Correction: 28 May 2015

www.sciencemag.org/cgi/content/full/science.aaa1160/DC1

Supplementary Material for

Exposure to ideologically diverse news and opinion on Facebook

Eytan Bakshy,* Solomon Messing, Lada Adamic

*Corresponding author. E-mail: ebakshy@gmail.com

Published 7 May 2015 on *Science* Express
DOI: 10.1126/science.aaa1160

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S6
Full Reference List

Correction: Several edits to text were made for clarification purposes in sections 1.1, 1.4, and 1.7 and in the Table S6 caption.

Supporting Materials for Exposure to Ideologically Diverse News and Opinion on Facebook

Eytan Bakshy* Solomon Messing
Lada Adamic
*Corresponding author. E-mail: eytan@fb.com

This PDF file includes:

Materials and Methods

Supplementary Text

Fig. S1-S10

Tables S1 to S7

Full Reference List

Other Supplementary Material for this manuscript includes the following:

The following code and data are archived in the Harvard Dataverse Network,

http://dx.doi.org/10.7910/DVN/LDJ7MS,

"Replication Data for: Exposure to Ideologically Diverse News and Opinion on Facebook".

R analysis code and aggregate data for deriving the main results (e.g., Table S5, S6)

Python code and dictionaries for training and testing the hard-soft news classifier

Aggregate summary statistics of the distribution of ideological homophily in networks

Aggregate summary statistics of the distribution of ideological alignment for hard content shared by the top 500 most shared websites.

1 Materials and Methods

1.1 Population

To construct our population, we consider active U.S. adults on Facebook who report their political affiliation. This includes U.S. Facebook users who are 18 or older, log in at least 4/7 days per week (i.e., 105/185 days during study period, July 7, 2014-January 7, 2014; this removes approximately 30% of users). We limit this population further to those who self-report their ideological affiliation, which comprises 25% of active U.S. adults as defined above. We also limit the population to those who have clicked on at least one link shared on Facebook that we classified as hard news/opinion over the course of the study. The final population includes 10.1 million users. See Table S1.

1.2 Political designations

All Facebook users can self-report their political affiliation; 9% of U.S. users over 18 do. We mapped the top 500 political designations on a five-point, -2 (Very Liberal) to +2 (Very Conservative) ideological scale; those with no response or with responses such as “other” or “I don’t care” were not included. 46% of those who entered their political affiliation on their profiles had a response that could be mapped to this scale. We validated a sample of these labels against a survey of 79 thousand U.S. users in which we asked for a 5-point very-liberal to very-conservative ideological affiliation; the Spearman rank correlation between the survey responses and our labels was 0.78.

1.3 Ideological homophily

Individuals vary with respect to how many friends they have of different affiliations. Figure S1 shows the proportion of ties to friends of different ideological affiliations for each user. Each panel indicates a focal user of a particular affiliation, and the distributions show the kernel density estimate of the percentage of ties to friends of different affiliations. Vertical lines indicate medians. Both conservatives and liberals exhibit substantial homophily in friend networks, while moderates maintain a similar distribution of ties to both liberal and conservative friends. See also Table S2.

1.4 Hard news and opinion data

The final dataset used in the analysis of the main text includes 226,310 hard news and opinion stories (which we refer to simply as "hard content" below). Because we wish to classify each article according to its ideological alignment, we only consider stories that had been shared by at least 20 U.S. users who self-report a mappable ideological affiliation. This set of links includes over 90% of all URLs labeled as hard content seen by individuals in our study. We describe the procedure used to classify hard content below.

1.4.1 Hard-soft classification

We build our hard-soft classifier using an approach often referred to in the Natural Language Processing literature as "bootstrapping" [28, 29, 30, 31] which entails using regular expressions to build a set of training labels (and should not be confused with bootstrapping in statistics).

We begin with URL content shared by at least 100 U.S. users. To extract features from the documents in question (text summaries of the articles sent to Facebook when a user shares content from an external website), we apply English stopwords; tokenize using unigrams, bigrams, and trigrams; and use tokens that have occurred in at least 2 and no more than half of all documents.

To construct positive training labels, we use stories from 81 of the most shared news sites on Facebook in 2012, among them *nytimes.com, foxnews.com, cnn.com*, and *latimes.com*, that contain indicators of explicit hard news/opinion topics in the URL (see `FitSVM.py` in replication materials for a complete list). These include the following strings: “politi”, “usnews”, “world”, “national”, “state”, “elect”, “vote”, “govern”, “campaign”, “war”, “polic”, “econ”, “unemploy”, “racis”, “energy”, “abortion”, “educa”, “healthcare”, “immigration.” We construct negative training cases by matching the URL to the following strings “sports”, “entertainment”, “arts”, “fashion”, “style”, “lifestyle”, “leisure”, “celeb”, “movie”, “music”, “gossip”, “food”, “travel”, “horoscope”, “weather”, “gadget.” Our training data consisted of 147,958 stories; of which 114,121 were labeled soft news and 33,837 were labeled hard news/opinion.

We then trained our classifier using a linear SVM with the standard L2 penalty and hinge loss using SciKitLearn. We classified 694,989 URLs as hard content and 6,929,907 as soft content in the complete set. Our classifier achieves ten-fold cross-validated accuracy of 97.1 percent. Limiting this set to URLs shared by at least 20 affiliated users yields 226,310 hard news and opinion stories, which we use in our analysis.

1.4.2 Measuring alignment

We measure the ideological *alignment* of these items, which constitutes a behavioral indicator of the extent to which partisan identifiers actively share content. We derive the alignment score A_u of an individual URL by averaging the political alignment of the set of people P_u who share the URL.

$$A_u = \frac{1}{|P_u|} \sum_{i \in P_u} a_i \tag{1}$$

As shown in Figure S2, the measure indicates substantial polarization among the content shared by individuals who provide an ideological affiliation, with the most frequently shared URLs coming from sources at the ends of the distribution of alignment. We note that while the alignment scores of URLs classified as hard content exhibit strong alignment with either the right or left, soft content generally does not show strong patterns of ideological separation (see

Figure S2a). We take quintiles of the measure and color the left- and rightmost quantiles blue and red, respectively.

Alignment scores of URLs, averaged over their respective websites, have substantial face validity. We note a few highly shared and well-known media sources in Table S3.

1.4.3 Validation

We validate this metric against other efforts to quantify ideology in media content. To produce these estimates, we matched domains from our study to those listed in three other studies: [32], [33], and [34].

We first compare our alignment measure to recent work by Budak et al. (2014) that utilizes crowd-sourced content analysis to quantify media bias in commonly visited online news sources. By employing humans to manually annotate articles for media bias, then using machine learning to infer ideology based on the vocabulary used in unseen news articles, Budak et al. generate measures of political slant. We matched all of the 15 sources in Budak et al. to domain-level alignment scores in our data and find the two scores to be highly correlated (Pearson's $\rho = 0.91$, 95% CI = $[0.750.97]$, see also Figure S3a).

Next, we examine the correlation between our measure and that of Groseclose and Milyo (2005), which leverages Americans for Democratic Action (ADA) ratings of the ideology of organizations such as Washington think tanks that news organizations cite as news sources [33]. We match 17 of the 20 sources in Groseclose and Milyo (matching hard news broadcasts to the same organization's website when the news source constituted a television show) to domains in our data; the correlation is -0.47 (Pearson's ρ, 95% CI = $[-0.78, 0.01]$), see also Figure S3c).

Finally, we validate our measure of alignment against Gentzkow and Shapiro's (2010) measure of media *slant*, which models the similarity between the language used by Democrats and Republicans in congressional proceedings and local newspapers in each congressional district

[34]. Of the 435 local papers scored by Gentzkow and Shapiro, we consider the 20 sources most shared on Facebook (this includes 4.6 percent of hard news/opinion shares from U.S. Facebook users; the remaining domains comprise just one percent). The correlation between alignment and Gentzkow and Shapiro's slant is 0.56 (Pearson's ρ, 95% CI = $[0.15, 0.80]$), see also Figure S3b).

1.5 Quantifying cross-cutting content

To measure whether content is cross-cutting, we then take the quintiles of the URL alignment score create five alignment categories: content shared by audiences that are on balance primarily liberal (-2), somewhat liberal (-1), bipartisan (0), somewhat conservative (1), and primarily conservative (2) (see also Figure S2). Finally, we show the proportion of stories published on websites of interest broken down by each alignment categories (Table S4).

1.6 Alignment and unaffiliated sharers

Our measure of content alignment uses data from individuals who self-report their political affiliation. It is important to note that many link shares are from users who do not report their political affiliation (Figure S4). Furthermore, both moderates and those who do not self-report their affiliation tend to share more liberally aligned content than conservative. Because liberals are more connected on Facebook (see Table S1) and because more individuals across the graph share liberal content, there is a higher likelihood that a given individual's friends have shared liberal content.

1.7 Relative risk in exposure probabilities

The main text demonstrates that individuals have the potential to be and are exposed to ideologically cross-cutting content. Although we do not identify the causal effects of a URL being cross-cutting on whether a user will see that URL in the New Feed, or select that URL if presented in the News Feed, it is still informative to look at the relative differences in the probability of exposure for ideologically consistent versus cross-cutting content. We summarize these relative differences in terms of risk ratios, e.g.,

$$1 - \frac{\text{Pr(click }|\text{exposed, cross-cutting)}}{\text{Pr(click }|\text{exposed, not cross-cutting)}}.$$

Such expressions can also be written in terms of the proportion of URLs that are cross-cutting at each stage in the consumption process. For example, the above equation is mathematically equivalent to $\frac{\pi_e(1-\pi_n)}{\pi_n(1-\pi_e)} - 1$, where π_e and π_n are the proportion of actual News Feed exposures and potential exposures (articles shared by friends), respectively, that are ideologically cross-cutting. We present these raw proportions in Table S5.

Table S6 presents these quantities at each stage: the relative likelihood of friends sharing cross-cutting content compared to the probability of sharing ideologically consistent content (given it was shared on Facebook at all), the relative likelihood an individual encounters cross-cutting content compared to consistent content in the News Feed given that friends' shared it, and the relatively likelihood an individual clicks on cross-cutting content compared to consistent content given exposure in the News Feed. In addition, we include a position-adjusted change for the selection probabilities (described in S1.8, below).

1.8 Position effects

In this section, we describe how the order in which the News Feed displays items could affect what information individuals are exposed to and select. In particular, we show the click rate on a link is negatively correlated with its position in the News Feed. Ignoring this variation could lead one to attribute differences in individual choice to differences induced by the feed ranking algorithm via ordering. We derive an adjusted estimate of the difference in the probability of selecting content, and show the sensitivity of the ratio reported in the main text to potential position effects.

If the News Feed ranking algorithm differentially placed ideologically cross-cutting content in lower positions (or higher positions), and being placed in a different position had a causal effect on click rates, then part of the observed difference in click-through rates for ideologically similar vs cross-cutting content could be explained by position effects alone.

We investigate this possibility by examining how exposure to ideologically diverse content and click through rates vary by position. Exposures to friends' link sharing behaviors are logged when viewers load the Facebook News Feed and the story renders in the visible portion of their Web browser or mobile device (i.e., a "validated viewport view").

Let $y_{ij} = 1$ when an individual i selects (clicks) a story j, and 0 otherwise. Let z_{ij} indicate whether a story appears on the individual's screen for more than 250 milliseconds, such that $z_{ij} = 1$ when i has a viewport view for story j, and 0 otherwise. Finally, we define $C(i, j)$ to be 1 when j is ideologically cross-cutting for user i, and 0 otherwise. We denote the total number of items seen by i, conditional on whether or not the content is crosscutting (c) as $n_i(c) = \sum_{j=1}^{M} z_{ij} C(i, j)$.

We then define $\bar{y}(c)$ to be the average probability that an individual clicks on an article, conditional on whether or not it is cross-cutting:

$$\bar{y}(c) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{n_i(c)} \sum_{j=1}^{M} y_{ij} z_{ij} C(i, j),$$

where N is the total number of individuals and M is the total number of distinct URLs that viewers could potentially be exposed to. Let the difference in probability of clicking on a link for ideologically similar and cross-cutting content be $\delta = \bar{y}(0) - \bar{y}(1)$.

Most users visit Facebook multiple times per day, so the same story may appear in different positions at different points throughout the day. We map each viewer-URL pair to the highest (i.e., minimum) position that a URL-share story rendered in each viewers' News Feed. We then examine the relationship between this position, ideological diversity, and click-through rates in Figure S5.

From Figure S5a, one can see that there is a strong correlation between position and click-through rates, and that furthermore, a slightly lower proportion of cross-cutting stories appear in positions toward the top of users' News Feeds for liberals, while conservatives encounter a slightly higher proportion of cross-cutting stories toward the top of their feeds (Figure S5b). Some positions—particularly the second position of the News Feed—are often allocated to sponsored content, which may include links to articles shared by friends which are associated with websites associated with a particular advertiser. Since we aim to characterize interactions with all hard content shared by friends, such links are included in our analyses. These links appear to be more ideologically consistent with the viewers; however further investigation is beyond the scope of this work. The decreasing relationship between click-through rate and position, combined with differences in ideological alignment that vary by position, could obscure the relative contribution of algorithmic ranking and selective exposure to decreased levels of ideologically cross-cutting news consumption.

One approach to adjust for the potential imbalance in story position is to stratify the difference in selection probabilities for ideologically similar vs cross-cutting content based on position. Using n_{ip} to denote the total number of stories for user i that appear in position p, $n_{i\cdot}$ to denote the total number of stories seen by a user, and $n_{\cdot\cdot}$ as the total number of stories seen by all users, we express the difference in click rates conditional on position as,

$$\tilde{\delta}_p = \bar{y}(0) - \bar{y}(1) \mid P = p,$$

And then write the stratified estimator of the δ as:

$$\tilde{\delta} = \frac{1}{n_{\cdot\cdot}} \sum_{p=1}^{\mathrm{P}} n_{\cdot p} \tilde{\delta}_p$$

As stated above, we provide these stratified estimates in Table **??**.

2 Supplementary Text

2.1 Facebook usage and exposure to cross-cutting content

Because the Facebook News Feed orders content shared by friends, the potential effects of ranking on exposure to cross-cutting content might be lower for highly active users who view many stories because these individuals scroll further down and see more of their friends' content. We investigate this possibility by examining the raw proportion of cross-cutting content that individuals could have potentially been exposed to, based on what their friend networks share, what they are exposed to via the Facebook News Feed, and what they select as a function of the number of stories viewed (Figure S6). We transform the number of stories viewed into deciles, using the distribution of stories viewed over all users, for ease of interpretation.

For conservatives, there is very little change in the diversity of content when moving from potential to exposed at each activity level. However, despite the lower proportion of cross-cutting content active conservatives encounter in News Feed, they select it at similar rates. It is possible that more active conservatives may be more open to consuming content from the other side, being younger and more female (both of which are associated with higher levels of openness [35]). There are other possible causes for this pattern, but a full analysis is beyond the scope of this work. Liberals, on the other hand, see less content from the other side, without much consistent variation related to activity.

2.2 Proportion of Individuals Exposed to Cross-Cutting Content

We also present the fraction of individuals in our sample who encounter at least one cross-cutting and aligned item at each stage, among those whose friends have shared said content (Figure S7). This provides an indication of the effects of algorithmic ranking and selectivity among those who are on the margins of being exposed to no cross-cutting content at all and allows comparisons between each stage with a common denominator (the total number of individuals in the sample). The figure shows that among those at the margin, selection choices play a greater role in determining whether individuals encounter ideologically cross-cutting content, compared to algorithmic ranking.

2.3 Sample Data

2.4 Classifications of political and nonpolitical URLs

A random sample of the text utilized by the hard-soft URL classifier. 20 URLs classified as hard-, and 20 classified as soft content, drawn from completely out of sample cases (e.g., URLs that were not in the training or test sets), are provided. When a user shares a URL from a news website, blog, or other site, these text summaries are sent to Facebook from the external website. They have been truncated and stripped of some types of punctuation.

2.4.1 Hard news and opinion content

- The government keeps backups of every federal record ever But it won't turn over Lois Lerner's missing emails ...
- On Sept 25th Governor Jerry Brown signed into law SB 1135 Prison Anti Sterilization bill authored by Senator Hannah Beth Jackson, sponsored by legal and human rights organization Justice Now and included bi partisan co authorship The bill went before the Governor after passing unanimously out ...

- COURT HALTS EXECUTION OF MENTALLY ILL TEXAS INMATE USA TODAY 1 32 P M WEDNESDAY, DECEMBER 3RD, 2014 AUSTIN A federal appeals court in New Orleans on Wednesday halted the execution of Texas killer Scott Panetti, whose case has spark...

- Foreigners love oil, gold, diamonds, and cheap labor of Islamic world They like the quarrels of the Middle East Believe me, they don t like us ...

- So far, the debate over the proposed Islamic center near Ground Zero has unfolded along predictable lines, with the man at the center of the project, Imam Feisal Abdul Rauf, drawing attacks from the right painting him as a terrorist sympathizer with ties to Hamas and the Muslim Brotherhood ...

- A Philadelphia police officer was caught on video cursing and threatening a teenager The video was posted on Facebook October 17, showing the unidentified officer following a teen boy as he walked home with ...

- The attack sent terrorized villagers fleeing into the bush in search of safety ...

- Philadelphia Police are looking for a woman who was seen on surveillance video stealing Halloween decorations from a home in South Philadelphia ...

- Malaysian Airlines fight MH17 carrying 295 passenger shot down by missile in Ukraine's east, Interfax reports ...

- As Congress and the White House mull a federal response to the surge of illegal immigration on the southwest border, here s a list of four Democratic governors who support letting illegal i...

- Pennsylvania Governor Tom Corbett on Tuesday signed into law a prisoner gag order that rights groups say is an affront to the First Amendment and a denial of all citizens' right to understand "an area of U S life physically removed from public scrutiny " More than 50 House Republicans have signed on to the strategy ...

- Australian Prime Minister Tony Abbott says authorities don t know the hostage taker s motivation yet The situation at the Lindt Chocolat Cafe has been going on for hours ...

- Aboard the papal plane, Nov 26, 2014 12 08 am (CNA) Pope Francis has said he aims to express the social doctrine of the Church, not the views of partisan political philosophies, suggesting it is reductionistic to say otherwise ...

- On a trip to Afghanistan during President Barack Obama's first term, Defense Secretary Robert Gates was stunned to find a telephone line at the military's special operations headquarters that linked directly back to a top White House national security official ...

- Dinesh D'Souza, a reliable producer of worthless garbage opinions, has another one The protesters in Ferguson, Missouri aren't so different from ISIS, the ruthlessly violent terrorist group currently wreaking havoc in Iraq ...

- United States will provide 47 million (35 million euros) in humanitarian aid to help Palestinians hit by Israel's campaign in the Gaza Strip, Secretary of State John Kerry pledged Monday ...

- Sen John McCain, the Republican nominee for president in 2008, joked on the Colbert Report that he might once again seek the nation s highest office The Comedy Central program s host, Stephen Colbert, asked the Arizona senator whether he would be interested in succeeding Chuck Hagel as secre...

- it is very likely that sometime next year, the Supreme Court will take up yet another major Texas redistricting case In 1991, the Democrats redrew the state s congressional map to create what the Almanac of American Politics called the shrewdest gerrymander of the 1990s with incre...

- A Florida woman claims she was beaten and choked by her husband of 4 years after a heated argument about not having enough fried chicken leftovers, according to an arrest affidavit obtained by The Smoking Gun ...

2.4.2 Soft content

- "The Power" un hit pop elettronico dal tedesco gruppo musicale Snap! dal loro album potenza mondiale It was released in January 1990 and reached number o ...

- Downtown Detroit is home to Quicken Loans, a company that sells mortgages The founder of Quicken Loans, Dan Gilbert is a one ...

- Garth Brooks LIVE in Missouri Posted by Clear99 on October 15, 2014 Experience the electrifying return of one of country music s most influential icons! Garth Brooks is returning to Missouri his first stop marks his first St Louis performance in over 18 years! He ll do two shows at the Scot...

- Official video of Blind Melon performing Tones of Home from the album Blind Melon Buy It Here ...

- Written by Bill Payne and Richie Hayward From the classic 1977 live album "Waiting For Columbus" Produced by Lowell George ...

- Special savings on Cyber Monday only ...

- 100+ positions available Licensed Insurance Agents, Sales Reps, and Customer Service Reps ...
- By now, there are very few Americans who haven t heard of the ...
- A quick clip from our new single “It’s A Revolution” The single drops with a new music video on September 15! Stay Tuned!
- Purchase Fam Jam (Fe Sum Immigrins) on iTunes ...
- Citando este pasaje del Antiguo Testamento, Jes s se alaba la diferencia entre la conducta externa y la vida interior del ser humano A qu se refer a el Se or? A la creencia de los fariseos de que cumpliendo las obras externas de la Ley de Mois s complac an a Dios, tales como lavarse las manos...
- In this post, we bring to you some of the most inspirational pictures quotes on Saying Images These quotes are about life, love, happiness & motivation We believe that you ll find some inspiration through this post & pictures quotes & they can change your life positively Inspirational Life qu...
- Janice and James Raffle and their three sons Angus, Barney and Joshua rushed to historic landmark when they heard the American president was in the area ...
- NEW YORK, United States For many, the market for wearables often called the next major technology battleground refers to gadgets worn on ...
- Fifth track of ”IN RAINBOWS”, disc 2, by Radiohead The band had worked on In Rainbows for more than two years, beginning in early 2005 In between recording ...
- Sandude Brewing Co of Turlock is trying to make a name for itself in the craft beer industry ...

- Protesters are expected to gather in downtown Greenville Sunday afternoon to stage a Die In along Main Street ...
- Help us reach 1,000,000 signatures today, telling LEGO to ditch Shell and their dirty Arctic oil!
- Join us for the expert webinar hosted by Zo Kessler on Wednesday, July 16, 2014 at 1 PM Eastern Time ...
- We live in a culture that says that the only way to get what you want is to DO DO DO and then DO some more Push Struggle Make it happen Put your nose to the grindstone Work your ass off Try And never ever give up I once believed this was true Surrender was ...

2.5 Ego networks

The figures in the main text are chosen as illustrative examples of liberal, moderate, and conservative networks; the proportion of links to liberals and conservatives fall within the interquartile ranges for each affiliation, shown in the main text. To illustrate the variety of ego network compositions in which Facebook users find themselves, we show separate random samples for conservatives, moderates, and liberals (Figures S8 - S10).

3 Figures and Tables

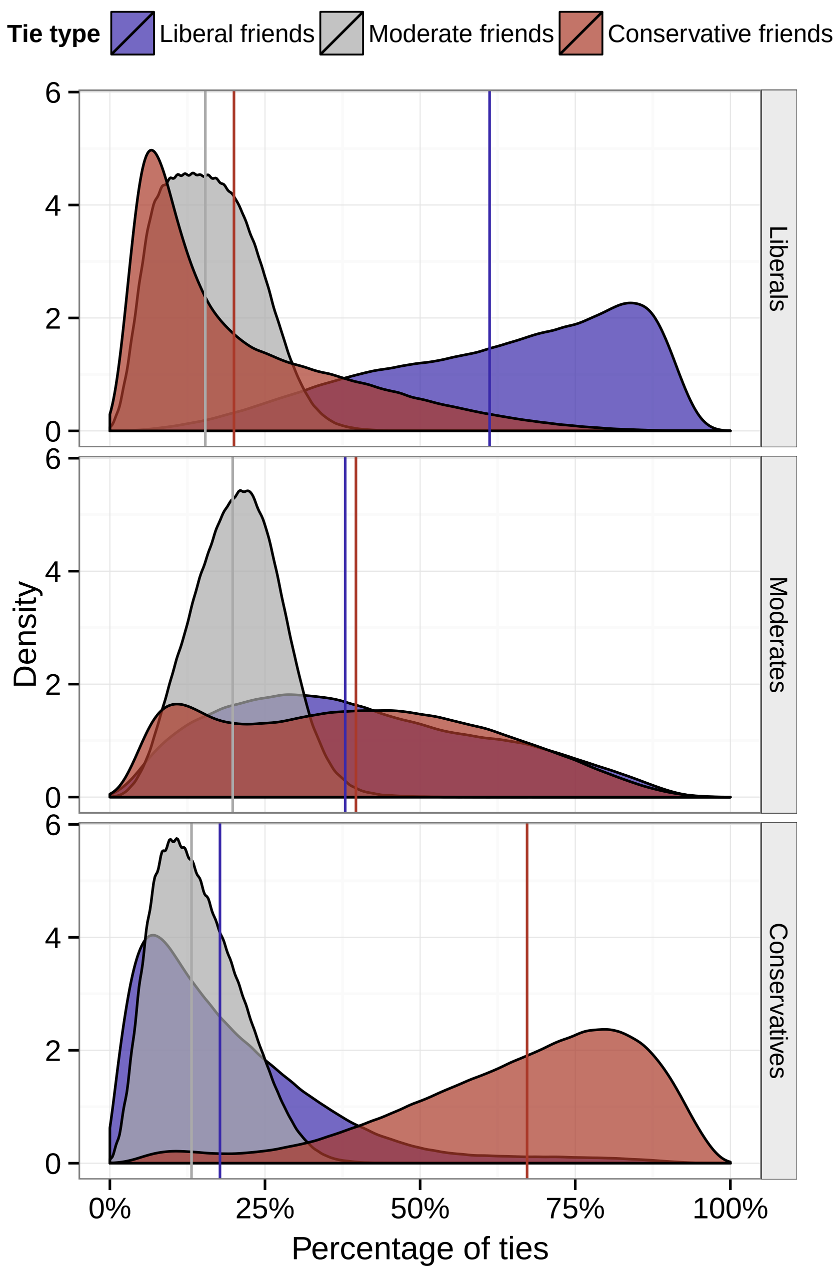
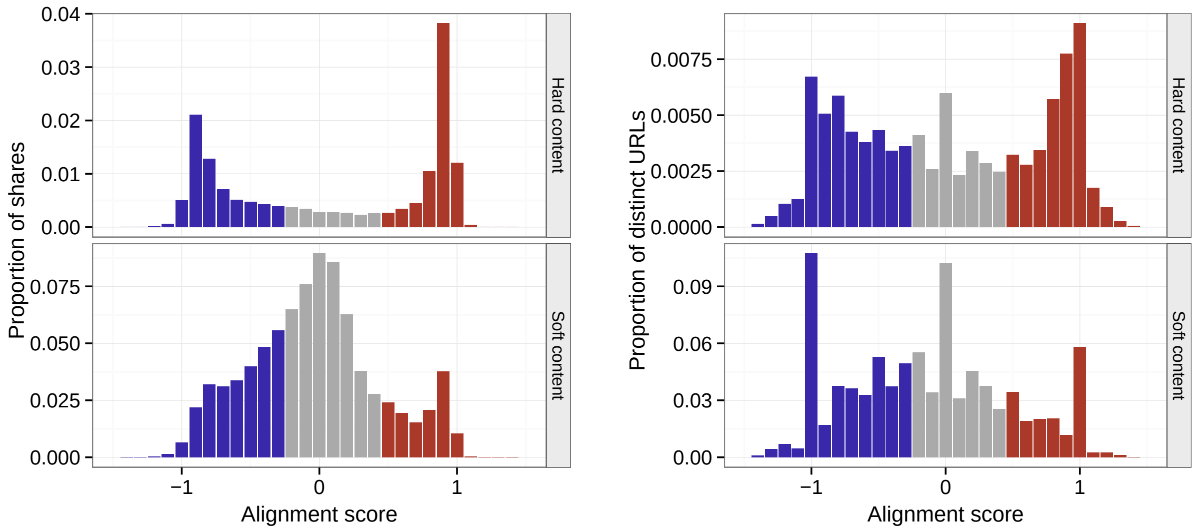
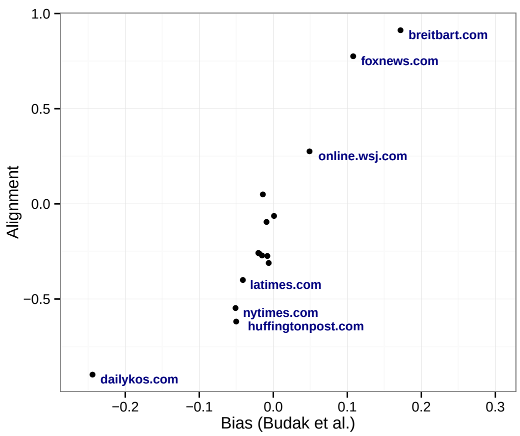


Figure S1: Kernel density plot of ties to friends of different affiliations, for liberals, moderates, and conservatives. Vertical lines indicate medians.

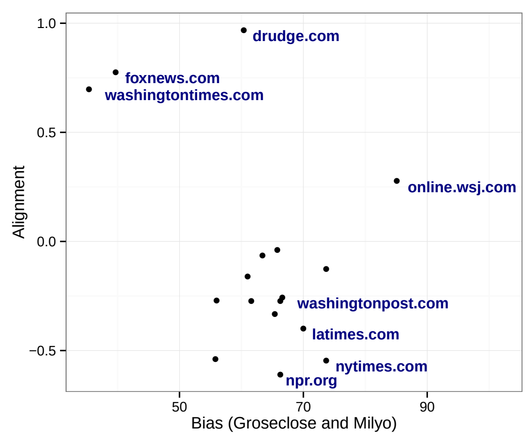


(a) (b)

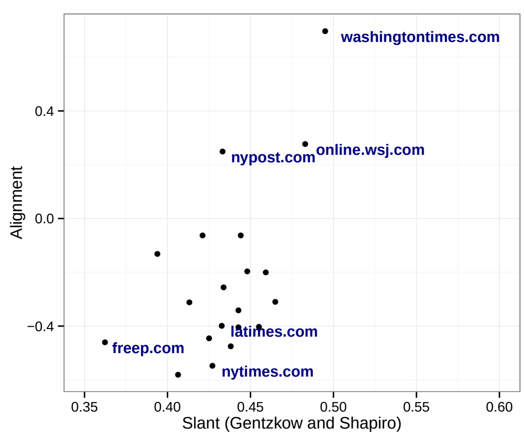
Figure S2: Distribution of alignment scores for hard news and opinion compared to soft content, (a) weighted by the total number of shares (b) weighted by the number of distinct URLs.



(a)

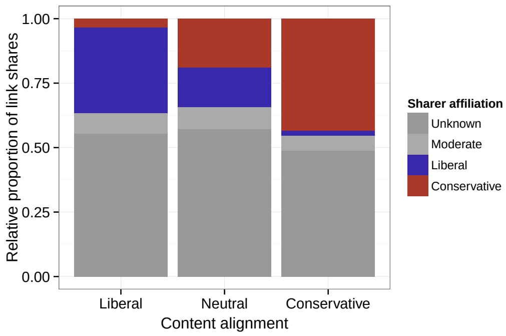


(b)

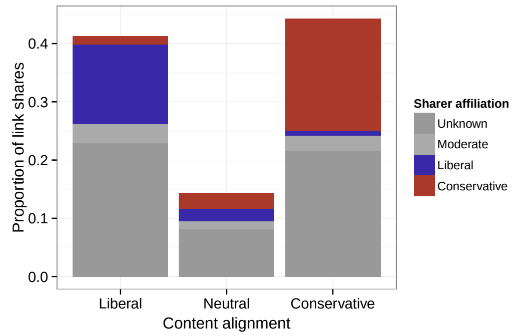


(c)

Figure S3: Average alignment by domain compared to (a) Budak et al.'s measure of media bias based on crowd-sourced annotations of partisan leanings; (b) compared to Groseclose and Milyo's measure of media bias, based on Americans for Democratic Action (ADA) ratings of the ideology of organizations that news organizations cite as sources; and (c) compared to Gentzkow and Shapiro's measure of media slant based on similarity to congressional records for 20 most shared domains.

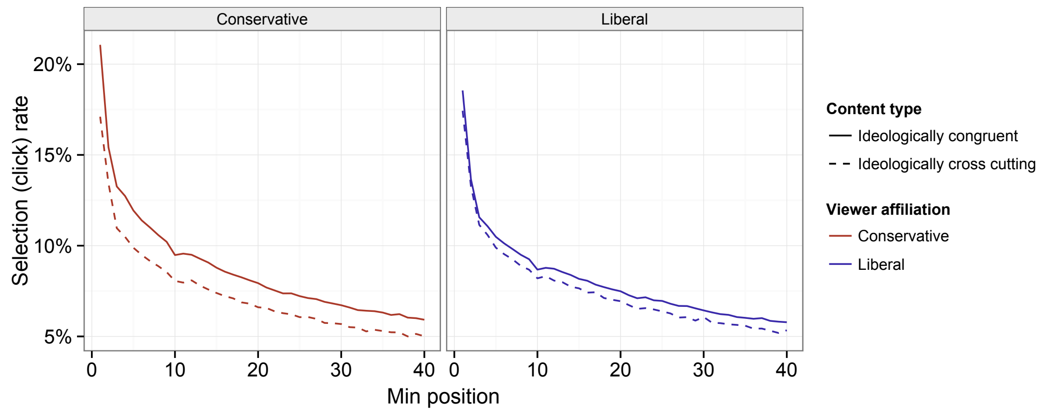


(a)

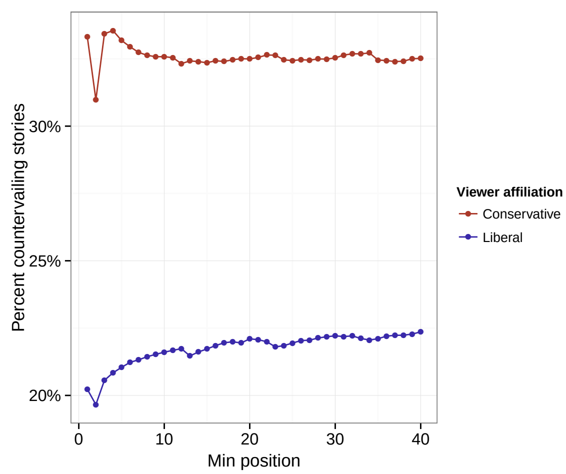


(b)

Figure S4: Ideological affiliations of sharers of liberal, neutral, and conservative hard content, including those who do not self-report their affiliation, as (a) the relative proportion of shares, from each alignment category (b) total number of link shares, as a proportion of all link shares



(a)



(b)

Figure S5: Relationship between story position and (a) click rate for ideologically congruent and cross-cutting content (b) percent of cross-cutting content shown in News Feed, for liberals and conservatives. Note that the relationship between click-through rate and position is both caused by relevance (including selective exposure) and individuals' tendencies to engage with content that is positioned toward the top of the News Feed.

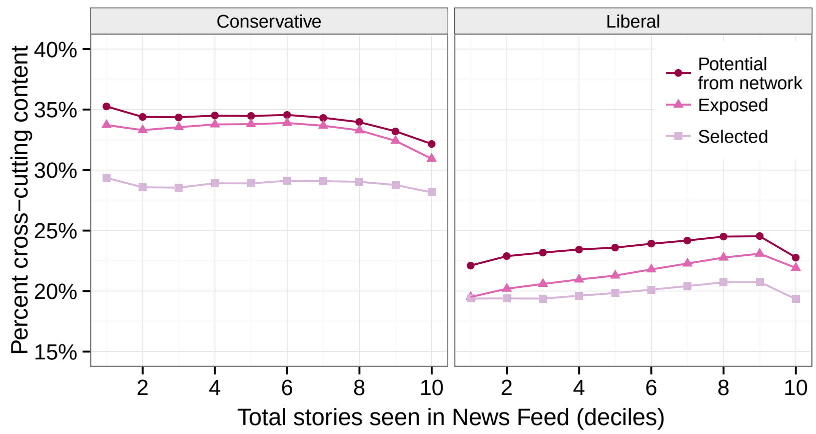


Figure S6: Proportion of ideologically cross-cutting content individuals could potentially be exposed to via their network, are exposed to via Facebook News Feed, and select (click), conditional on the decile-transformed total number of stories they are exposed to on Facebook.

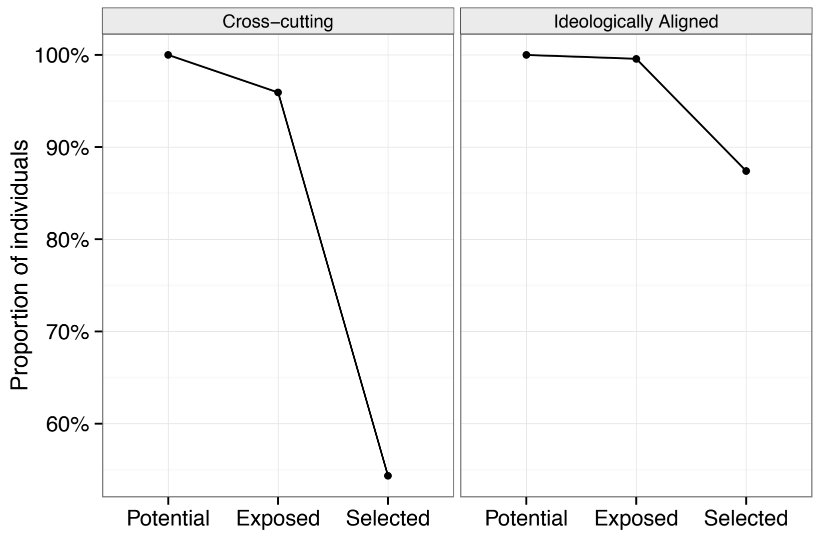


Figure S7: Proportion of individuals with at least one cross-cutting story (1) shared by friends (potential), (2) actually appearing in peoples' News Feeds (exposed), (3) clicked on (selected). The proportion of individuals with at least one ideologically aligned item is provided for comparison. Among individuals at the margins, choices about what to consume are more important in determining exposure to cross-cutting content than algorithmic ranking.

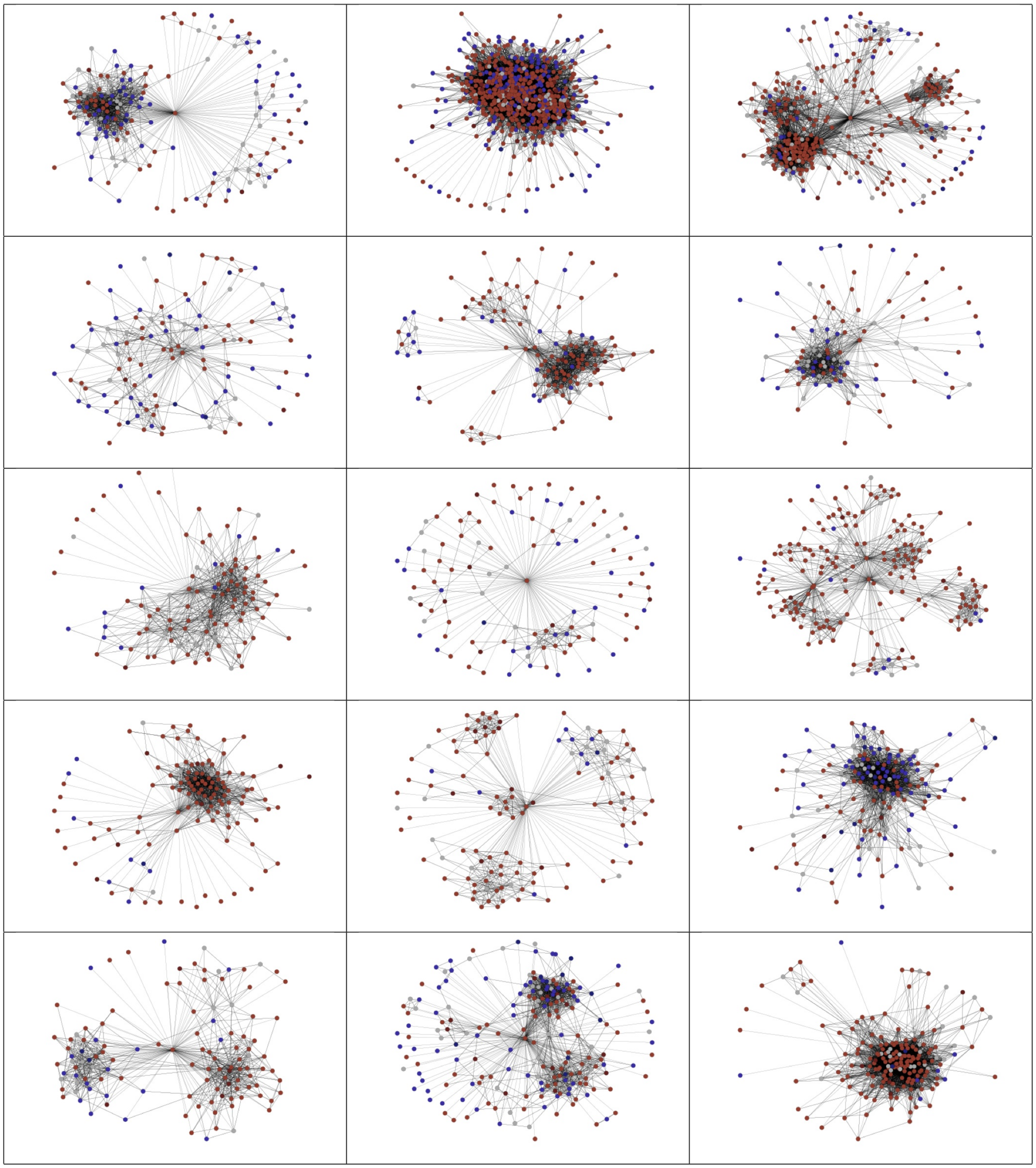


Figure S8: Ego networks of a random sample of conservatives, each having at least 100 friends with a declared political affiliation that was either conservative (red), moderate (gray) or liberal (blue). Only friends with political affiliation are shown.

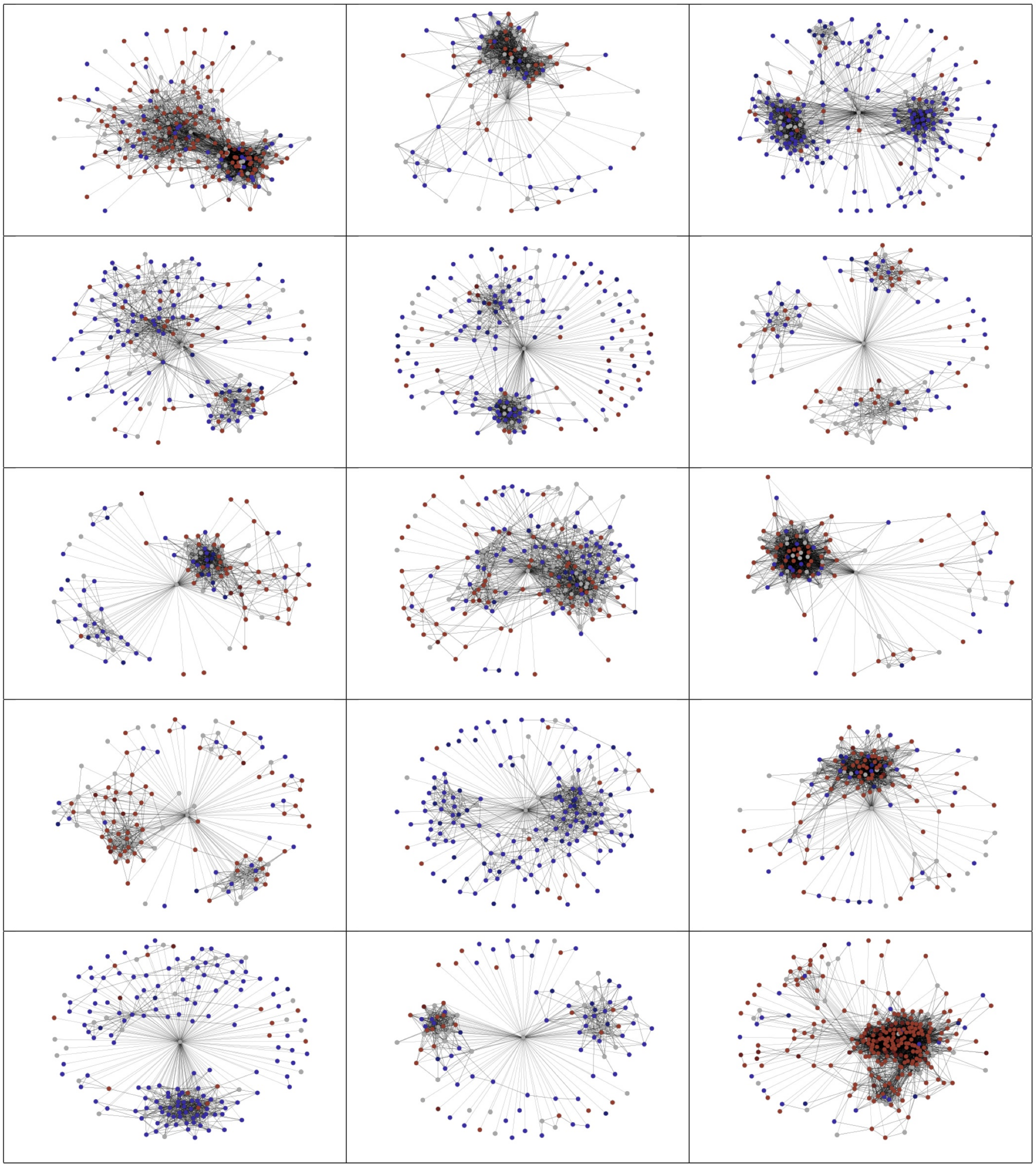


Figure S9: Ego networks of a random sample of moderates, each having at least 100 friends with a declared political affiliation that was either conservative (red), moderate (gray) or liberal (blue). Only friends with political affiliation are shown.

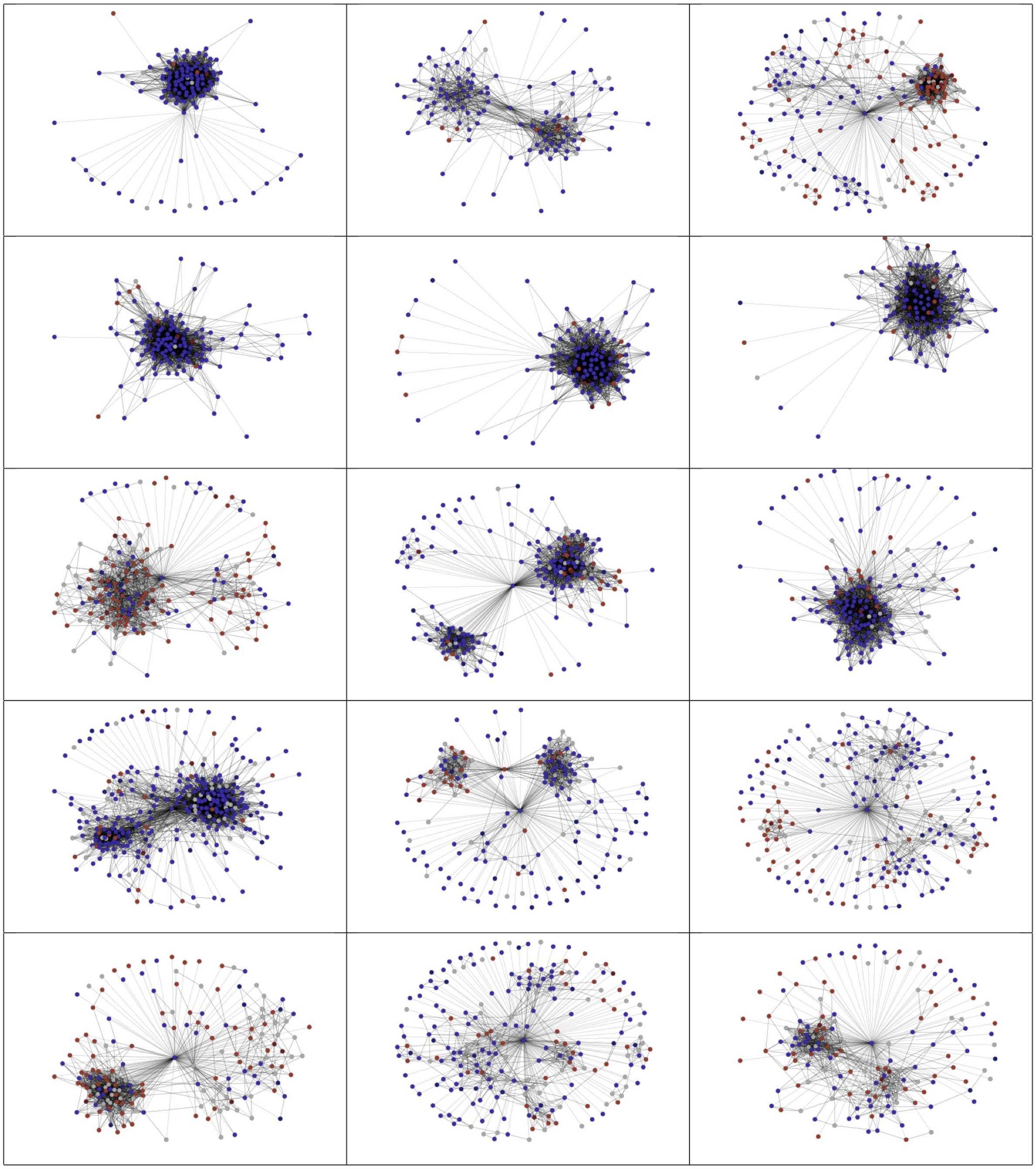


Figure S10: Ego networks of a random sample of liberals, each having at least 100 friends with a declared political affiliation that was either conservative (red), moderate (gray) or liberal (blue). Only friends with political affiliation are shown.

Variable	Viewer affiliation	Mean	25th perc.	Median	75th perc.
Age	Liberal	36.21	26	32	44
	Moderate	36.42	26	31	44
	Conservative	39.13	26	36	50
Female	Liberal	0.61	0	1	1
	Moderate	0.48	0	0	1
	Conservative	0.53	0	1	1
Login days	Liberal	171.82	167	182	185
	Moderate	172.64	169	182	185
	Conservative	172.83	169	182	185
Num. friends	Liberal	551.00	198	369	683
	Moderate	487.45	192	355	620
	Conservative	475.46	194	350	615

Table S1: Summary statistics for the population in our study. N = 4,063,793 (liberal), 1,602,164 (moderate), 4,469,394 (conservative).

User affiliation	Friend affiliation	Mean	5%	25%	50%	75%	95%
Liberal	Liberal	0.60	0.24	0.44	0.61	0.77	0.90
	Moderate	0.16	0.03	0.10	0.15	0.22	0.31
	Conservative	0.24	0.02	0.09	0.20	0.36	0.60
Moderate	Liberal	0.40	0.10	0.24	0.38	0.54	0.76
	Moderate	0.20	0.06	0.14	0.20	0.26	0.35
	Conservative	0.40	0.07	0.23	0.40	0.56	0.76
Conservatives	Liberal	0.21	0.02	0.09	0.18	0.30	0.52
	Moderate	0.14	0.02	0.08	0.13	0.19	0.29
	Conservative	0.65	0.28	0.52	0.67	0.80	0.91

Table S2: Summary statistics for the distribution of the proportion of ties to friends of different affiliations, for liberals, moderates, and conservatives.

Domain	Avg. alignment
www.dailykos.com	-0.90
www.huffingtonpost.com	-0.62
www.nytimes.com	-0.55
www.cnn.com	-0.27
www.washingtonpost.com	-0.26
www.foxnews.com	0.78
www.theblaze.com	0.89
www.tpnn.com	0.93

Table S3: Domain-level alignment for a sample of well-known media sources. Site alignment scores are obtained by averaging the alignment of URLs from a particular domain.

Domain	-2	-1	0	1	2
www.dailykos.com	0.967	0.029	0.000	0.002	0.000
www.huffingtonpost.com	0.361	0.330	0.291	0.014	0.001
www.nytimes.com	0.428	0.368	0.125	0.074	0.002
www.cnn.com	0.037	0.513	0.394	0.045	0.009
www.washingtonpost.com	0.208	0.361	0.213	0.159	0.056
www.foxnews.com	0.001	0.002	0.008	0.532	0.454
www.theblaze.com	0.000	0.000	0.004	0.309	0.686
www.tpnn.com	0.000	0.000	0.000	0.019	0.980

Table S4: Proportion of links from popular news outlets that are shared by primarily liberal (-2), somewhat liberal (-1), bipartisan (0), somewhat conservative (1), and primarily conservative (2) audiences.

Viewer affiliation	π_r	π_n	π_e	π_s
Liberal	0.454	0.237	0.222	0.211
Conservative	0.403	0.347	0.337	0.296

Table S5: Proportion of content that is ideologically cross-cutting for content that is shared by random others (π_r), within individuals' networks (π_n), was displayed in the News Feed (π_e), and got clicked on (π_s).

Viewer affiliation	Random $\rightarrow$ Potential	Potential $\rightarrow$ Exposed	Exposed $\rightarrow$ Selected	Exposed $\rightarrow$ Selected*
Liberal	-0.626	-0.080	-0.063	-0.065*
Conservative	-0.212	-0.046	-0.172	-0.165*

Table S6: Relative risk in probability of encountering cross-cutting versus consistent content at each transition (minus 1, see S1.7 and S1.8). * indicates position-adjusted estimate.

References and Notes

1. K. Olmstead, A. Mitchell, T. Rosenstiel, *Navigating news online*. Pew Research Center. Online at: http://www. journalism. org/analysis_report/navigating_news_online (2011).

2. W. L. Bennett, S. Iyengar, A New Era of Minimal Effects? The Changing Foundations of Political Communication. *J. Commun.* **58**, 707–731 (2008). doi:10.1111/j.1460-2466.2008.00410.x

3. E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (Penguin Press, London, 2011).

4. S. Messing, S. J. Westwood, *Communic. Res.* (2012).

5. E. Bakshy, I. Rosenn, C. Marlow, L. Adamic, *Proceedings of the 21st international conference on World Wide Web Pages* **1201.4145** (2012).

6. C. R. Sunstein, *Republic.com 2.0* (Princeton University Press, 2007).

7. N. Stroud, Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure. *Polit. Behav.* **30**, 341–366 (2008). doi:10.1007/s11109-007-9050-9

8. S. Kull, C. Ramsay, E. Lewis, Misperceptions, the Media, and the Iraq War. *Polit. Sci. Q.* **118**, 569–598 (2003). doi:10.1002/j.1538-165X.2003.tb00406.x

9. S. Flaxman, S. Goel, J. M. Rao, Ideological segregation and the effects of social media on news consumption, *SSRN Scholarly Paper ID 2363701*, Social Science Research Network, Rochester, NY (2013).

10. T. Groeling, Media Bias by the Numbers: Challenges and Opportunities in the Empirical Study of Partisan News. *Annu. Rev. Polit. Sci.* **16**, 129–151 (2013). doi:10.1146/annurev-polisci-040811-115123

11. M. Gentzkow, J. M. Shapiro, Ideological Segregation Online and Offline. *Q. J. Econ.* **126**, 1799–1839 (2011). doi:10.1093/qje/qjr044

12. M. J. LaCour, A balanced information diet, not echo chambers: Evidence from a direct measure of media exposure, *SSRN Scholarly Paper ID 2303138*, Social Science Research Network, Rochester, NY (2013).

13. E. Lawrence, J. Sides, H. Farrell, Self-Segregation or Deliberation? Blog Readership, Participation, and Polarization in American Politics. *Perspect. Polit.* **8**, 141 (2010). doi:10.1017/S1537592709992714

14. D. O. Sears, J. L. Freedman, Selective Exposure to Information: A Critical Review. *Public Opin. Q.* **31**, 194 (1967). doi:10.1086/267513

15. N. A. Valentino, A. J. Banks, V. L. Hutchings, A. K. Davis, Selective Exposure in the Internet Age: The Interaction between Anxiety and Information Utility. *Polit. Psychol.* **30**, 591–613 (2009). doi:10.1111/j.1467-9221.2009.00716.x

16. L. A. Adamic, N. Glance, *Proceedings of the 3rd International Workshop on Link Discovery* (ACM, New York, 2005), pp. 36–43.

17. S. Iyengar, K. S. Hahn, Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *J. Commun.* **59**, 19–39 (2009). doi:10.1111/j.1460-2466.2008.01402.x

18. M. Duggan, A. Smith, Social media update 2013. Pew Research Center. Online at: http://www.pewinternet.org/2013/12/30/social-media-update-2013/ (2013). (2013).

19. M. D. Conover *et al*., *Fifth International AAAI Conference on Weblogs and Social Media* (2011).

20. D. C. Mutz, J. J. Mondak, The Workplace as a Context for Cross-Cutting Political Discourse. *J. Polit.* **68**, 140 (2006). doi:10.1111/j.1468-2508.2006.00376.x

21. S. Goel, W. Mason, D. J. Watts, Real and perceived attitude agreement in social networks. *J. Pers. Soc. Psychol.* **99**, 611–621 (2010). Medline doi:10.1037/a0020697

22. D. C. Mutz, Cross-cutting Social Networks: Testing Democratic Theory in Practice. *Am. Polit. Sci. Rev.* **96**, 112 (2002). doi:10.1017/S0003055402004264

23. B. Bishop, *The Big Sort: Why the Clustering of Like-Minded America is Tearing Us Apart* (Houghton Mifflin Harcourt, New York, 2008).

24. D. C. Mutz, P. S. Martin, *Am. Polit. Sci. Rev.* **95**, 97 (2001).

25. T. Mendelberg, Political Decision Making. *Deliber. Particip.* **6**, 151–193 (2002).

26. R. Huckfeldt, J. M. Mendez, T. Osborn, Disagreement, Ambivalence, and Engagement: The Political Consequences of Heterogeneous Networks. *Polit. Psychol.* **25**, 65–95 (2004). doi:10.1111/j.1467-9221.2004.00357.x

27. R. Bond, S. Messing, Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook. *Am. Polit. Sci. Rev.* **109**, 62–78 (2015). doi:10.1017/S0003055414000525

28. E. Riloff, *Proceedings of the Thirteenth National Conference on Artificial Intelligence* (AAAI Press, Portland, Oregon, 1996), vol. 2 of AAAI'96, pp. 1044–1049.

29. E. Riloff, *Case-Based Reasoning Research and Development*, K. D. Ashley, D. G. Bridge, eds., no. 2689 in Lecture Notes in Computer Science (Springer Berlin Heidelberg, 2003), pp. 4–4.

30. S. Gupta, C. D. Manning, *Proceedings of the SIGNLL Conference on Computational Natural Language Learning* (SIGNLL, Baltimore, MD, 2014), vol. 98.

31. S. Gupta, C. D. Manning, *Proceedings of the ACL 2014 Workshop on Interactive Language Learning, Visualization, and Interfaces* (ACL-ILLVI) (Baltimore, MD, 2014), p. 38.

32. C. Budak, S. Goel, J. M. Rao, Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis (2014); available at http://ssrn.com/abstract=2526461.

33. T. Groseclose, J. Milyo, A Measure of Media Bias. *Q. J. Econ.* **120**, 1191–1237 (2005). doi:10.1162/003355305775097542

34. M. Gentzkow, J. M. Shapiro, What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica* **78**, 35–71 (2010). doi:10.3982/ECTA7195

35. S. K. Whitbourne, Openness to experience, identity flexibility, and life change in adults. *J. Pers. Soc. Psychol.* **50**, 163–168 (1986). Medline doi:10.1037/0022-3514.50.1.163