MIS Quarterly

RESEARCH ARTICLE

NETWORK STRUCTURE AND PATTERNS OF INFORMATION DIVERSITY ON TWITTER[1]

Jesse Shore
Questrom School of Business, Boston University, 595 Commonwealth Avenue,
Boston, MA 02215 U.S.A. {jccs@bu.edu}

Jiye Baek
Hong Kong University of Science and Technology, Clear Water Bay,
Kowloon HONG KONG {jiyebaek@ust.hk}

Chrysanthos Dellarocas
Questrom School of Business, Boston University, 595 Commonwealth Avenue,
Boston, MA 02215 U.S.A. {dell@bu.edu}

Appendix A

Robustness Check on Recent Data

Summary

We conducted an extensive robustness check on recent (February–March, 2017) data. We calculated our own measure of slant by analyzing patterns of co-following among people who follow news sources directly. We then assembled a sample of Twitter users and analyzed incoming versus outgoing slant. To the extent that we were able to conduct analogous analyses, our main result stands with current data.

Calculation of Slant Score

In our main analysis we had an externally validated measure of slant that corresponded to the same period of time as our data was gathered. We felt that the measures from 2009 were much too old to rely on for current data, and that many important news sources (for example, Wikileaks and fivethirtyeight.com) had arisen in the intervening years. We thus devised our own measure that was closely related in principle to the measure from Gentzkow and Shapiro (2011) for the 2009 data. In particular, we use their notion that bias is captured by audience segregation. In their methods, a news site was deemed conservative (i.e., having a positive slant) if it was relatively heavily read by conservative people and had relatively few liberal people among its audience. News sites with centrist scores (near zero) were read by both conservative and liberal people. We adapt this idea to Twitter data as detailed here.

We began by identifying news sources to include in our slant calculations. We created a list of 186 different sources, including most of the sources from the 2009 data (we excluded sources without at least 10,000 followers on Twitter, as well as blogtalkradio.com, because it is a platform for others to share their content and at this point in time does not have a predictable slant). We then extended the original list by searching for curated lists of political news sources and added all of those as well (limiting our sources to those that had at least 10,000 followers on Twitter). We also separated CNN politics from the rest of CNN, *The Wall Street Journal* opinion section from the rest of *The Wall Street Journal*, and the Politics and Black Voices sections from the rest of the Huffington Post, as those divisions are thought to have a different slant from the rest of the reporting on those sites, according to multiple comments we read.

We then identified the main twitter handle of each of these news sources and used the Twitter API to download the lists of direct followers of these news sources. This step took approximately 7 weeks of continuous querying of the API in order to collect 330,000,000 follower relationships to our list of 186 news sources (these 7 weeks were divided across multiple developer accounts so the actual time spent collecting data was closer to 2 weeks). Collecting data at this scale necessitated building a collection script that was robust to undocumented behaviors of the API.

Having collected the lists of followers, we then calculated a 186 × 186 affinity matrix, the entries of which represent how much each pair of news sources overlapped in terms of followers. For example, at the time of data collection, *The New York Times* had approximately 33,000,000 followers on Twitter, and *TheBlaze* (Glenn Beck) had approximately 616,000 followers; about 165,700 people followed both sites, which is 26.9% of the size of the smaller audience. We thus use 0.269 as the entry in our affinity matrix for the strength of the connection between *The New York Times* and *TheBlaze*. In general, denoting our affinity matrix as A,

$$\mathrm{A_{ij}} = \frac{\left|\mathrm{Followers_i} \cap \mathrm{Followers_j}\right|}{\min\left(|\mathrm{Followers_i}|, |\mathrm{Followers_j}|\right)}$$

where |X| denotes the cardinality of set X.

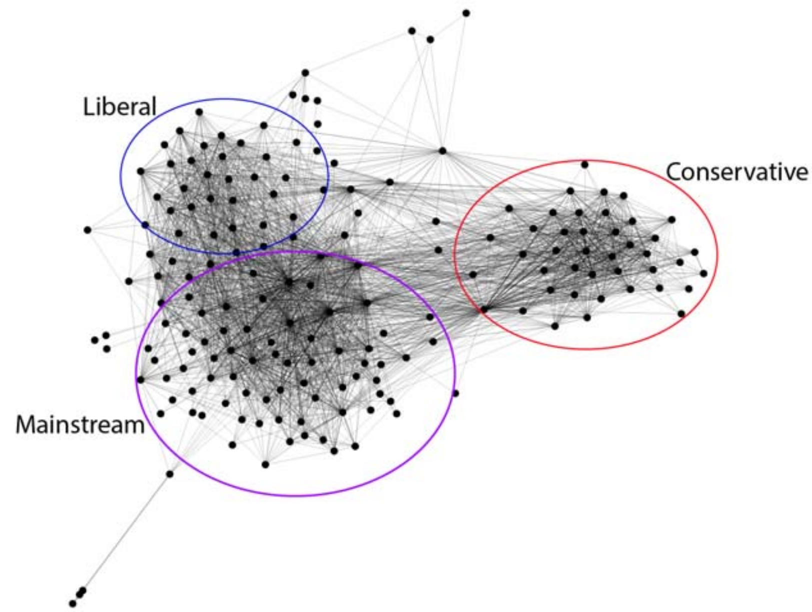
Ultimately, our slant scores are calculated from this raw affinity matrix. However, analogous to Gentzkow and Shapiro's method, we needed to first identify subsets of news sources that were clearly embedded in a partisan cluster such that we had a basis for approximating partisan audiences. The intention behind this is to use co-following of sites in the conservative cluster as an indicator of conservatism, and co-following of sites in the liberal cluster as an indicator of liberalism.

However, as others have recently pointed out (e.g., see the preliminary research report by Benkler et al. 2017), the conservative news cluster is clearly distinct, while the liberal news cluster is more integrated with the mainstream news cluster (that is, people who follow news sources with a liberal slant also follow mainstream news sources more than do people who follow news sources with a conservative slant), as Figure A1 illustrates. Because of this overlap, community detection algorithms run on the raw affinity matrix tend to divide the news site co-following network into only two rather than three communities. However, in order to calculate an analogous measure of slant to the one in our main analysis, we needed to identify which sites were in the less distinct, smaller cluster of clearly liberal-leaning sites.

As an intermediate step, we analyzed a scaled version of the matrix rather than the raw matrix to emphasize the community structure and provide a clearer signal of partisan separation. In particular, we reasoned that some co-following is to be expected if only based on random chance. For example, for a very small niche news site, it wouldn't be terribly surprising if a substantial fraction of its audience also followed CNN or *The New York Times*. Intuitively, the smaller the first site, the less surprising it would be if a large fraction of its audience also followed larger news sites as well. Thus, we scale the entries in the affinity matrix by the ratio of the natural logs of the number of followers for each news site. This penalizes co-following with sites that are much larger and emphasizes co-following that is more unexpected with respect to relative audience size. After thus scaling the matrix, we removed all edges with a scaled weight less than 0.3 and trimmed nodes with resulting degree less than 5. These latter steps emphasize stronger connections and greater embeddedness in communities. We then ran the spin glass-based community detection algorithm (Reichardt and Bornholdt 2006) on this preprocessed network to find three clear communities.

Having thus identified nodes belonging to a liberal and a conservative cluster, we return to the raw affinity matrix to calculate slant scores (that is, the preprocessed version of the matrix was only used to detect membership in the three clusters). For each news site, we calculated the mean co-following between that site and sites in the liberal cluster divided by the mean co-following with all sites in the network and multiplied that by negative one to represent the liberal-leaning tendency. We then added that to the mean co-following between that site and sites in the conservative cluster divided by the mean co-following with all sites in the network to represent the conservative leaning slant. This sum represents the estimated slant for each site about which we collected data.

$$Slant_i = -1 \times \left(\frac{\frac{1}{|liberal|} \sum_{j \in liberal} A_{ij}}{\frac{1}{n} \sum_{j \neq i} A_{ij}} \right) + 1 \times \left(\frac{\frac{1}{|conservative|} \sum_{j \in lconservative} A_{ij}}{\frac{1}{n} \sum_{j \neq i} A_{ij}} \right)$$



Note: For purposes of visualization only, weak ties (those less than 0.25) are omitted.

Figure A1. Visualization of Raw Affinity Matrix

Our resulting estimated slant scores correspond fairly well to other published measures of slant, such as Pew Research Center's 2014 report on political polarization. The correlation between our slant scores and Gentzkow and Shapiro's 2009 scores is 0.79; the two news sources with the greatest difference between our score and Gentzkow and Shapiro's scores are the hotairblog.com and breitbart.com. Both were relatively new at the time of Gentzkow and Shapiro's data collection, and Breitbart has been documented to have evolved substantially since 2009.[1]

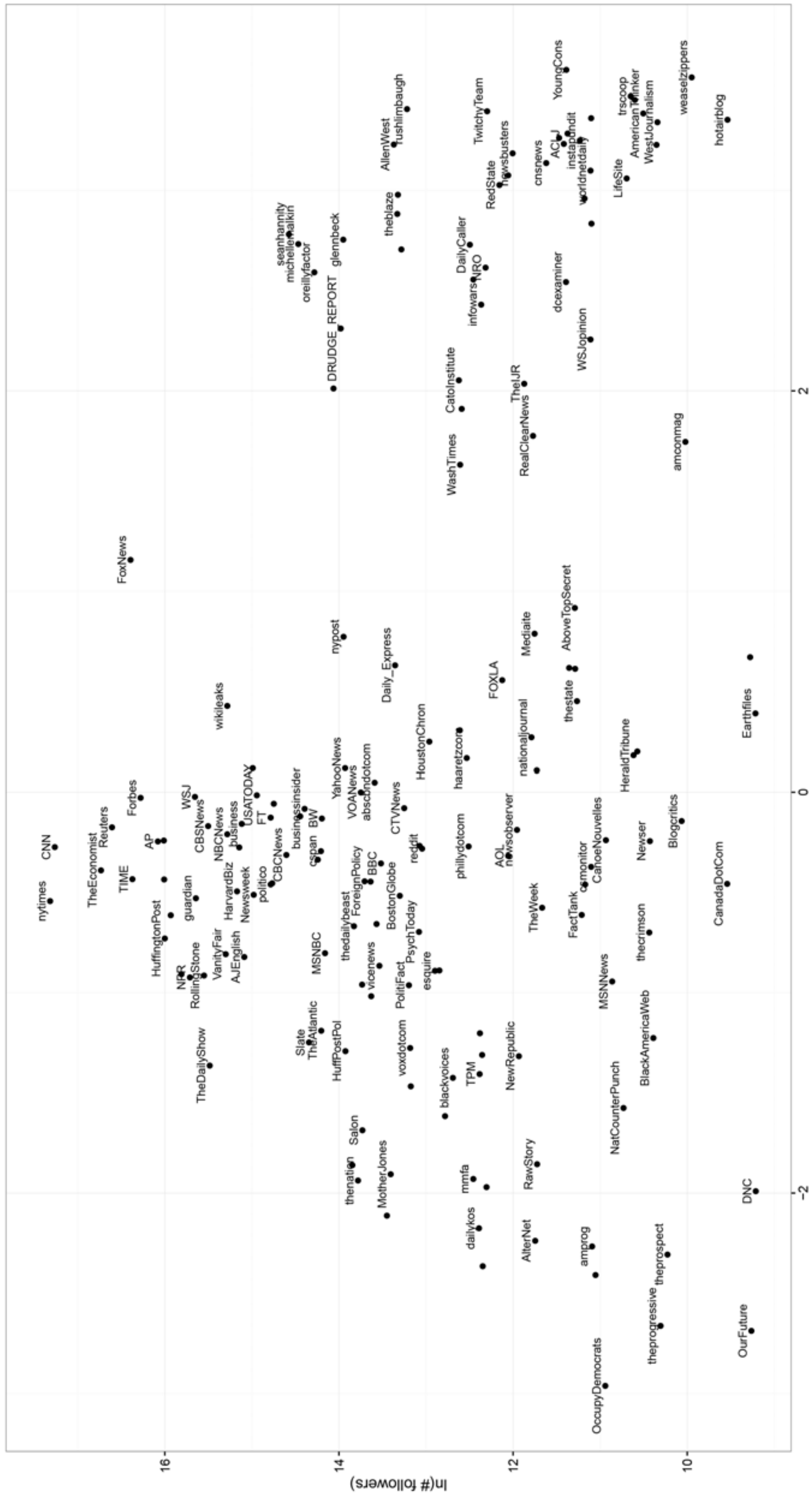
Because liberals are more likely to be exposed to mainstream news sources, a truly unbiased source would not necessarily have a slant score of zero. Note, for example, that C-SPAN, a source that largely presents footage of congress without commentary, has an estimated slant score of approximately -.29 on the scale for our 2017 data (which is on a different scale than Gentzkow and Shapiro's slant scores). Estimated slant scores are plotted in Figure A2 and reported in full in Table A1.

Collection of Data on Individual Accounts

As indicated above, only a minority of Twitter accounts actually tweet links to news articles; sampling accounts at random (to the extent that is possible) to find those that both received and sent tweets containing links to news articles would be prohibitively expensive in terms of time. In order to find accounts more efficiently, we instead began by searching for tweets containing links to news stories, found the accounts that sent these tweets, and then searching among the followers of these accounts until we found those that also tweeted at least one link. Using this method, we obtain the full recent timeline (up to 3,200 tweets, including only those less than or equal to 2 weeks old) for both the upstream followee and the downstream follower. We use the mean slant of the followee's tweets as a predictor variable and the mean slant of the follower as the dependent variable in regressions that we use as a robustness check of our main result.

---

[1]See, for example, "From Agitator to Enforcer: The Evolution of Breitbart," *The Washington Post*, February 17, 2017 (at at https://www.washingtonpost.com/video/politics/from-agitator-to-enforcer-the-evolution-of-breitbart/2017/02/19/9714b4ce-f3b8-11e6-9fb1-2d8f3fc9c0ed_video.html.



Note: The logarithm of the number of followers is on the vertical axis

Figure A2. Estimated Slant Scores for Robustness

Table A1. Estimated Slant Scores for 2017 Data

Twitter Handle	URL	Estimated Slant
ABC	http://ABCNews.com	-0.241
AboveTopSecret	http://www.abovetopsecret.com	0.916
abscbndotcom	http://www.ABS-CBN.com	0.048
ACLJ	http://www.ACLJ.org	3.232
ACLU	http://www.aclu.org	-1.861
AJEnglish	http://aljazeera.com	-0.823
AllenWest	http://www.allenbwest.com	3.229
AlterNet	http://www.AlterNet.org	-2.238
amprog	http://www.americanprogress.org	-2.267
AmericanThinker	http://www.americanthinker.com	3.384
AOL	http://www.aol.com	-0.317
AP	http://www.ap.org	-0.245
azcentral	http://www.azcentral.com	0.306
BBC	http://www.bbc.co.uk	-0.355
billmaher	http://www.billmaher.com/	-1.415
oreillyfactor	http://www.billoreilly.com	2.591
BlackAmericaWeb	http://www.blackamericaweb.com	-1.227
Blklivesmatter	http://www.blacklivesmatter.com	-1.971
Blogcritics	http://blogcritics.org/	-0.144
business	http://www.bloomberg.com	-0.158
BoingBoing	http://www.boingboing.net	-1.310
BostonGlobe	http://bostonglobe.com	-0.517
bostonherald	http://www.bostonherald.com	0.617
BreitbartNews	http://breitbart.com	2.978
businessinsider	http://businessinsider.com/	-0.083
BW	http://www.businessweek.com	-0.130
BuzzFeedNews	http://www.buzzfeed.com/news	-0.657
cspan	http://www.c-span.org	-0.294
CanadaDotCom	http://www.canada.com	-0.456
CatoInstitute	http://www.cato.org/	2.053
CBCNews	http://www.cbc.ca/news/	-0.313
CBSNews	http://CBSNews.com	-0.169
chicagotribune	http://chicagotribune.com	-0.446
chicksonright	http://chicksontheright.com	3.452
CTmagazine	http://christianitytoday.com	1.910
HoustonChron	http://www.chron.com	0.250
CNBC	http://cnbc.com	-0.057
CNET	http://www.cnet.com	-0.108
CNN	http://www.cnn.com	-0.273
CNNPolitics	http://cnn.com/politics	-0.648
cnsnews	http://facebook.com/cnsnewscom	3.137
NatCounterPunch	http://counterpunch.org	-1.575
crooksandliars	http://www.crooksandliars.com/	-2.409
csmonitor	http://www.CSMonitor.com	-0.372
CTVNews	http://www.ctvnews.ca	-0.079
DailyCaller	http://www.dailycaller.com	2.728
dailykos	http://www.dailykos.com	-2.176

Table A1. Estimated Slant Scores for 2017 Data (Continued)

Twitter Handle	URL	Estimated Slant
demunderground	http://www.democraticunderground.com	-2.227
DNC	http://democrats.org/	-1.990
DRUDGE_REPORT	http://www.DRUDGEREPORT.com	2.312
Earthfiles	http://www.earthfiles.com	0.390
TheEconomist	http://www.economist.com	-0.390
EFF	http://www.eff.org	-0.884
esquire	http://www.esquire.com	-0.891
MichaelSalla	http://exopolitics.org/	0.179
Daily_Express	http://www.express.co.uk	0.630
YahooFinance	http://finance.yahoo.com	0.138
firstdraftnews	http://firstdraftnews.com	-0.865
FiveThirtyEight	http://www.fivethirtyeight.com	-1.017
Forbes	http://forbes.com	-0.028
ForeignPolicy	http://www.foreignpolicy.com	-0.444
FOXLA	http://www.foxla.com	0.556
FoxNews	http://www.foxnews.com	1.157
CanoeNouvelles	http://fr.canoe.ca/	-0.239
FreeBeacon	http://FreeBeacon.com	2.960
Snowden	http://freedom.press	-0.355
FreeRepublicTXT	http://www.freerepublic.com	0.176
fpmag	http://frontpagemag.com	3.287
FT	http://www.ft.com/	-0.127
GallupNews	http://www.gallup.com	0.613
glennbeck	http://www.glennbeck.com	2.754
globeandmail	http://www.globeandmail.com	-0.336
GOP	http://gop.com	2.011
haaretzcom	http://www.haaretz.com	0.168
seanhannity	http://hannity.com	2.781
HarvardBiz	http://hbr.org	-0.494
HeraldTribune	http://www.heraldtribune.com	0.182
Heritage	http://heritage.org	2.705
hotairblog	http://hotair.com	3.352
HuffingtonPost	http://www.huffingtonpost.com	-0.730
blackvoices	http://www.huffingtonpost.com/black-voices/	-1.425
HuffPostPol	http://www.huffingtonpost.com/politics	-1.291
HumanEvents	http://www.HumanEvents.com	3.228
TheIJR	http://ijr.com	2.036
infowars	http://www.infowars.com	2.430
zittrain	http://www.jz.org	-0.930
latimes	http://latimes.com/	-0.454
lessig	http://lessig.org	-1.347
LifeNewsHQ	http://www.LifeNews.com	3.075
LifeSite	http://www.lifesitenews.com	3.060
MarketWatch	http://www.marketwatch.com/	0.119
mashable	http://mashable.com	-0.614
Mediaite	http://www.Mediaite.com	0.788

Table A1. Estimated Slant Scores for 2017 Data (Continued)

Twitter Handle	URL	Estimated Slant
mmfa	http://mediamatters.org/	-1.929
PolitiFact	http://membership.politifact.com	-0.963
michellemalkin	http://www.michellemalkin.com	2.732
MotherJones	http://www.motherjones.com	-1.906
MoveOn	http://MoveOn.org/	-2.365
MSNBC	http://msnbc.com	-0.803
nationaljournal	http://www.nationaljournal.com/	0.271
NRO	http://www.NationalReview.com	2.615
NatureNews	http://www.nature.com/news	-0.654
NBCNews	http://NBCNews.com	-0.210
NewRepublic	http://www.newrepublic.com	-1.316
MSNNews	http://news.msn.com	-0.944
YahooNews	http://news.yahoo.com	0.118
newsbusters	http://www.newsbusters.org/	3.185
Newser	http://www.newser.com	-0.243
newsmax	http://www.newsmax.com	2.834
newsobserver	http://www.newsobserver.com	-0.188
Newsweek	http://www.newsweek.com	-0.512
NewYorker	http://www.newyorker.com	-0.908
NPR	http://www.npr.org	-0.924
nypost	http://www.nypost.com	0.773
nytimes	http://www.nytimes.com/	-0.543
OccupyDemocrats	http://www.OccupyDemocrats.com	-2.962
OurFuture	http://www.ourfuture.org	-2.687
NewsHour	http://pbs.org/newshour	-0.959
FactTank	http://www.pewresearch.org/fact-tank	-0.612
phillydotcom	http://www.philly.com	-0.270
instapundit	http://pjmedia.com/instapundit/	3.248
politico	http://www.politico.com	-0.460
powerlineUS	http://www.powerlineblog.com	3.429
theprogressive	http://www.progressive.org	-2.662
ProPublica	http://www.propublica.org/	-1.467
theprospect	http://prospect.org/	-2.306
PsychToday	http://www.psychologytoday.com	-0.697
RawStory	http://www.rawstory.com	-1.856
RealClearNews	http://www.realclearpolitics.com	1.775
reddit	http://reddit.com	-0.281
RedState	http://www.redstate.com	3.027
Reuters	http://www.reuters.com	-0.175
RollingStone	http://www.rollingstone.com	-0.915
RSPolitics	http://www.rollingstone.com/politics	-2.026
rushlimbaugh	http://www.rushlimbaugh.com	3.406
Salon	http://www.Salon.com	-1.686
SELFmagazine	http://www.self.com	-0.268
SFGate	http://sfgate.com	-0.889
shadowproofcom	http://shadowproof.com/	-1.744

Table A1. Estimated Slant Scores for 2017 Data (Continued)

Twitter Handle	URL	Estimated Slant
Slate	http://www.slate.com/	-1.249
TPM	http://www.talkingpointsmemo.com/	-1.406
technorati	http://technorati.com	-0.461
amconmag	http://www.theamericanconservative.com	1.745
TheAtlantic	http://www.theatlantic.com	-1.190
theblaze	http://www.TheBlaze.com	2.883
thecrimson	http://thecrimson.com	-0.700
thedailybeast	http://thedailybeast.com	-0.668
TheDailyShow	http://thedailyshow.com	-1.363
gatewaypundit	http://www.thegatewaypundit.com/	3.360
guardian	http://www.theguardian.com	-0.528
thehill	http://www.thehill.com	-0.119
MattWalshBlog	http://themattwalshblog.com/	3.284
thenation	http://thenation.com	-1.938
theolympian	http://www.theolympian.com	-0.554
TPCblog	http://thepoliticalcarnival.net	-3.346
trscoop	http://therightscoop.com	3.469
thestate	http://www.thestate.com	0.452
TheWeek	http://www.TheWeek.com	-0.576
thinkprogress	http://www.thinkprogress.org	-2.112
ThisAmerLife	http://www.thisamericanlife.org	-1.610
TIME	http://www.time.com	-0.434
TODAYshow	http://TODAY.com	-0.274
Topix	http://topix.com	0.573
townhallcom	http://townhall.com/	3.262
TreeHugger	http://www.treehugger.com	-1.616
TwitchyTeam	http://www.twitchy.com	3.395
UPI	http://www.upi.com	0.201
USATODAY	http://www.usatoday.com	-0.016
usnews	http://www.usnews.com	0.107
VanityFair	http://www.vanityfair.com	-0.808
vicenews	http://www.vicenews.com	-0.866
villagevoice	http://www.villagevoice.com	-1.202
VOANews	http://voanews.com/	-0.001
voxdotcom	http://vox.com	-1.276
dcexaminer	http://www.washingtonexaminer.com	2.543
washingtonpost	http://washingtonpost.com	-0.435
WashTimes	http://www.washingtontimes.com	1.632
weaselzippers	http://www.weaselzippers.us	3.564
weeklystandard	http://www.weeklystandard.com	2.555
WestJournalism	http://www.westernjournalism.com	3.341
wikileaks	http://wikileaks.org	0.428
worldnewsdotcom	http://wn.com/	0.671
worldnetdaily	http://www.wnd.com	3.099
WSJ	http://wsj.com	-0.023
WSJopinion	http://wsj.com/opinion	2.256
YoungCons	http://youngcons.com/	3.602

Because of the extremely slow process of finding accounts that fit our criteria, our largest set of data contains only the slant of this single followee, rather than the entire incoming time line of a given follower to predict that follower's outgoing slant. This obviously introduces substantial error (including attenuation bias) that was not present in our initial analysis, in which we had the entire incoming time line for every account in our dataset. We have two reasons to believe that the error introduced by this method may not be as problematic as it sounds. First, the followees that we find by sampling on tweets (i.e., searching for tweets with the search API) are likely to be more representative of the time line than a followee picked at random; this is because sampling on tweets is more likely to find especially active tweeters, whose tweets make up a majority of a follower's incoming time lines. Second, for a subset of data, we also collected expanded incoming time lines and found generally consistent regression results.

Our full algorithm for collecting individual-level data is given below, but before detailing that, we begin by defining a subroutine—"PROCESS URLS"—used repeatedly throughout our data collection. For example, we frequently needed to determine the slant of the most recent 3,200 tweets sent by a given account. URLs in tweets are always shortened links; that is, the original full URL is converted into a shorter version by Twitter (using the t.co domain) and possibly other shortening services such as bit.ly. This means that the links in any tweet data we collected are totally opaque with respect to where they point.

> PROCESS URLS: Given a set of tweets containing URLS, we begin by discarding any that are more than 2 weeks old. We then follow each link until there are no further redirects and collect the URL at the final destination. We then strip away protocol and server names (http:// or https:// and www or other leading information) and match to the list of news domains we created earlier, discarding any tweets without links to a news source on our list.

Collection Algorithm

> We query the search API for tweets containing hyperlinks. We then PROCESS URLS (as above) for these tweets and call remaining tweets "seed tweets." For each seed tweet, we look up the account that sent the tweet, designating this account as "followee." If the followee has greater than 10,000 followers, we stop and move on to the next tweet without recording data. We query the REST API for the most recent 3,200 tweets sent by the followee and then PROCESS URLS and calculate the average slant of news tweets sent by the followee. We then query the API for the list of accounts that follow the followee. Until we find a follower that sent a news link, we loop over the following steps: {randomly select one follower, query the API for the most recent 3,200 tweets sent by that follower, PROCESS URLS} (if no followers tweeted a news link, stop and move on to the next seed tweet). Calculate the average slant of the follower. Store data and move on to the next seed tweet.

Further Robustness Checks

The above collection algorithm allows us to regress a follower's outgoing slant on the outgoing slant of an account that they follow. Two substantial concerns with this approach follow: (1) in 2017, the number of "bots" (accounts that are not run by a human, but rather by a piece of computer code) has reportedly increased on Twitter and (2) it is unclear how representative one followee account would be of the entire incoming time line (notwithstanding our note that because the followee is sampled on tweet activity, they are likely to be more representative than the average followee). To check the magnitude of any effects caused by these issues, we conducted further robustness checks on a subset of data.

Bot check: The Truthy team at Indiana University recently released an API to their "Bot or not" service (http://truthy.indiana.edu/botornot/). Essentially, this is a machine-learning based classifier of Twitter accounts that produces a score, ranging from zero to one, with one being highest certainty that the account in question is a bot. We took a subset of followers from our main collection algorithm and used the botornot API to provide a bot score. We then kept only those accounts that scored less than 0.4.

Expanded time line: With the subset of accounts unlikely to contain many bots, we then expanded the number of followees in the "incoming slant" calculation. We did not attempt to include all followees, because of feasibility constraints imposed by Twitter's rate limits and the requirement to PROCESS URLS. Instead, using insight from the main analysis of 2009 data that heavier tweeters tend to tweet less centrist material, we obtained time lines from only the top-20 heaviest tweeters from among each follower's followees. This also has the advantage that the more an account tweets, the more representative of the whole incoming slant they are, all else equal.

Results

Results from these robustness checks are presented in Table A2. Our main result, that on average outgoing slant is more moderate than incoming slant, stands with these current data. These new data are much less complete than our main analysis and necessarily introduce some sampling bias of unknown magnitude. However, they were collected at a time of greater maturity of the Twitter platform and a time of much greater apparent political polarization—during the first months of the Trump administration—than the data for the main analysis.

Because of the different approaches to data collection, a detailed interpretation of differences between parameter estimates from 2009 to 2017 is generally not warranted. However, one substantial difference is attributable to the sampling approach and we note it here. The parameter for the interaction between incoming slant and number of followers was large and significant in our main analysis, but insignificant here. This is almost certainly because we only collected data for accounts with fewer than 10,000 followers for this robustness check; thus, there is much less variation in this variable by construction.

Table A2. Relationship between Incoming and Outgoing Political Slant

	DV: Mean Slant of Sites in Outgoing Tweets	
	I	II‡
Mean slant, sites in incoming tweets†	0.586*** 0.010	0.751*** 0.032
ln(Count of outgoing tweets)	0.064*** 0.010	0.007 0.034
ln(#followers)	0.048*** 0.014	0.094 0.048
ln(#followers ÷ *ln*(#followees)	-0.027 0.014	-0.016 0.048
Incoming slant × *ln*(Count outgoing tweets)	0.154*** 0.010	0.093** 0.033
Incoming slant × *ln*(#followers)	0.009 0.014	0.038 0.048
Incoming slant × *ln*(#followers) ÷ *ln*(#followees)	0.036* 0.017	0.011 0.057
Intercept	-0.049*** 0.010	-0.009 0.033
# of Twitter Accounts	5966	445
Adjusted R^2	0.419	0.589

Notes: Standard errors are printed below parameter estimates.
***p < 0.001; **p < 0.01; *p < 0.05
Except for mean incoming slant and mean outgoing slant, all variables are centered and scaled to unit variance.
†For model I, incoming tweets originate from one sampled followee. For model II, incoming tweets originate from the top 20 most-frequent tweeters among that account's followees.
‡Model II includes a subset of accounts that have a low probability of being bots and have fuller incoming slant information.

References

Benkler, Y., Faris, R., Roberts, H., and Zuckerman, E. 2017. "Study: Breitbart-Led Right-Wing Media Ecosystem Altered Broader Media Agenda," *Columbia Journalism Review* (available at http://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php).

Gentzkow, M., and Shapiro, J. M. 2011. "Ideological Segregation Online and Offline," *The Quarterly Journal of Economics* (126:4), pp. 1799-1839.

Pew Research Center. 2014. "Ideological Placement of Each Source's Audience," Journalism & Media (available at http://www.journalism.org/2014/10/21/political-polarization-media-habits/pj_14-10-21_mediapolarization-08/).

Reichardt, J., and Bornholdt, S. 2006. "Statistical Mechanics of Community Detection," *Physical Review E*, (74:DOI: 016110).