Supplementary Information for

Polarized information ecosystems can reorganize social networks via information cascades

Christopher K. Tokita, Andrew M. Guess*, Corina E. Tarnita*

Christopher K. Tokita.
E-mail: christopher.tokita@gmail.com

This PDF file includes:

Supplementary text
Figs. S1 to S8 (not allowed for Brief Reports)
Table S1 (not allowed for Brief Reports)
SI References

Supporting Information Text

Supplemental Methods

Monitoring the follower networks of Twitter users. For each of our four news sources of interest, we used the Twitter API to sample 3,000 random Twitter users that followed that news source's account. We then used the *tweetscores* R package to estimate the ideology of each of the 12,000 sampled users (1).

From the original pool of approximately 12,000 sampled users who follow a news outlet of interest, we monitored the follower networks of 1,000 liberal followers of CBS News, 1,000 conservative followers of USA Today, 1,000 liberal followers of Vox, and 1,000 conservative followers of the Washington Examiner. When selecting the 1,000 random users to monitor for a given news outlet, we first conducted filters based on self-reported geolocation and other Twitter profile attributes to reasonably filter down to US-based users who primarily tweet in English. We then pulled the complete follower network of each of our 4,000 monitored users at the beginning and end of a 6-week period from August to September 2020, allowing us to assess who unfollowed these users over this period of time. Finally, using the initial follower networks of each monitored user, we estimated the ideology of up to 50 random followers to create a baseline for the ideological composition of each user's follower network.

Using the initial and final follower networks of each individual, we calculated the rate of unfollows by opposite-ideology users in their follower network. To determine whether an unfollow event was breaking a cross-ideology social tie, users and their unfollowers were classified simply as either liberal (ideology score < 0) or conservative (ideology score > 0). We excluded from analysis unfollowers for whom we could not estimate ideology, either because the account had been suspended, deleted, or made private, or because they did not follow any of the political, news, or cultural accounts needed to do the estimation. We then compared the proportion of unfollowers that were of the opposite ideology (i.e., conservative unfollowers if the focal user is liberal) against the estimated proportion of opposite-ideology followers in the focal user's initial follower network. This allowed us to account for the fact that many user's follower networks were not ideologically balanced and set the baseline expectation that if unfollows were random then the proportion of unfollows by opposite-ideology users should match the proportion of followers that were of the opposite ideology.

Calculating point estimates of group mean in the observational study. We used Bayesian inference to calculate the mean frequency of cross-ideology unfollows in the observational study. Relative to typical frequentist statistics, Bayesian statistical methods allow us to estimate statistics, like group mean, while also estimating our uncertainty of this estimate. Moreover, hypothesis testing is more intuitive in that it allows us to directly compare the evidence for two competing hypotheses—in our case, whether the frequency of cross-ideology unfollowing was higher in the low-correlation information ecosystem or the high-correlation information ecosystem.

We estimated group means for the observational treatments—information ecosystem at a coarse level and news source at a finer level—using Bayesian generalized linear modeling in the *brms* package in R. We assumed a Gaussian likelihood function and a prior distribution that had the same mean and standard deviation as the population. For each estimation, we ran four Markov chains, each with a warmup of 5,000 iterations and 15,000 total iterations, totaling 40,000 total samples of the posterior. We calculated the credible intervals—the Bayesian counterpart to confidence intervals—as the highest-density intervals of the posterior distribution.

Analyzing the information correlation of news sources. In our model, the primary parameter of interest, γ, dictates how similarly news sources assign importance to the same topic/story. In the real world, importance is signaled by the amount of coverage a news source gives to a topic and the language it uses to describe the topic. If news sources are highly correlated in how they assign importance to topics, then we would expect that they cover topics in similar proportion and with similar language. Therefore, to generate an empirical measurement of γ for news sources, we can look at the similarity of language over 24-hour news cycles.

Using the Twitter API, we pulled all available tweets from *@AP*, *@CBSNews*, *@USATODAY*, *@voxdotcom*, and *@dcexaminer* at three separate time points: December 7, 2020; December 14, 2020; and December 22, 2020. Each pull included the 3,200 most recent tweets from a given account. The Twitter API limits downloads to the most recent 3,200 tweets from a given account, and given that each account does not tweet at the same frequency, the time span of tweets in our dataset ranged from 34 days for the Washington Examiner to 76 days for Vox.

Next, we analyzed the text posted on each news outlet's Twitter account in a given 24-hour news cycle. For a given Twitter account, we first cleaned and stemmed all words in a given tweet. To then count the frequency of words shared in a given 24-hour news cycle, we grouped tweets by day and created a term-frequency matrix for each day of each Twitter account.

Finally, to estimate the information correlation γ of each news source, we calculated the average cosine similarity between a news source and the Associated Press (AP). Since the AP provides much basic-fact reporting for news outlets and does not partake in editorials, we assumed that the AP provides an accurate proxy for the baseline information available to a news consumer. Thus, when considering the correlation of a news source to the general information ecosystem, we compare the similarity between that news source's tweets and the tweets of the AP.

Since cosine similarity takes values in the range $[0, 1]$ and most empirical values in our dataset fall below ~ 0.3, we converted the estimated empirical γ to the standardized range $[-1, 1]$ (as found in our model). To first standardize the empirical γ values, we divided each empirical γ value by the mean self-similarity of the AP to itself. Since it stands to reason that no news source could be more correlated than it is to itself, we randomly divided each of AP's tweets into two pools and then calculated the average cosine similarity of a given day. This process was repeated 200 times and the maximum mean cosine similarity was taken as the self-similarity of AP to itself. Lastly, we rescaled the now standardized empirical γ values from the range $[0, 1]$ to the range $[-1, 1]$ according to the formula $2\gamma - 1$.

Table S1. Parameter settings for model. Unless stated otherwise, these are the default values used in all simulations.

	Description	Values or range in simulation
T	Simulation length in time steps	3.0×10^6
N	Number of individuals	200
k	Average degree of social network	8
ψ	Fraction of society that directly samples news sources each round (drawn randomly each round)	0.1
γ	The information ecosystem as measured by the correlation in coverage between media sources	$[-1, 1]$

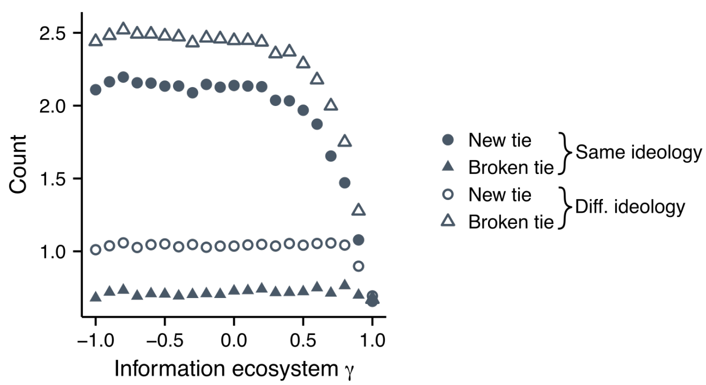

Fig. S1. The average number of changes in initial social ties (i.e., ties present at $t = 0$) over the course of the simulation. Tie changes are categorized by comparing an individual's social ties in the initial and final social networks.

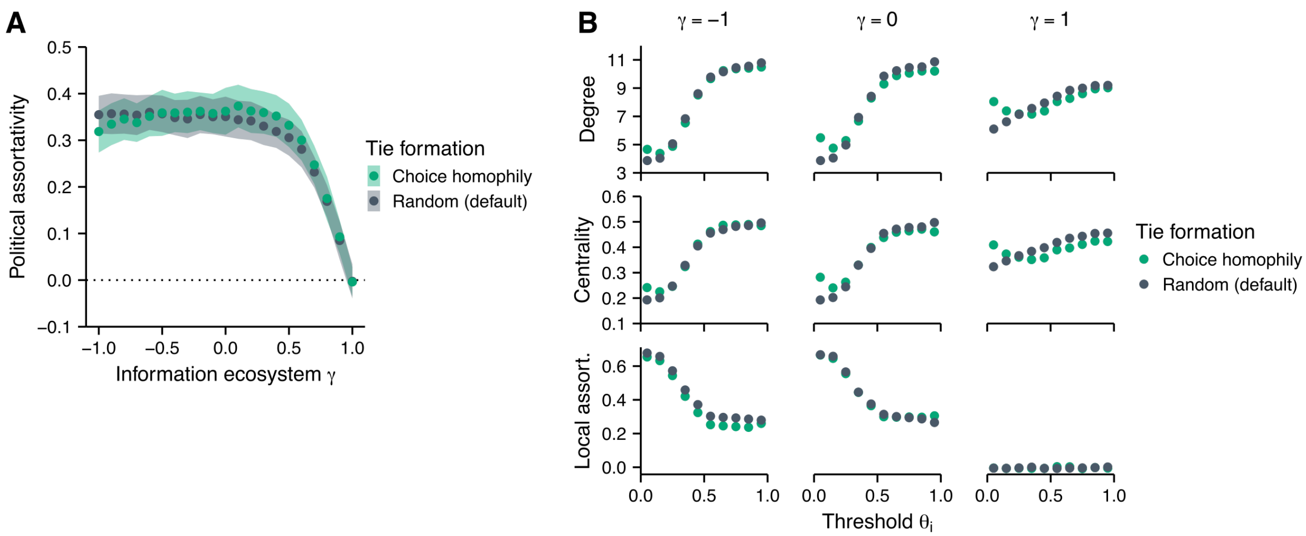

Fig. S2. The effect of adding choice homophily to the tie formation process in the model: during the tie formation step of the model, when there has been a broken tie and another individual is randomly selected to form a new tie, the individual forms a tie randomly to one individual they see as reacting "correctly" to the news. Individuals determine the correct behavior of others based on their own threshold value θ_i and the significance of the story s from their preferred news source. For example, if the tie-forming individual finds that $s > \theta_i$, then she forms a tie randomly with one active individual. (A) Mean political assortativity ($\pm$ SD) of the final $t = T$ social networks as a function of the information ecosystem and the tie formation process. (B) Mean degree, centrality, and local assortativity of individuals broken out by threshold and information ecosystem. Each point is a binned average of individuals with that threshold value.

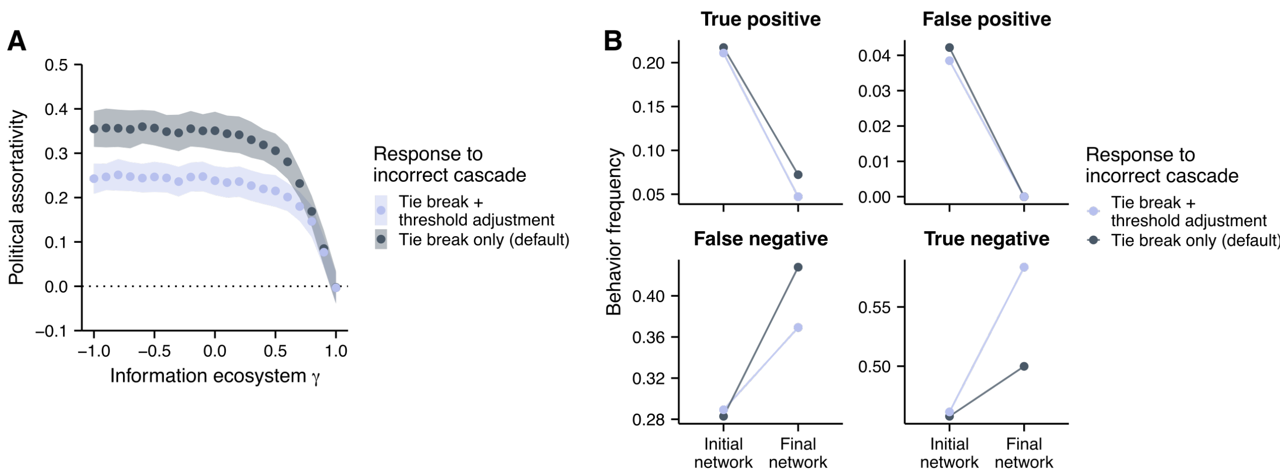

Fig. S3. The effect of allowing individuals to become insensitive to information in response to cascades. In an extension of the model, individuals respond to cascades by adjusting both social ties and their personal thresholds: when an individual finds that they incorrectly got swept up in a cascade, she both breaks a social tie with an "incorrect" neighbor and increases her threshold by amount $\Delta = 0.05$. Thus, in response to reacting to "unimportant" news, individuals become insensitive to information over time. (A) Mean political assortativity ($\pm$ SD) of the final $t = T$ social networks as a function of the information ecosystem and how individuals respond to incorrect cascades. (B) Mean behavior rates of individuals in the base model and the extension in which individuals also adjust their thresholds in response to participating in "incorrect" cascades. When individuals become insensitive to information over time, the biggest change in behavior is that more individuals find news unimportant and remain inactive, as evidenced by the rise in true negative behavior and drop in false negative behavior.

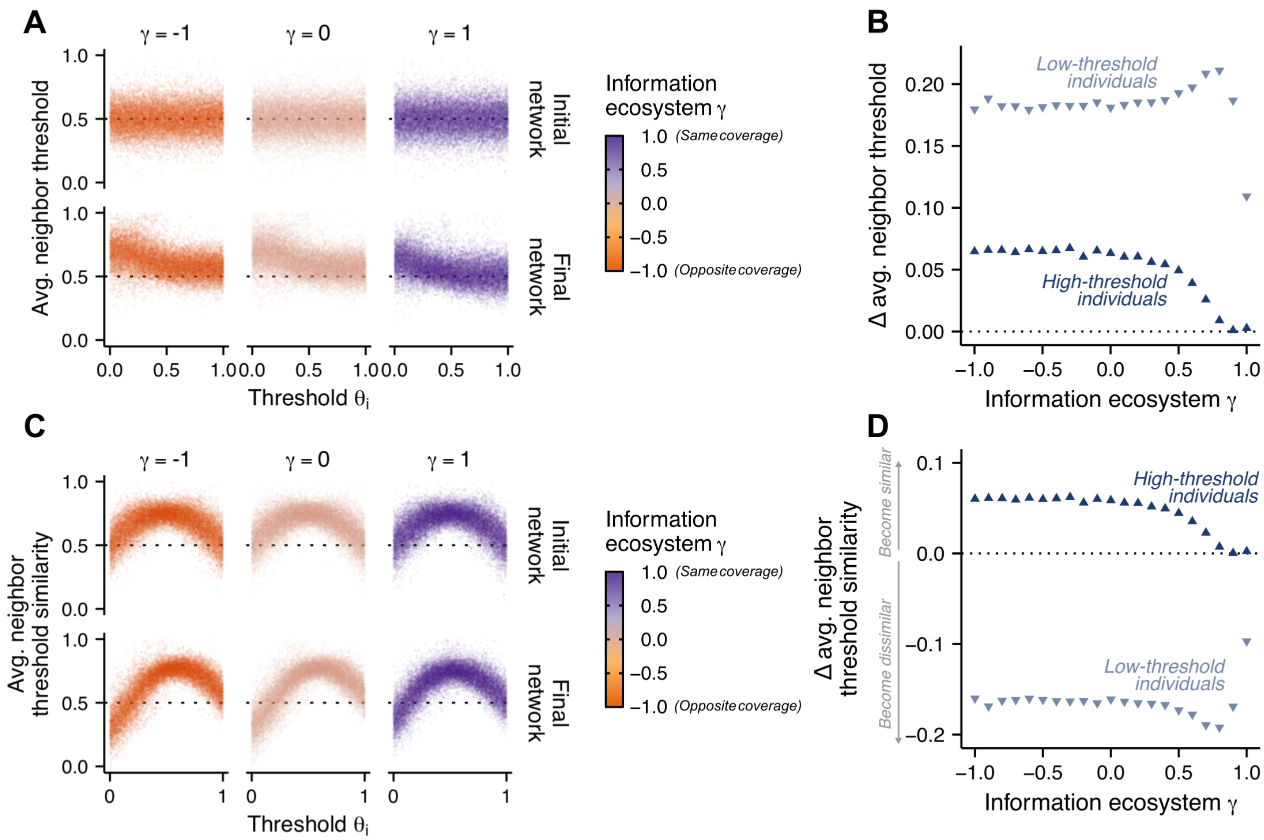

Fig. S4. (A) Relationship between an individual's threshold θ_i and the average threshold of their neighbors, broken out by initial/final network and information ecosystem. Points are individuals across all 100 replicate simulations. (B) Mean change in the average threshold of an individual's neighbors between the initial and final social network configuration for low-threshold ($\theta_i < 0.25$) and high-threshold ($\theta_i > 0.75$) individuals. Positive values mean the average threshold of neighbors increased over the course of the simulation. Of note, threshold values do not change in our model; only social network ties change. Therefore, an increase in mean neighbor thresholds means that individuals tend to gain connections to higher-threshold individuals and/or lose connections to lower-threshold individuals. (C) Relationship between an individual's threshold θ_i and the average threshold similarity of their neighbors, broken out by initial/final network and information ecosystem. Threshold similarity between an individual i and a neighbor j was calculated as $1 - |\theta_i - \theta_j|$, where $\theta_i - \theta_j$ is the distance between the two individual's thresholds. Points are individuals across all 100 replicate simulations. (D) Mean change in the threshold similarity between an individual and their neighbors, broken out for low-threshold ($\theta_i < 0.25$) and high-threshold ($\theta_i > 0.75$) individuals.

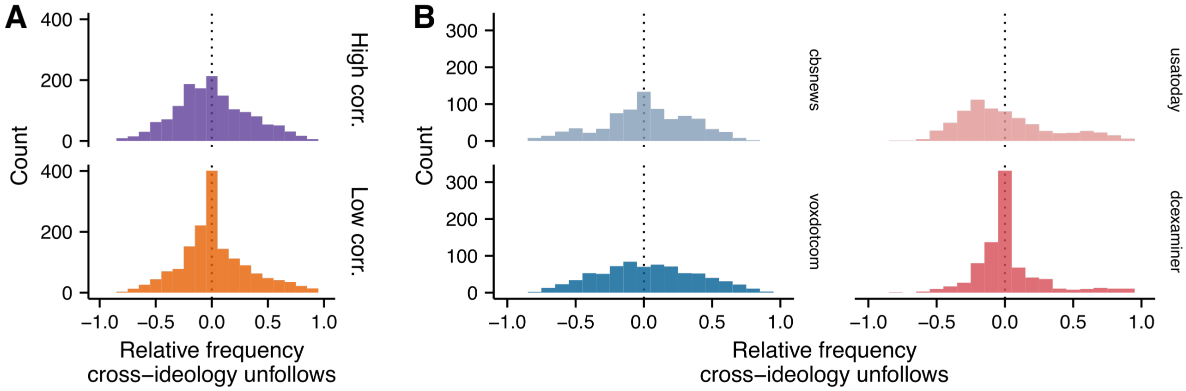

Fig. S5. Histogram of the cross-ideology unfollows in the observational Twitter study, broken out by (A) information ecosystem and (B) news source.

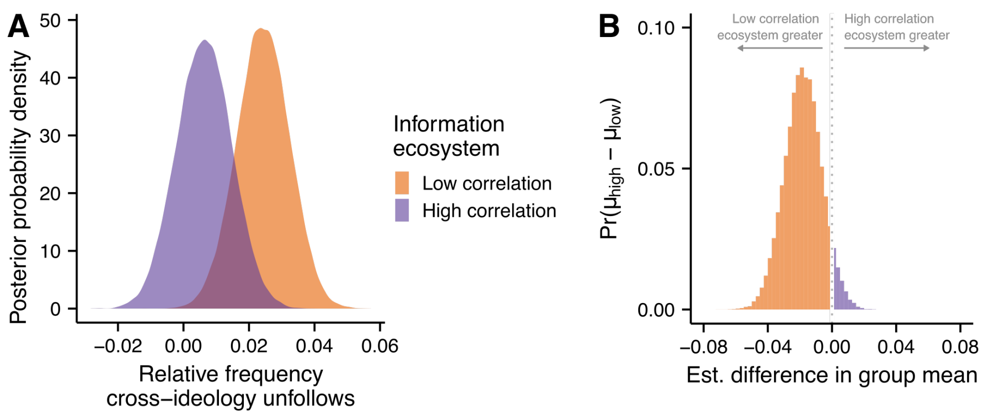

Fig. S6. (A) Posterior distribution for the mean relative frequency of cross-ideology unfollows. (B) Posterior of the difference in mean between the high- and low-correlation information ecosystems. 94.0% of the distribution lies below zero, which indicates a high likelihood that the mean in low-correlation information ecosystem is higher.

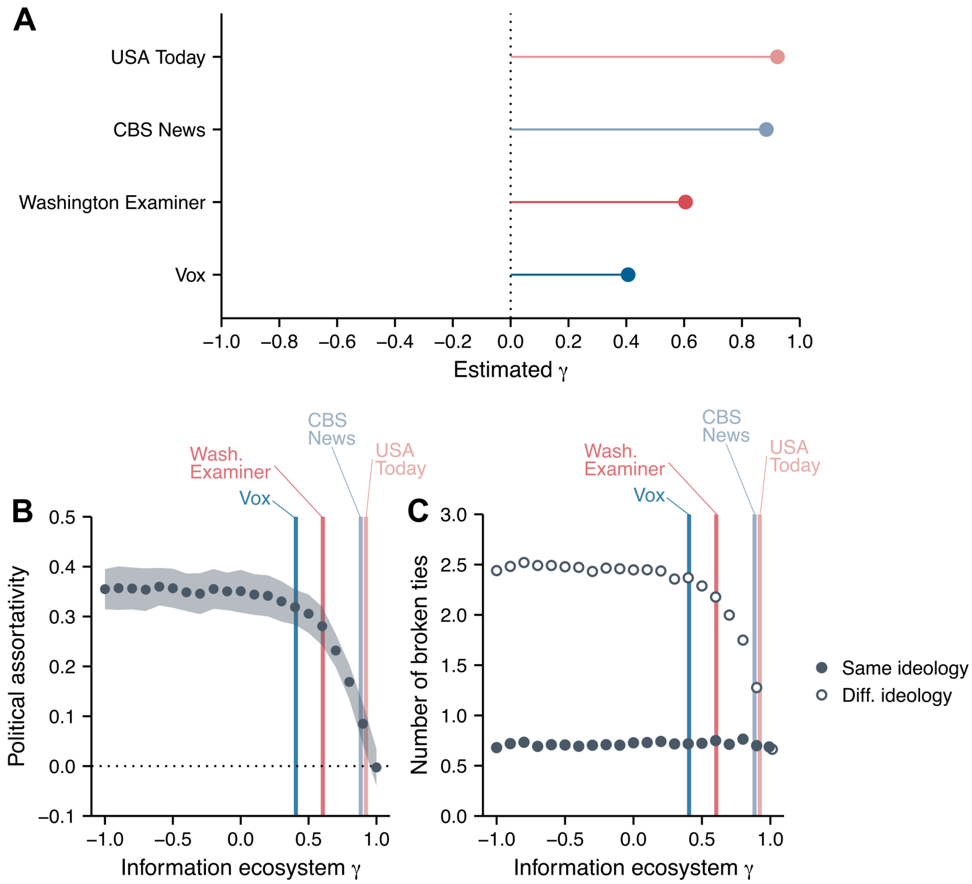

Fig. S7. (A) Estimated information correlation γ of each news source that we followed in our observational study. (B) Model predictions for final network assortativity with the estimated γ for each news source overlaid. (C) Model predictions for the number of broken ties with the estimated γ for each news source overlaid

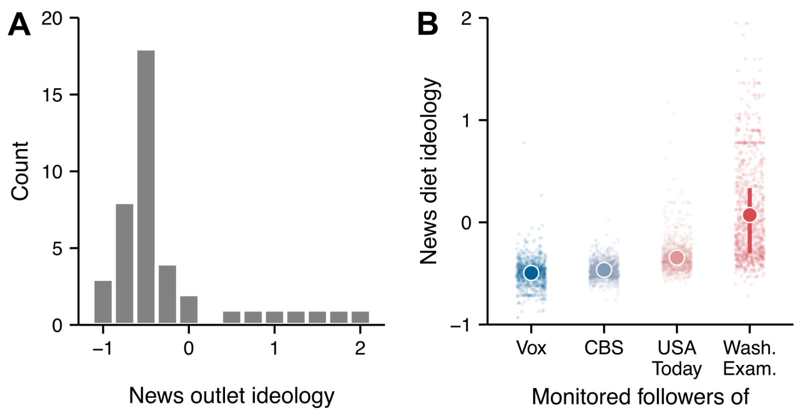

Fig. S8. (A) Histogram of news outlet ideology in our news diet data set. The data set was composed of the estimated ideology/slant of 42 major news sources. (B) News diet ideology of each of our 4,000 monitored users. Each point is the media diet of one user in our observational study, and the large points represent the mean ($\pm$ interquartile range). For a given user, her news diet ideology was calculated by taking the average ideology of the news sources she follows on Twitter.

References

1. P Barberá, JT Jost, J Nagler, JA Tucker, R Bonneau, Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychol. Sci.* **26**, 1531–1542 (2015).