Supplementary Information

Exposure to News Grows less Fragmented with an Increase in Mobile Access

Tian Yang[1], Sílvia Majó-Vázquez[2], Rasmus K. Nielsen[2], Sandra González-Bailón[1, *]

[1]University of Pennsylvania

[2]University of Oxford

* Corresponding author: Sandra González-Bailón

Email: sgonzalezbailon@asc.upenn.edu

This PDF file includes:

Sections

Figures

Tables

www.pnas.org/cgi/doi/10.1073/pnas.2006089117

1. Data

We obtain our observational data from the following sources:

1. Comscore's Media Metrix for PC and Multi-Platform online usage. This panel offers estimates of media use for desktop-only and multi-platform (desktop, mobile and tablet), provided as monthly aggregates at the domain level. The estimates of audience reach (or market share) for a given domain rely on panels that are representative of the online population. In the description of methodology, Comscore reports that an average of ~ 250,000 panelists for desktop and ~35,500 for mobile are active on a given month in the US. Their measurement methodology relies on a combination of panel-based tracking (i.e., what panelists do while online) and census-based site analytics (i.e. traffic information generated at the site level). We compiled the list of news sites by retrieving the set of domains Comscore classifies in the category "News/Information". We then manually checked the list to exclude portals, search engines, and social media sites. Table SI1 shows the list of news sites that we use in our analyses. We analyze audiences visiting these sites for the period October 2014 to December 2018 (data on a small subset of sites are only available since June 2016, these sites are excluded from the analyses that relate to the longer longitudinal trends). Across all these years, we collected data for the US population aged 18 and older, except for the year 2014 when our data targets the population aged 15 and older.

2. Comscore's Plan Metrix Multi-Platform. This panel, a subset of the larger Media Metrix panel, connects attitudes and other self-declared information with online behavior (i.e., websites visited) for the US population aged 18 and older. The survey responses include the questions about party affiliation and political outlook that we use in our analyses. In the description of their methodology, Comscore reports an average of ~ 20,000 panelists complete the survey each month. We analyze data for users that visit the news sites listed in table SI1 for the period September 2017 to December 2018.

In order to get a sense of how our list of news sites compares to the lists used in recent prior work (1-4), we calculated the overlap of news sites across the five studies. We display the results in figure SI1. Panel A shows the percentage of outlets included in each study (rows) that also appear in the other studies (columns). As the figure shows, the five studies have, in general, low overlap in the news domains they consider. Budak et al's study focuses on the 15 top news domains according to views among Bing users. These domains appear in all five studies (with the exception of Yahoo News, which we excluded from our data to be consistent with our decision not to analyze portals; this explains why our list does not have 100% overlap). Our list has the largest overlap with Peterson et al.'s paper, which makes sense since that study also considers web browsing behavior (the studies by Bakshy et al. and Grinberg et al. analyze Facebook and Twitter sharing behavior, respectively). The bars in panel B visualize the proportion of outlets used in our study that were included in the other studies. What this panel shows is that about half of our sites appear in at least one other study and about half appear only in our list (most of these are local news domains, e.g., 13wmaz.com, miamiherald.com).

Overall, the discrepancy among studies in the news sites analyzed derives from the fact that the lists of news sources are compiled from the bottom-up, e.g., only domains that gain some traction in terms of visits (on the web) or shares (in social media) end up in the data analyzed. In section

5 we run robustness tests to determine whether our main findings change if we exclude a subset *N* of news outlets from our data (they do not).

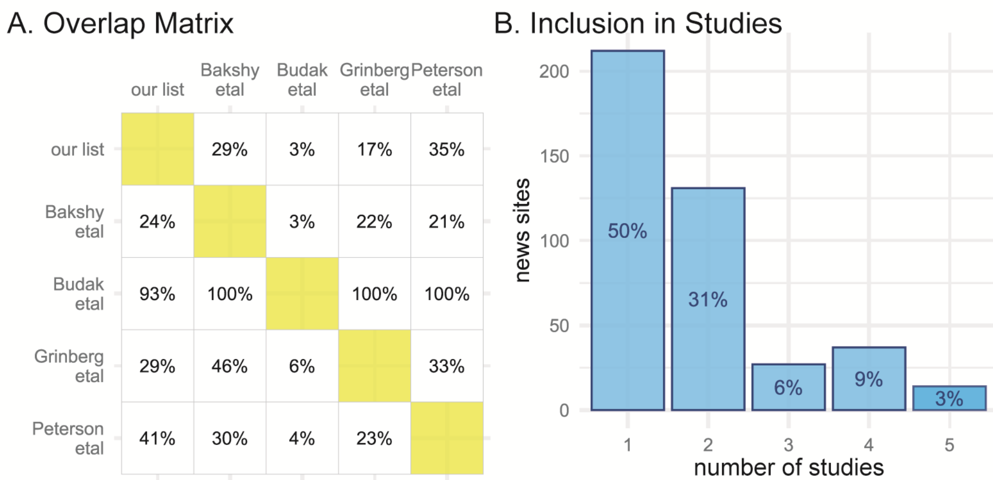

Figure SI1. Overlap in News Sites Considered across Studies. In this figure, we assess the overlap across the lists of news sources used in recent research. Panel A shows that the five studies have, in general, low overlap in the news domains they consider. Panel B visualizes the proportion of outlets used in our study that were also included in the other studies: about half of our sites appear in at least one other study and about half appear only in our list.

2. Co-Exposure Networks

We build the co-exposure networks using the observed audience overlap between news sites. Comscore provides statistics of audience overlap, which we use to build the monthly networks for both Desktop and Multi-Platform access. Because the data is likely to contain random browsing behavior and noise, we filtered the networks according to two techniques. One relies on the pointwise mutual information measure of association (PMI), which we use to identify overlap that occurs to a greater extent than expected by chance. For a pair of news sites (*i*, *j*), the PMI measure is defined as

$$PMI(i,j) = log\frac{p(i,j)}{p(i)p(j)},$$

where $p(i,j) = \frac{online\ audience\ consuming\ i\ and\ j}{total\ online\ audience}$, $p(i) = \frac{online\ audience\ consuming\ i}{total\ online\ audience}$, and

$p(j) = \frac{online\ audience\ consuming\ j}{total\ online\ audience}$.

The filtered networks only retain the edges with a PMI larger than zero (with weighted edges).

Table SI1. List of News Sites included in the Analyses (October 2014 – December 2018)

10tvweb	campusreform.org	deseretnews	idahostatesman	lawnewz
11alive	cantonrep	desmoinesregister	ifyouonlynews	legalinsurrection
12news	cbn	digitaljournal	ijr	lifenews
13newsnow	cbslocal	dispatch	independent	lifesitenews
13wmaz	cbsnews	dowjones	indystar	lifezette
9news	centredaily	drudgereport	inforum	liveaction.org
abclocal	charismanews	eaglerising	inquirer	liverpoolecho
abcnews	chicagotribune	economist	islandpacket	ljworld
about	chicksonright	elitedaily	itvnews	lohud
abqjournal	chron	elnuevodia	jacksonville	louderwithcrowder
activistpost	chronicle	endingthefed	johnstonpressplc	macon
Adn	cincinnati	euronews	joongangilbo	mailonline/dailymail
ahbelo	citizen-times	expressnews	journalnow	mainetodaymedia
americanthinker	clarionledger	fayobserver	jsonline	manchestereveningnews
americasfreedomfighters	clashdaily	firehouse	judicialwatch.org	marketwatch
ammoland	cnbc	firstcoastnews	kait8	mashable
anonews.co	cnn	firstthings	kansas	mcclatchydc
aplus	cnsnews	floridatoday	kansascity	mediageneral
argusleader	coed	forbes	kare11	mediaite
arkansasonline	collective-evolution	foxnews	katc	menstrait
asburyparkpress	coloradoan	freebeacon	kcra	metro
associatedpress	commercialappeal	freerepublic	kens5	miamiherald
atlantablackstar	conservativenewsandviews	fresnobee	kentucky	mic
australianbroadcastingcorp	conservativereview	frontpagemag	kgw	michigan
azcentral	conservativetribune	ft	khou	mirroronline
ba-bamail	countercurrentnews	fusion.net	khq	modbee
baynews9	courierpostonline	gainesville	king5	moonbattery
bbcnews	csmonitor	gazette	kltv	morningnewsusa
belfasttelegraph	dailycaller	gothamist	kmbc	mrc.org
bhmediagroup	dailyherald	grabien	knoxnews	msnbc
billoreilly	dailykos	greenbaypressgazette	krem	msnewsnow
bizpacreview	dailymulligan	greenvilleonline	kron4	mynews13
blackamericaweb	dailyrecord	hamptonroads	krqe	myrtlebeachonline
bloomberg	dailysignal	hannity	ksbw	naplesnews
Bnd	dailystar	havenews	ksdk	nationalreview
boston	dailytidings	heraldextra	ksl	navytimes
bostonglobe	dailyvoice	heraldtribune	ktuu	nbcnews
bostonherald	dailywire	historynet	ktvb	nccommunitygroup
bradenton	dataomaha	hngn	kvue	newsadvance
breitbart	deadstate.org	hotair	kxan	newsbusters.org
burlingtonfreepress	delawareonline	huffingtonpost	lancasteronline	newschannel10
businessinsider	delmarvanow	humanevents	lasvegassun	newscorpaustralia
buzzfeed	democratandchronicle	ibtimesuk	latimes	newsday

Table SI1. List of News Sites included in the Analyses (continued)

news-journalonline	qz	sunherald	toprightnews	wdtn
newsmax	rare.us	sun-times	townhall	westernjournalism
newsnow	readingeagle	takepart	trend-chaser	wfaa
newsobserver	realclearpolitics	tallahassee	tri-cityherald	wgal
newsok	reason	tampabay	trinitymirrorgroup	wgrz
newsone	recordonline	tbo	truthandaction.org	whas11
newspapers	redstate	tcpalm	truthrevolt.org	wishtv
news-press	registerguard	techcrunch	tucson	wisn
newstatesman	rendermedia	telegram	twcnews	wistv
newyorker	reuters	telegraphmediagroup	twitchy	wivb
nola	revealnews.org	tennessean	upi	wkow
noozley	reverbpress	theadvocate	upworthy	wkyc
northjersey	reviewjournal	theamericanconservative	usatoday	wlox
npr	rgj	theantimedia.org	usherald	wlwt
nwfdailynews	rushlimbaugh	theatlantic	usnews	wmur
nydailynews	sacbee	theblaze	usuncut	wnd
nypost	sanluisobispo	thecharlotteobserver	vcstar	wndu
nytimes	savannahnow	theconservativetreehouse	veteranstoday	woodtv
oann	scoopnest	thedailybeast	vicemedia	worldnewsdailyreport
observermedia	scotsman	thedesertsun	voanews	worldnewsnetwork
occupydemocrats	scrippsnationalnews	thediplomat	voiceofeurope	wpbf
ocregister	sctimes	theepochtimes	vox	wpri
ogdennewspapersinc.	seacoastonline	thefederalist	walb	wrcbtv
ohio	seattletimes	thegatewaypundit	walesonline	wric
onenewsnow	sfexaminer	theguardian	wane	wsbt
onenewspage	sj-r	thehill	washingtonexaminer	wsfa
onlineathens	skynews	theledger	washingtonpost	wsj
ozock	slate	thenewamerican	washingtontimes	wthr
pbsfrontline	sltrib	theneworleansadvocate	wate	wtnh
pbsnewshour	southbendtribune	thenewstribune	wattsupwiththat	wtol
pe	southeastmissourian	theolympian	wave3	wtsp
philly	spectator	thepatriotnation.net	wavy	wwlp
pjmedia	spectator.org	therightscoop	wbaltv	wwltv
pjstar	spectrumlocalnews	thestate	wbay	wxyz
pnj	spokesman	thesun	wbir	wyff4
politico	standard	thetab	wboc	wzzm13
postcrescent	standardnews	thetimes	wbtw	yournationnews
post-gazette	starnewsonline	theweek	wcax	
poughkeepsiejournal	star-telegram	thinkprogress.org	wchstv	
powerlineblog	startribunenetwork	threepercenternation	wcnc	
pressdemocrat	statesmanjournal	time	wcpo	
propublica.org	stripes	timesfreepress	wcsh6	
providencejournal	success	toledoblade	wdbj7	

The second technique is based on backbone extraction, which is well established in network science (5). This technique operates on the assumption of a null model of random weight distribution at the node level used to assess the significance of tie strength in the observed network. Significance is determined through a parameter alpha that can be interpreted as the conventional *p*-value. We use $\alpha = 0.05$, a threshold under which many more ties are eliminated compared to the PMI approach. The main text shows the networks filtered according to these two techniques. Figure SI2 shows the same trends according to the raw, unfiltered data. All three networks give a consistent picture of decreasing density for desktop data, and increasing density for multi-platform data.

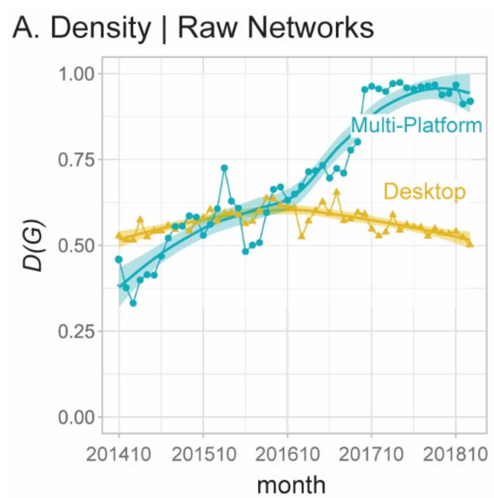

Figure SI2. Density Trends according to Unfiltered Data. This figure shows the evolution of density scores for the co-exposure networks prior to any thresholding eliminating the weakest ties.

Figure SI3 offers a summary of changes in other network statistics. In addition to density, we also looked at centralization, which is a graph-level measure of inequality in node connectivity (6); transitivity, which is a measure of clustering at the local, triadic level (7); and modularity, a measure of clustering at the meso-scale based on community detection (8). The co-exposure networks we analyze are very dense (especially those built with multi-platform data) so, as expected, centralization scores are low and transitivity scores are high. The modularity scores allow us to determine if the networks can be characterized by the existence of communities, that is, subsets of news outlets that are more densely connected internally than externally. The community detection algorithm that we use (8) is based on a random walker and it is particularly well suited for weighted networks like ours. The modularity scores being so close to zero suggest there is no clear community structure characterizing these networks. Most of the information in the network data we analyze lies on the edge weight distribution (e.g., the cell values in the adjacency matrix capturing the strength of audience overlap). The MRQAP models we discuss in the main text use this adjacency matrix as the dependent variable and they determine how relevant ideological similarity is in explaining the variance in the strength of those network ties.

Figure SI3. Changes in Network Statistics. The panels in this figure show changes in three network statistics for the PMI and backbone networks: centralization (upper row), transitivity (middle row), and modularity (lower row).

To give some additional context to the changes in network density we depict in the main text (which suggests mobile access significantly diversifies news diets), figure SI4 summarizes survey data showing the rapid increase in mobile access to news in the past five years. These data come from two different surveys: Pew and the Reuters Digital News Report; but they are consistent in identifying how rapidly mobile access has increased for news consumption. These trends actually look qualitatively similar to the patterns we identify with our observational data: mobile is on the rise, to the detriment of desktop-only access. Our argument is that mobile access is diversifying news exposure in terms of sources, hence the rising density in the co-exposure networks we analyze. We show that if we only use desktop data to assess fragmentation, we miss this trend – and this is important since desktop browsing data is the predominant type of data used in research until now (even though, as survey data also suggests, desktops are becoming increasingly less relevant as the main device used to consume news).

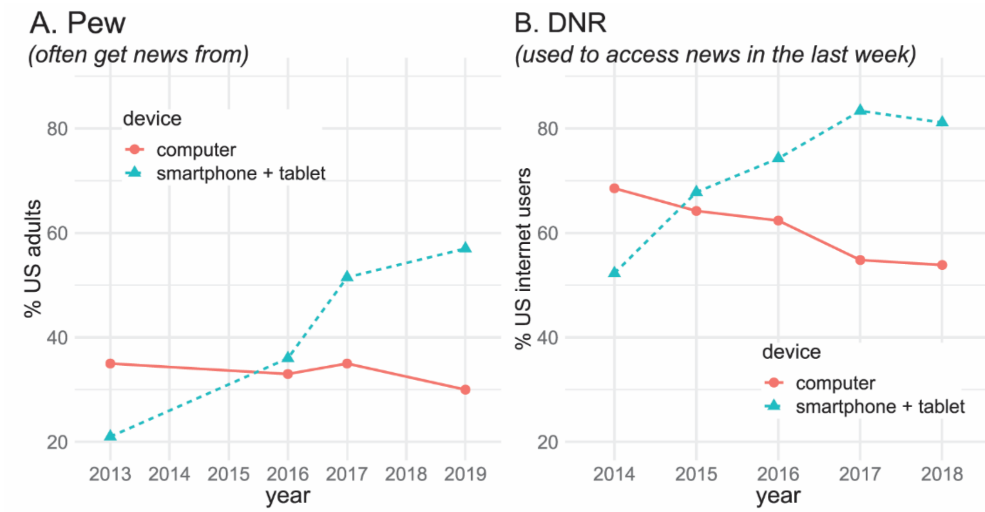

Figure SI4. Survey Data on Device Used to Access News. Panel A shows changes in the percentage of the US population who often gets news using computer and mobile devices (smartphone and tablet), according to Pew data. Panel B shows changes in the percentage of internet users who access news using computers and mobile devices, according to the Digital News Report data.

3. Ideological Measures

We assess the statistical significance of the segregation scores for desktop and multi-platform data (figure 3, panels A and B in the main text) with a Wilcoxon signed rank test. This test confirms that the differences between desktop and multi-platform data are statistically significant: desktop data overestimates the segregation score for both partisanship (p = .003) and political outlook (p < .001). The (pseudo) median value for the discrepancy is .010 for partisanship (which is 36% of the median value of this measure for multi-platform data across all months, .027), and .009 for political outlook (which is 38% of the median value of this measure for multiplatform data across all months, .025). In other words, desktop data overestimates more than 35% of the segregation levels compared to multi-platform data.

Figure SI5 compares the favorability scores (9) of news sites to the total online population. To calculate the favorability score for political outlook, we aggregated the responses ‘very liberal’ and ‘somewhat liberal’ as ‘liberal’; and the responses ‘very conservative’ and ‘somewhat conservative’ as ‘conservative’. For the most part, the audiences consuming news identify to a lesser degree with Democrats than the total online population, and they lean towards more conservative views.

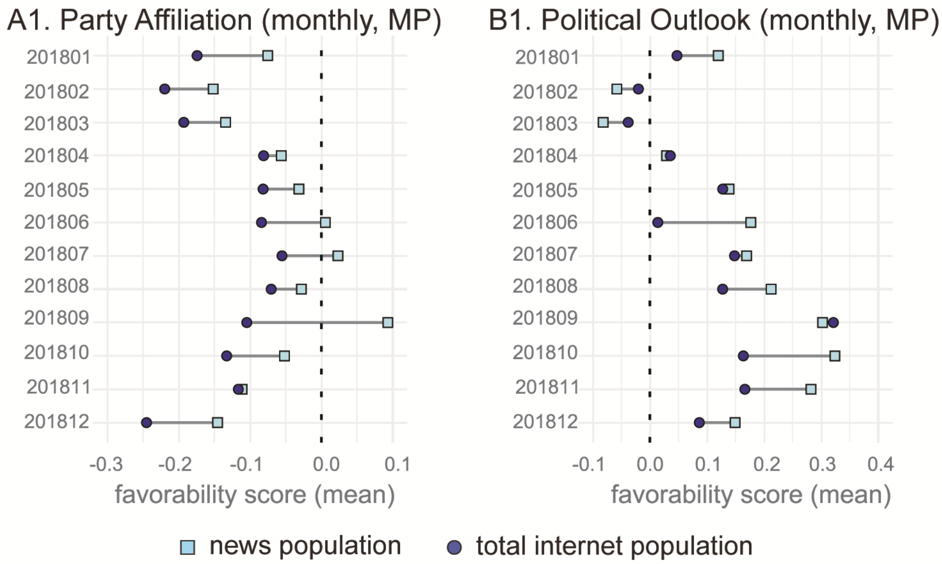

Figure SI5. Comparison of News Consuming Population Total Online Population. Panel A compares monthly ideological scores based on party affiliation of the news consuming population with the scores for the total internet population. Panel B compares monthly scores based on political outlook.

Figure SI6 compares Comscore's measures of party affiliation with Gallup measures. The figure does not show serious disagreements.

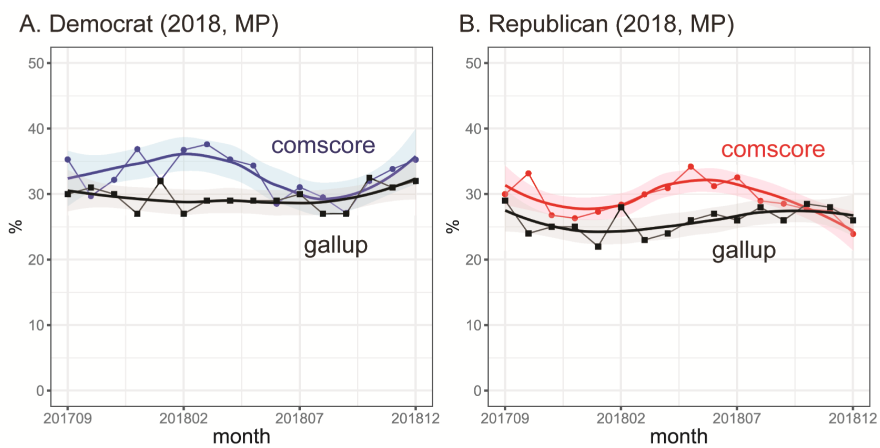

Figure SI6. Comparison of Comscore and Gallup Measures of Party Affiliation. Panel A shows changes in the percentage of the population who identify as democrat according to Comscore and Gallup estimates. Panel B shows changes in the percentage who identify as Republican.

Figure SI7 shows the monthly distributions of the favorability scores. They confirm that, overall, online news consumption audiences tend to lean Democratic and towards conservative views.

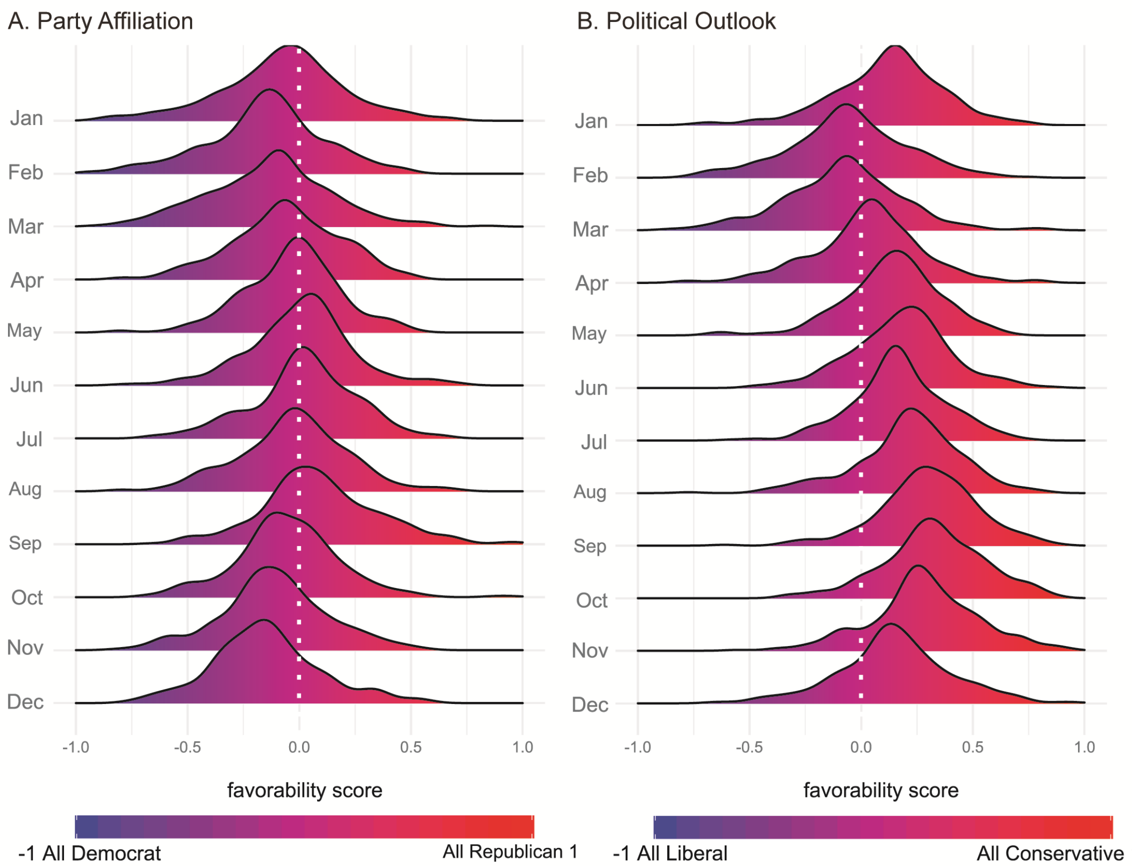

Figure SI7. Monthly Distributions of Favorability Scores (2018, MP). Panel A shows monthly distributions for the favorability scores of news domains according to party affiliation. Panel B shows monthly distributions for political outlook.

Even though the top news sites in terms of audience reach converge towards the median values of the favorability scores, figure SI8 shows that there is no correlation between the ideology measures and the percentage reach of news sites or the average time users spend on those domains. In section 6 we show the results of robustness tests in which we exclude from our data news sites where users spend less than the median average time. Our main results remain qualitatively unchanged.

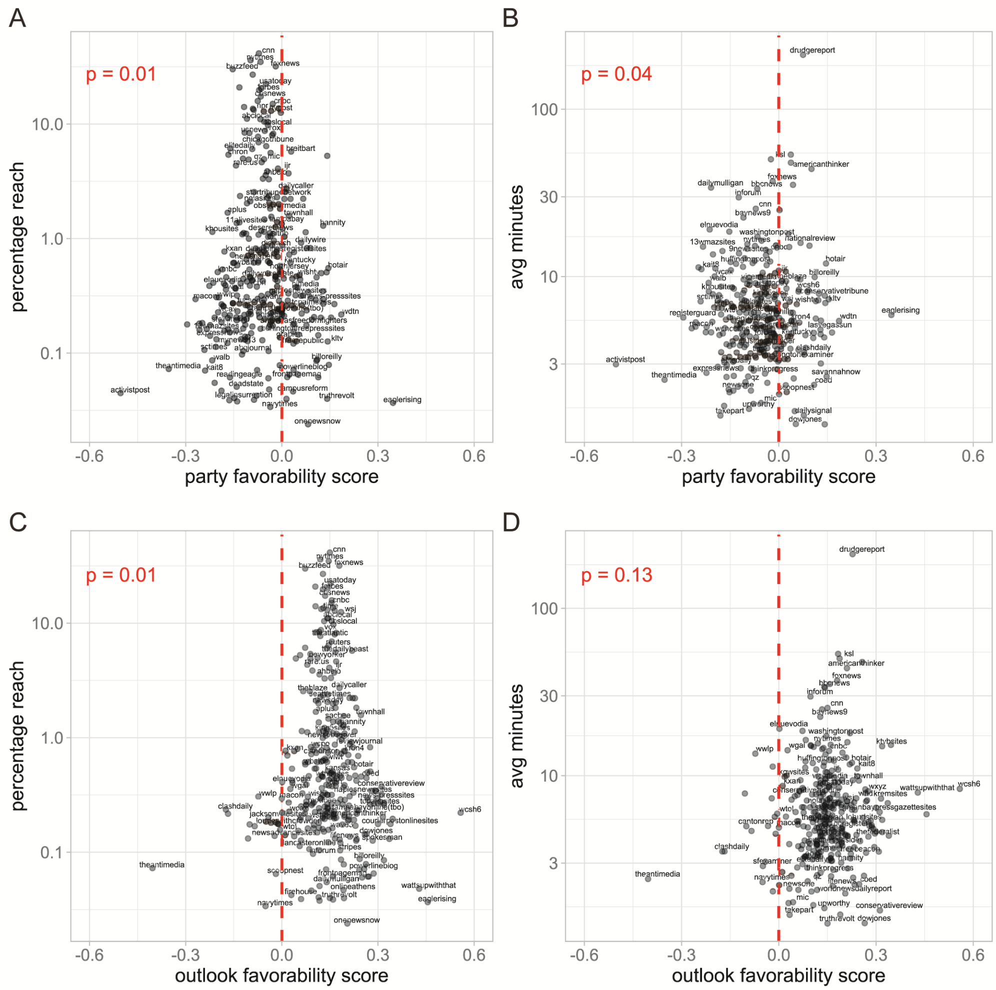

Figure SI8. Correlation of Favorability Scores with Audience Reach and Engagement. The panels in this figure show the lack of correlation between the percentage reach of news sites and the average time spent of their audiences with the favorability scores based on party affiliation (A-B) and political outlook (C-D).

4. QAP correlations and models

The analysis of co-exposure networks precludes the use of standard statistical techniques that assume independence in the units of observation. Therefore, to test the selective exposure hypothesis we make use of multiple regression quadratic assignment procedures (MRQAP). This involves fitting a linear model for square matrix data:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + E,$$

where Y, X, Z, and E are $n \times n$ matrices where the rows and columns correspond to news outlets. We use three different matrices as the dependent variable Y: the raw co-exposure network; the PMI version, and the backbone version (all derived from MP access, log-transformed). Across the three models, the matrices X and Z measure the same thing: pairwise distance in the party affiliation and political outlook scores, respectively. We assume linear dependence between the variables and, as usual, the null hypothesis is that β_1 and β_2 will not be statistically significant from zero (10). The permutation tests give us strong evidence that we can reject the null hypothesis.

Figure SI9 shows the QAP correlations for the raw network data.

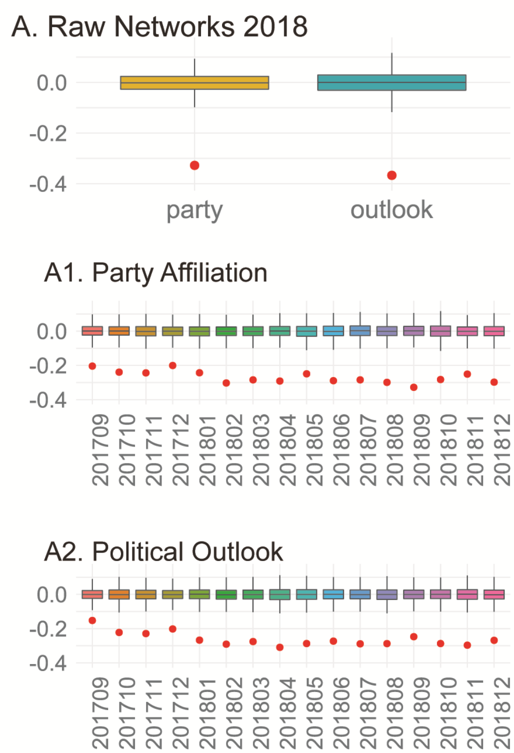

Figure SI9. Correlation of Audience Overlap and Ideological Scores for Unfiltered Data. The figure shows yearly and monthly correlations between pairwise distance in the favorability score distributions and audience overlap between news sites.

5. Node Removal Robustness Tests

In order to determine how much our results would change if a subset *N* of sites were removed from our data, we reproduced our main findings under the random removal of N = 25, 50, 75 and 100 news sites. Figures 10 to 14 show that results remain qualitatively unchanged (the confidence intervals in these figures are based on 1,000 random iterations).

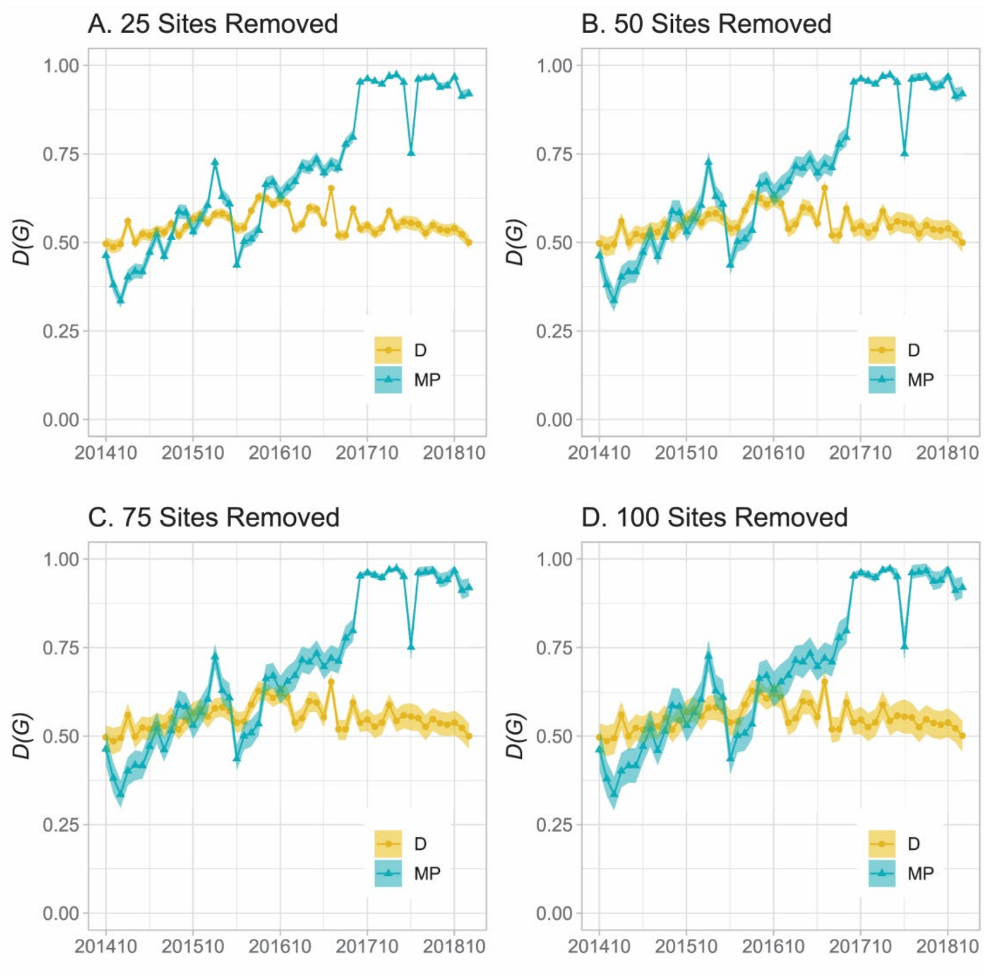

Figure SI10. Changes in Network Density for Different Node Removal Conditions (Raw Data). The figure shows changes in network density as a different subset of nodes *N* are randomly removed from the networks. Confidence intervals are based on 1,000 random iterations.

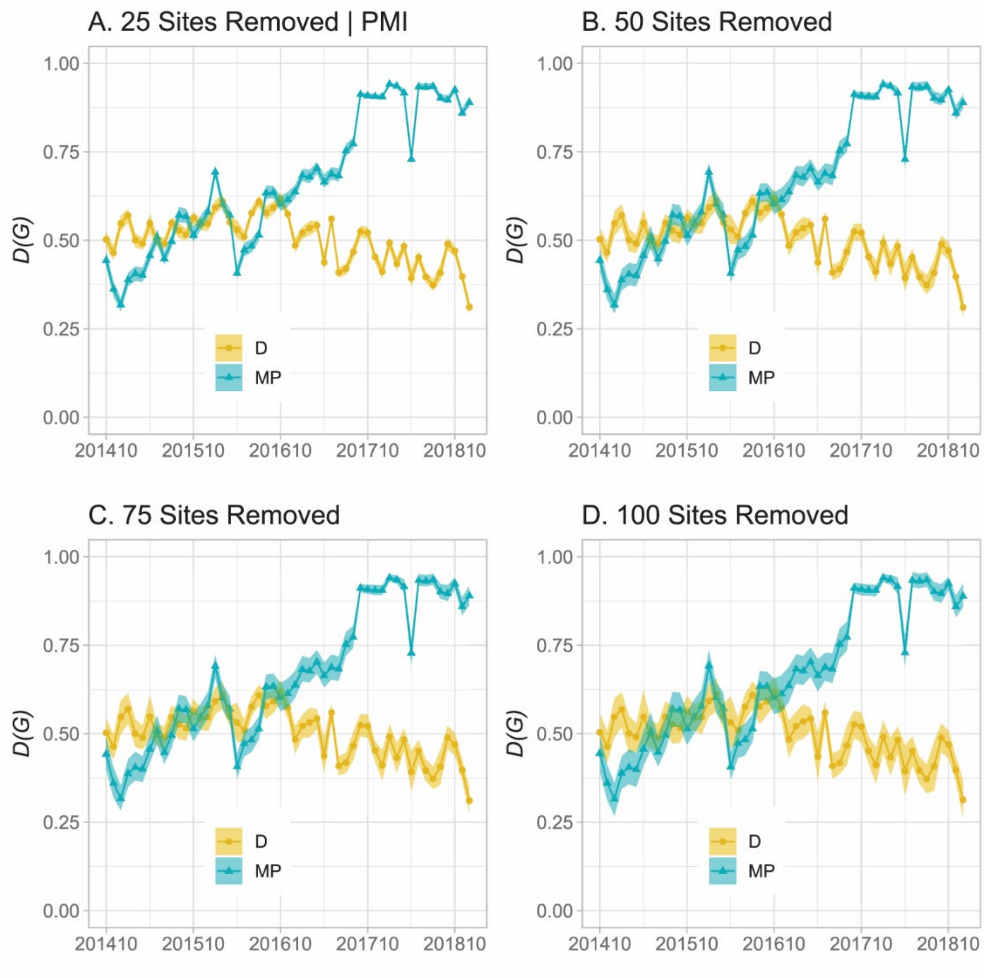

Figure SI11. Changes in Network Density for Different Node Removal Conditions (PMI Data). The figure shows changes in network density as a different subset of nodes *N* are randomly removed from the networks. Confidence intervals are based on 1,000 random iterations.

Figure SI12. Changes in Network Density for Different Node Removal Conditions (Backbone Data). The figure shows changes in network density as a different subset of nodes *N* are randomly removed from the networks. Confidence intervals are based on 1,000 random iterations.

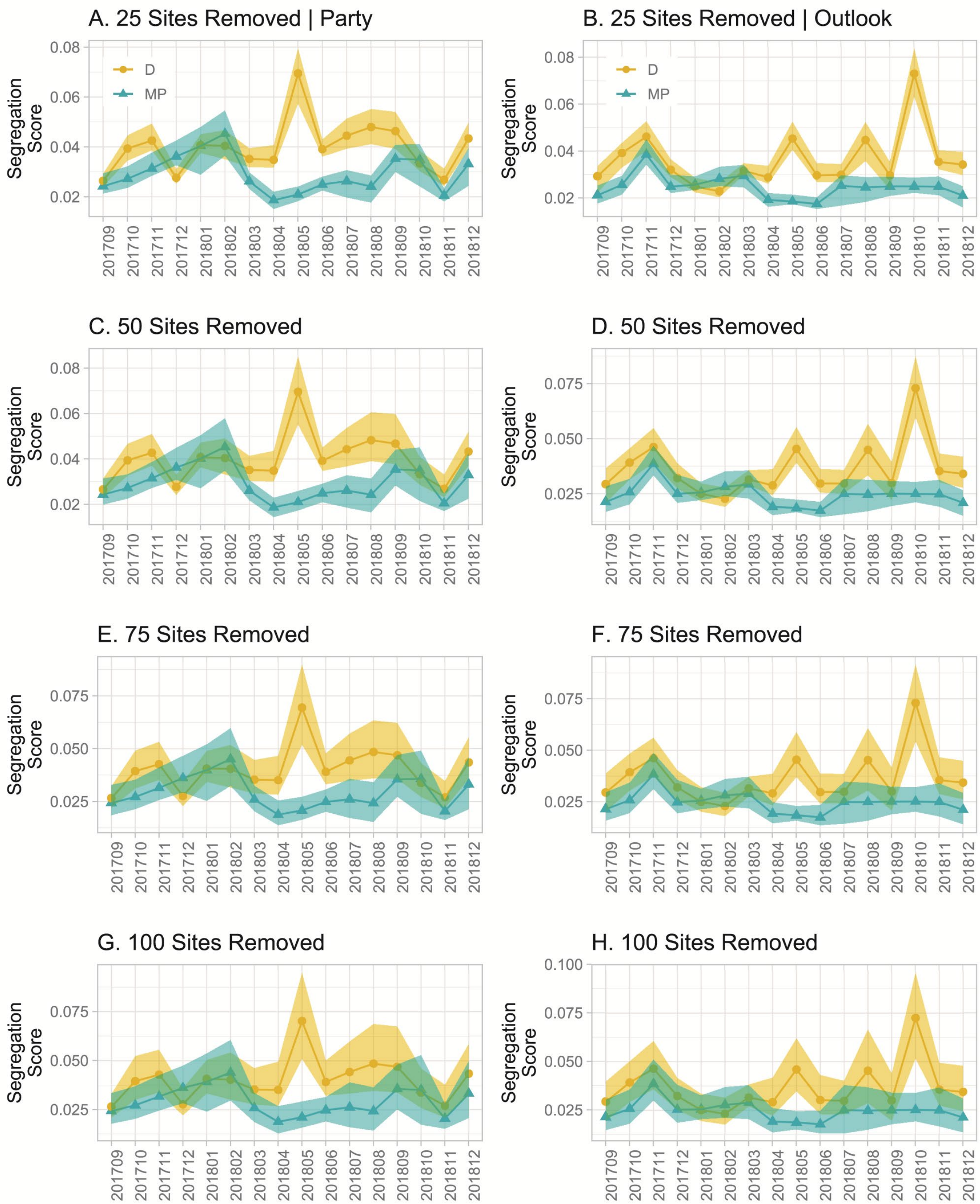

Figure SI13. Changes in Segregation Scores for Different Node Removal Conditions. The figure shows changes in the segregation scores as a different subset of nodes *N* are randomly removed from the networks. Confidence intervals are based on 1,000 random iterations.

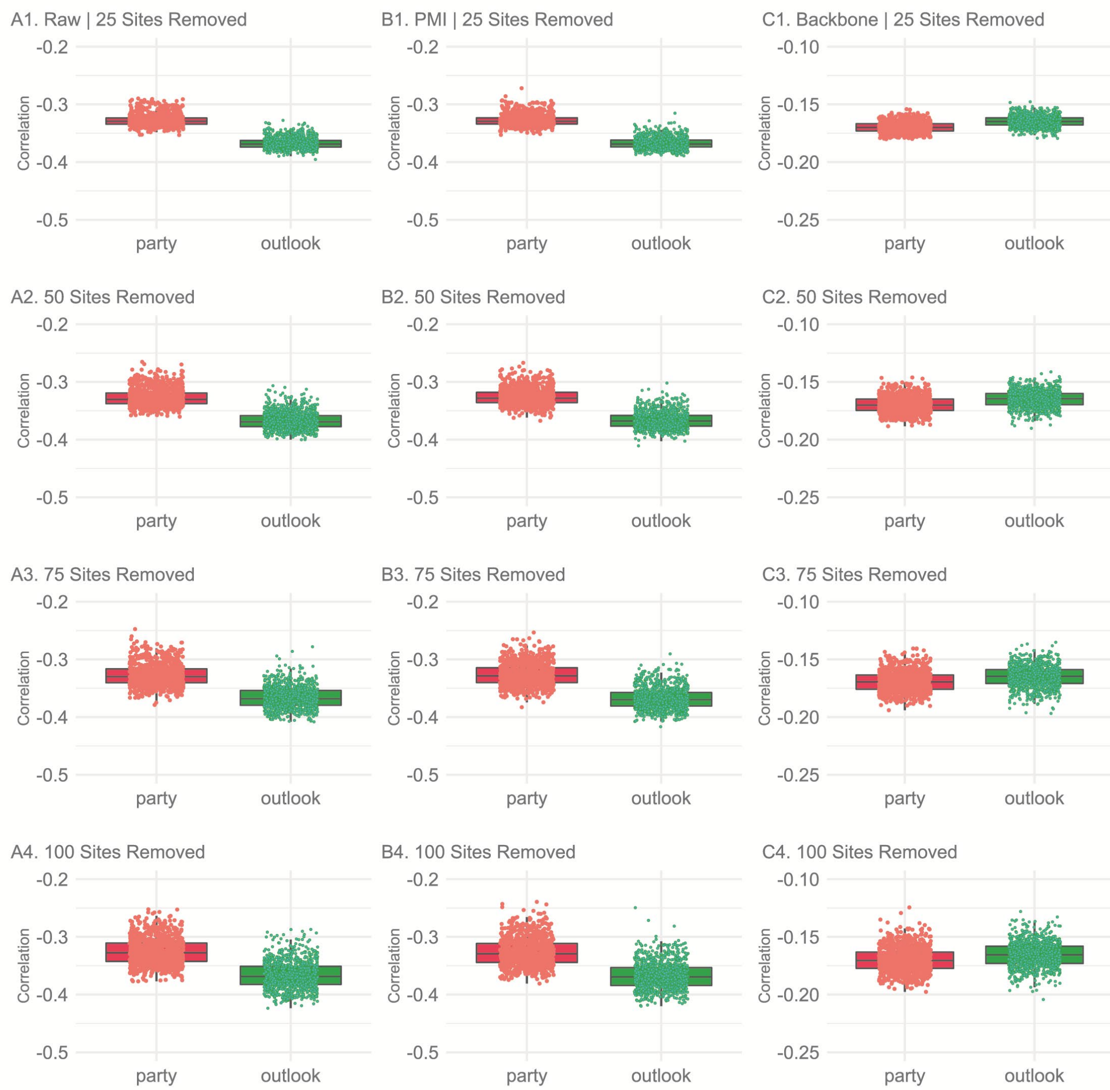

Figure SI14. Changes in QAP Correlation for Different Node Removal Conditions. The figure shows changes in the QAP correlations between audience overlap and pairwise distance in the favorability score distributions as a different subset *N* of nodes are randomly removed from the networks. Plotted values are drawn from 1,000 random iterations.

6. Core Network Robustness Tests

In order to determine if our main findings hold when we only pay attention to news outlets in which users spend more time (a measure that we take as a proxy to intensive use or "serious reading"), we reproduced all our analyses excluding news sites below the median time. In other words, we focus on sites on the higher half of the "average time spent" distribution. As a shorthand, we call this subset of news outlets "core network". Figures SI15 to SI17 show the outputs of this test, which reveal that the main trends remain qualitatively unchanged. (The

Wilconsox test for figure SI16 also suggests statistically significant differences in the median: p = .013 for partisanship; p = .002 for political outlook).

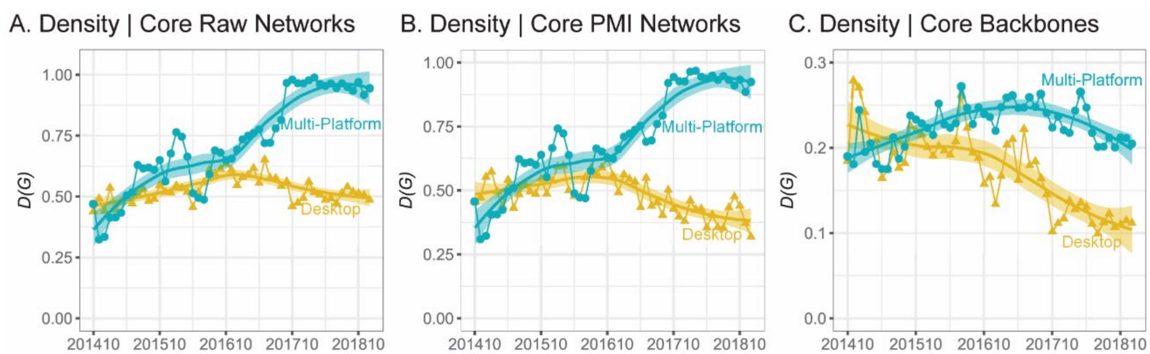

Figure SI15. Density Trends for Core Networks. The panels plot changes in network density for the raw network (A), the PMI networks (B), and the backbones (C) without news outlets in which audiences spend less than the median time.

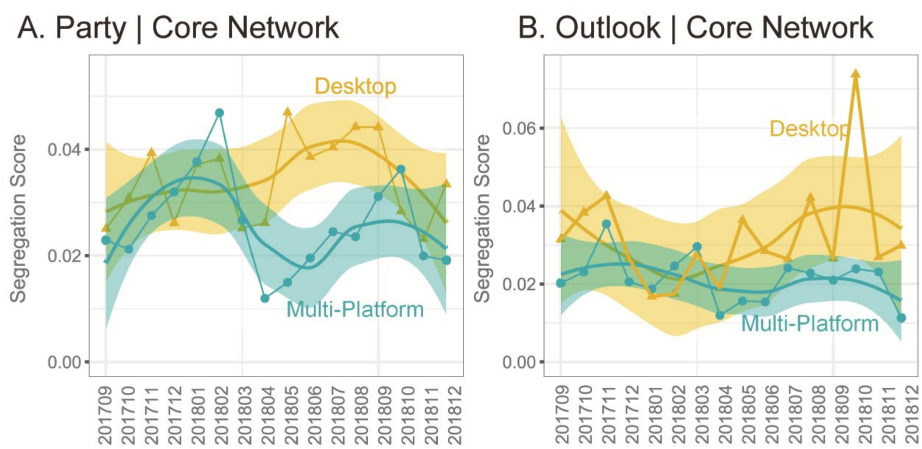

Figure SI16. Segregation Scores for Core Networks. The panels plot changes in the segregation score for news outlets in which audiences spend more than the median time.

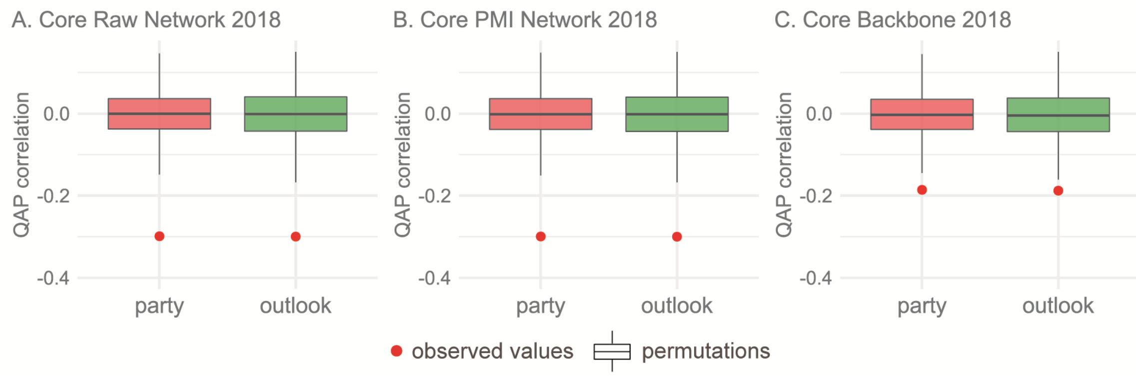

Figure SI17. QAP Correlations for Core Networks. The figure shows correlations between pairwise distance and favorability scores for party affiliation and political outlook considering only news sites in which audiences spend more than the median time.

7. References

1. Budak C, Goel S, & Rao JM (2016) Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis. *Public Opinion Quarterly* 80(S1):250-271.

2. Bakshy E, Messing S, & Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* 348(6239):1130-1132.

3. Peterson E, Goel S, & Iyengar S (2019) Partisan selective exposure in online news consumption: evidence from the 2016 presidential campaign. *Political Science Research and Methods*:1-17.

4. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, & Lazer D (2019) Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363(6425):374-378.

5. Serrano MÁ, Boguñá M, & Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences* 106(16):6483-6488.

6. Freeman LC (1977) A Set of Measures of Centrality based upon Betweenness. *Sociometry* 40:35-41.

7. Barrat A, Barthélemy M, Pastor-Satorras R, & Vespignani A (2004) The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* 101(11):3747-3752.

8. Pons P & Latapy M (2006) Computing Communities in Large Networks Using Random Walks. *Journal of Graph Algorithms and Applications* 10(2):191-218.

9. Tyler M, Grimmer J, & Iyengar S (2019) Partisan Enclaves and Information Bazaars: Mapping Selective Exposure to Online News. *Working Paper*.

10. Dekker D, Krackhardt D, & Snijders TAB (2007) Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika* 72(4):563-581.