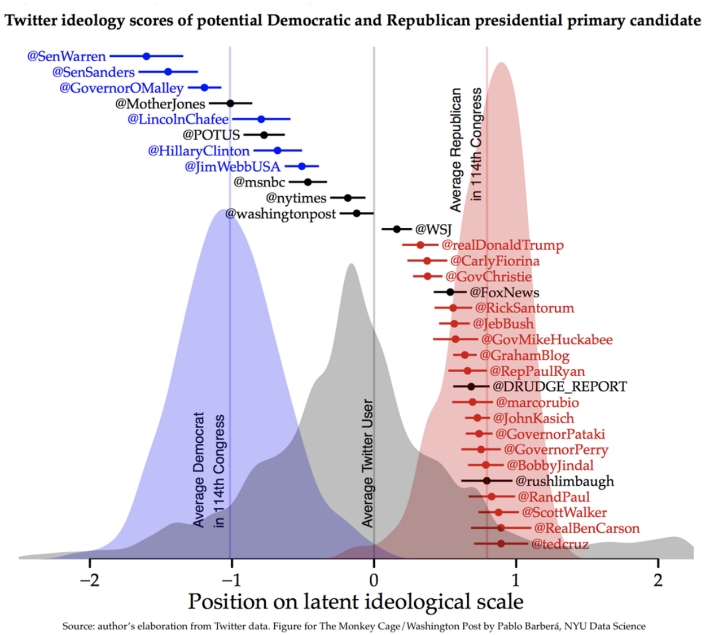
Measuring Partisan Identity

From voter file:

- Precise measure for every matched users

What about their friends:

- Ideology estimation method employed by Barbera, 2015
- Homophily assumption: Following relationships between users and political elites to estimate ideology.



Measuring Partisan Segregation: Offline

$$\text{Offline Outgroup Proximity}$$

$$\frac{\sum_{k=1}^{1000} \frac{1}{d+1}(p_k = q_i)}{\sum_{k=1}^{1000} \frac{1}{d+1}}$$

Where:

- i is a matched voters
- k is a given neighbor
- d is the distance in meters between the neighbor and the individual
- p_k is the partisanship of the neighbor
- q_i is the opposite party of the individual whose exposure is being measured.

Online Partisan Segregation

$$\text{Online Outgroup Proximity}$$

$$\frac{\sum_{k=1}^{n} \log(a+1)(p_k = q_i)}{\sum_{k=1}^{n} \log(a+1)}$$

Where:

- i is a matched voter
- k is a given neighbor
- a is the number of interactions between the friend and a user i
- p_k is the partisanship of the neighbor
- q_i is the opposite party of the individual whose exposure is being measured.

Measurement Assumptions

- We are interested in the probability of the voter **i** being exposed to an outgroup **j** online and offline. We assume this probability increases with the number of outgroup neighbors and users on user **i** networks.
- Two way to measure:
    - **Naive:** count the number of nearest neighbors and users I follow.
    - **Weighted:** use some transformation function.
- We assume proximity matters:
    - **Offline:** geographical (inverse distance)
    - **Online:** online interactions (retweets)

As a consequence:

- These measures are not on the same scale. We cannot say .52>.50
- We argue they are however roughly proportional.
- And allow us to understand differences in orders of magnitude and their common correlation.

Results

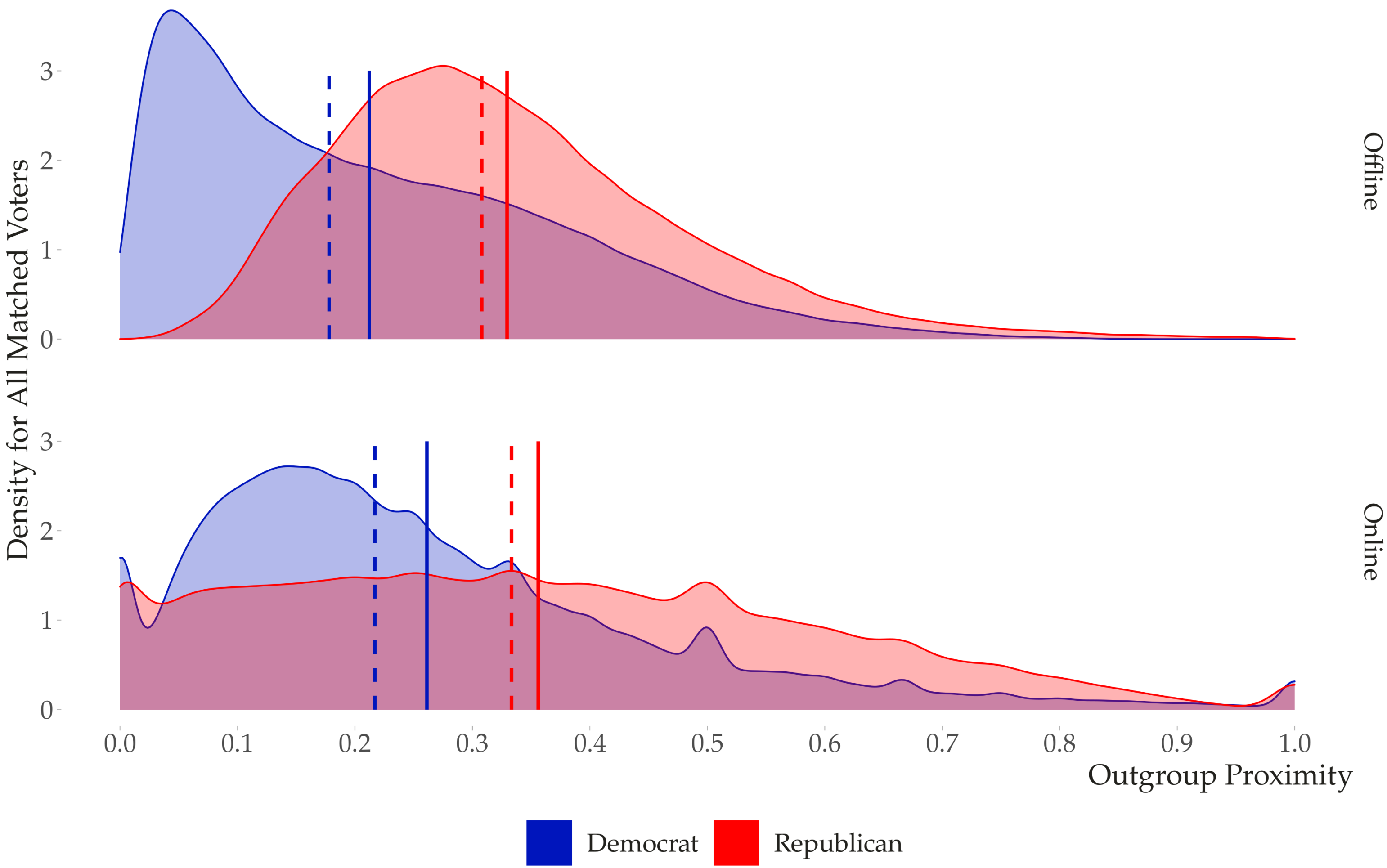
Demographics of Twitter Panel

Table 1 Comparing Demographics:Twitter-L2 Panel vs. ANES 2020

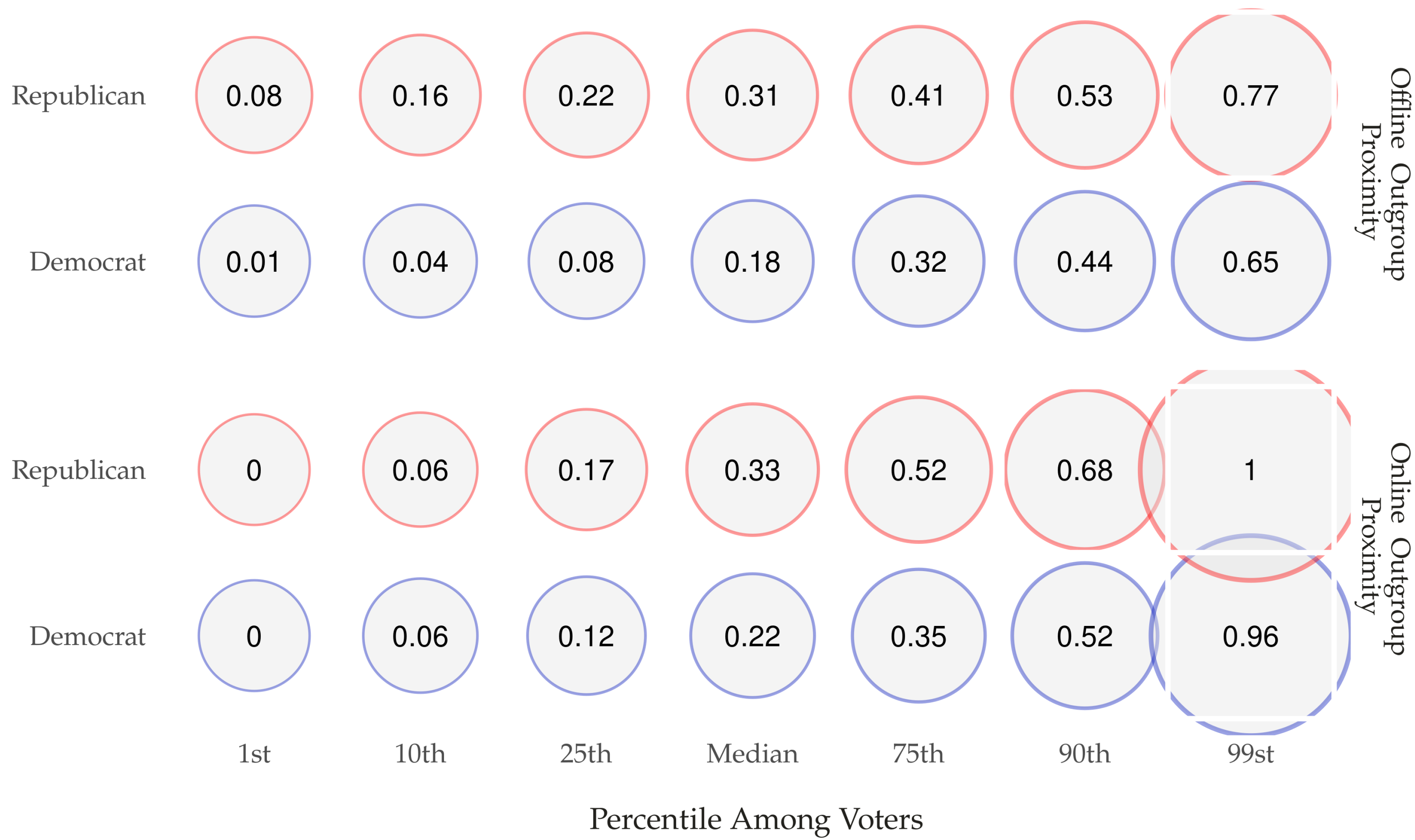
Variable	Twitter-Voter Panel	ANES 2020	Difference	z-score
Age (Mean years)	41.8(0.015)	48.4(0.34)	6.5	19.155⋆⋆⋆
Gender (% Female)	37.8(0)	51.5(0.8)	13.8	17.621 ⋆⋆⋆
Ethnicity (% White)	66.2(0)	65.8(0.7)	-0.004	-0.535
Partisanship (% Democrats)	42.6(0.1)	35.3(0.8)	-7.3	-9.564⋆⋆⋆

Note: Significance of differences between the datasets were tested with a two-sided mean difference t-test (⋆⋆⋆ means $p-value < 0.001$). Sociodemographic population data from the ANES 2020 survey. Standard errors for the values are in parentheses.

Online vs Offline Exposure



Online vs Offline Exposure by Quantiles



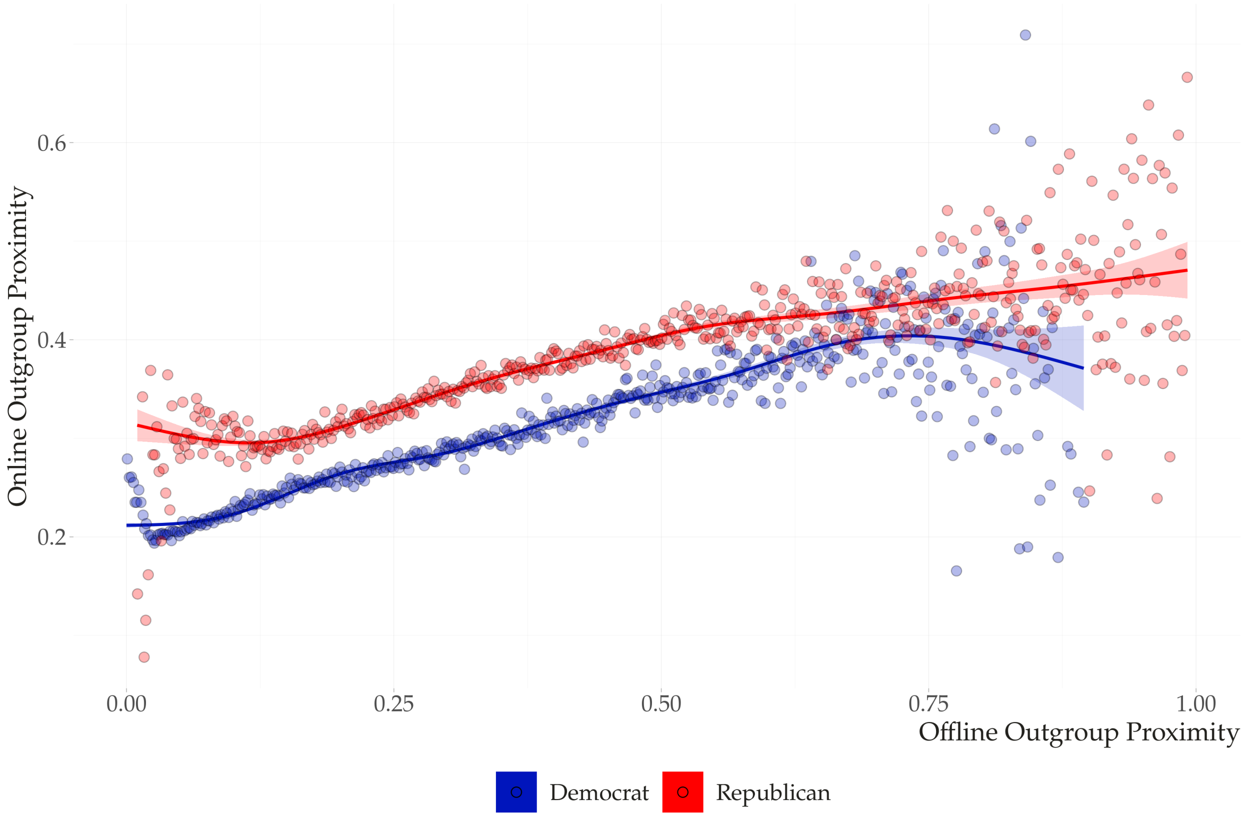
Comparing Offline and Online Exposure Across Subgroups

Table 2 Comparing Online and offline outgroup proximity Across Subgroups

Variable	Online Outgroup Proximity	Offline Outgroup Proximity	Paired-difference	Z-Score(p-Value)
Sociodemographics				
Male	0.306	0.259	0.047	151.094 ⋆⋆⋆
Female	0.280	0.249	0.030	79.614 ⋆⋆⋆
White	0.300	0.278	0.022	76.691 ⋆⋆⋆
Non-White	0.287	0.212	0.074	178.788 ⋆⋆⋆
Age: <35	0.297	0.242	0.055	141.319 ⋆⋆⋆
Age: 35 - 60	0.296	0.259	0.037	112.312⋆⋆⋆
Age: +60	0.289	0.274	0.013	18.051 ⋆⋆⋆
Political Variables				
Democrat	0.261	0.212	0.050	144.023 ⋆⋆⋆
Republican	0.356	0.329	0.023	44.661 ⋆⋆⋆
Blue States	0.256	0.205	0.049	147.366 ⋆⋆⋆
Swing States	0.304	0.270	0.033	66.905 ⋆⋆⋆
Red States	0.331	0.293	0.036	83.729 ⋆⋆⋆

Note: Significance of differences between the datasets were tested with a two-sided mean difference paired z-test(⋆⋆⋆ means $p - value < 0.001$)

Correlation between online and offline exposure

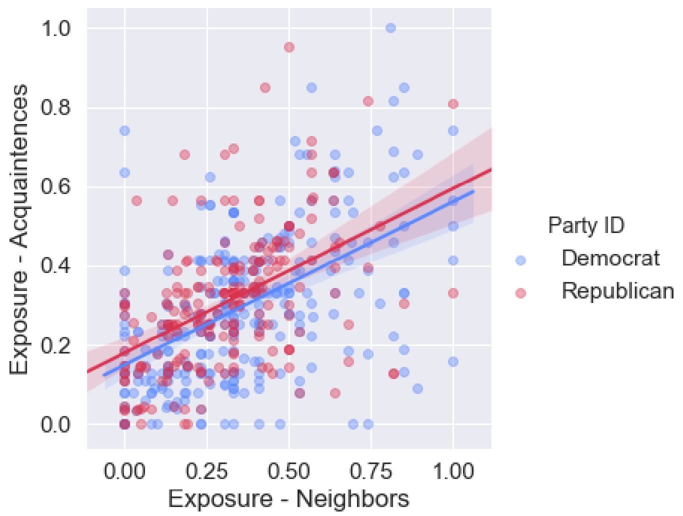


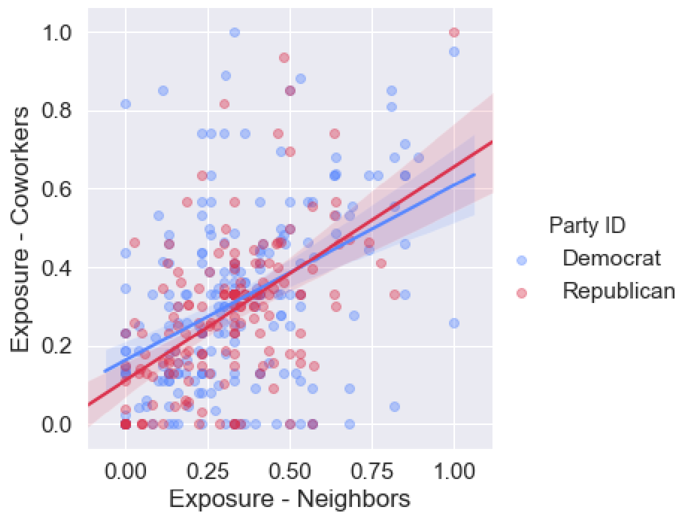
Modeling Online Echo Chambers

Dependent Variable:	Online Proximity		
Model:	(1)	(2)	(3)
Variables			
Age: 35-60	-0.0126**	-0.0122***	-0.0103***
	(0.0059)	(0.0023)	(0.0012)
Age: +60	0.0094	0.0114*	0.0132***
	(0.0170)	(0.0058)	(0.0019)
Male	-0.0260***	-0.0247***	-0.0243***
	(0.0082)	(0.0028)	(0.0011)
White	0.0193***	0.0161***	0.0141***
	(0.0043)	(0.0020)	(0.0012)
Republican	0.0113	0.0143***	0.0165***
	(0.0133)	(0.0033)	(0.0013)
Log of Friends on Twitter	-0.0406***	-0.0427***	-0.0431***
	(0.0022)	(0.0008)	(0.0005)
Offline Proximity	-0.3205***	-0.2593***	-0.2361***
	(0.0523)	(0.0129)	(0.0041)
Fixed-effects			
State-Level (50)	Yes		
Congressional Districts (500)		Yes	
census_tract (47,680)			Yes
Fit statistics			
Observations	599,707	599,707	599,707
R^2	0.05349	0.06785	0.15011
Within R^2	0.03496	0.03166	0.02992

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Robustness Check : Yougov Survey





Challenges

- Joint Scaling Problem
  - Our assumption is that transformation functions provide a connection between composition and potential exposure.
  - While these are not on the same scale, we argue these are comparable approximations.
- Imputation Process
  - Offline: easy. Use Bayesian Process with Precinct data as priors
  - Online: Other strategies to measure partisan identity with Twitter data?
- Correlation between offline on online behavior:
  - What would you be interested in seeing here?

Next Steps

- Focus on the effects of partisan segregation on online behavior
  - Toxicity
  - Outgroup Hostility
  - Sharing of low-quality content
- More methodological articles
  - Comparing different strategies to estimate ideology with online data
  - Use LLMs as an imputation technique and compare to traditional models.

Thank you!