A Online Appendix for "Homophily, Group Size, and the Diffusion of Political Information in Social Networks"

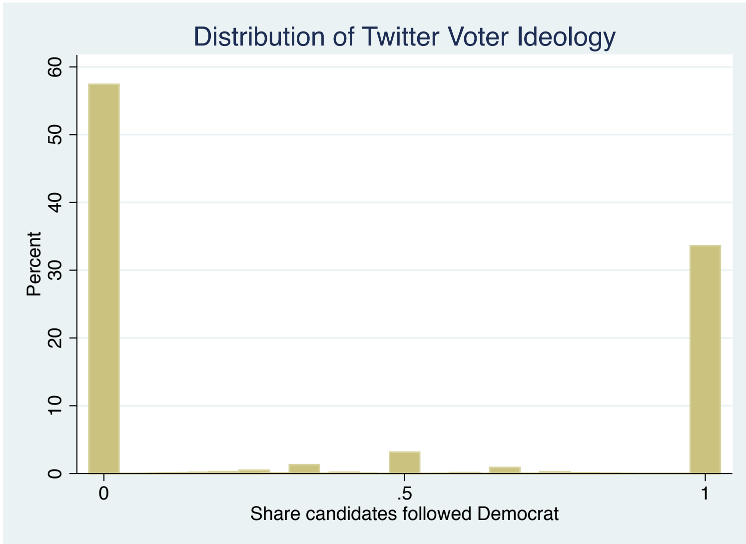


Figure A1: Histogram of Twitter Voters by Share of Democrats Followed

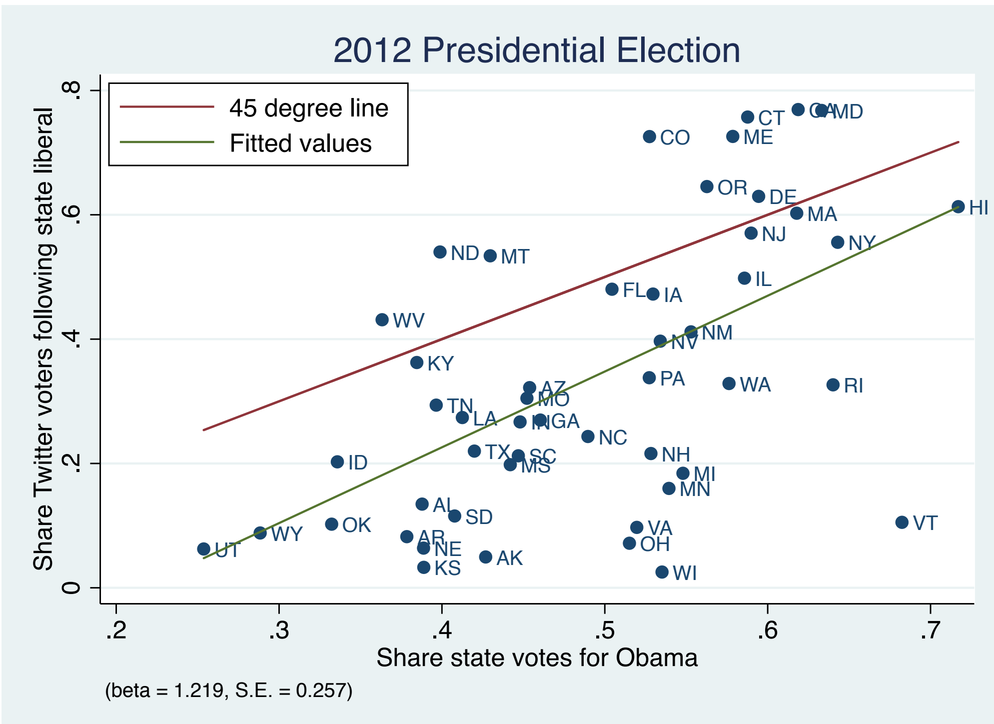
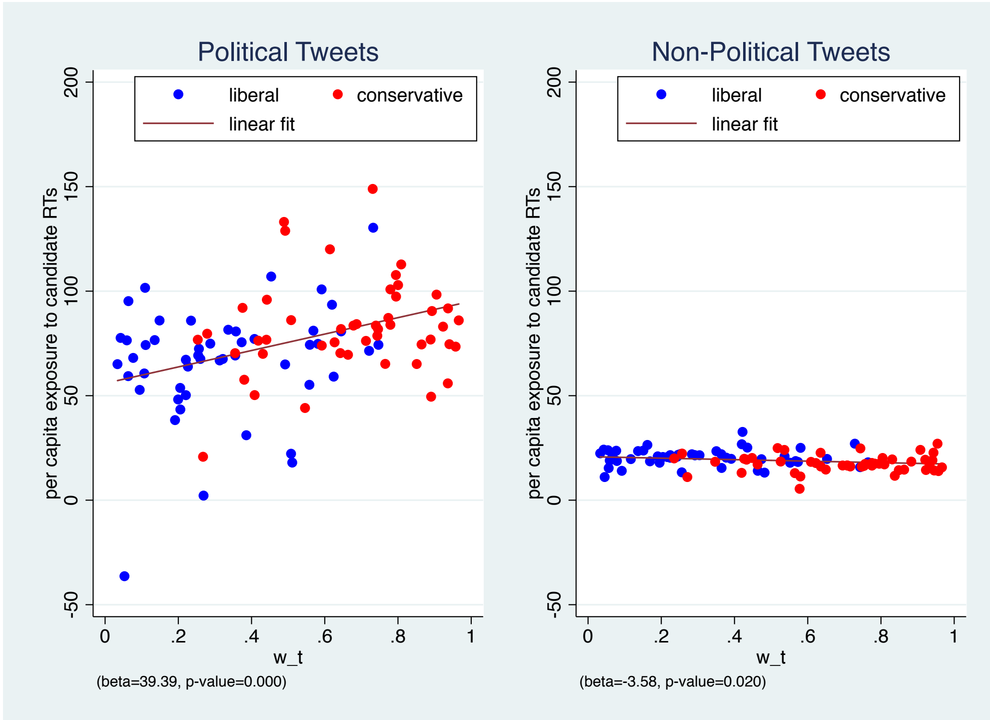
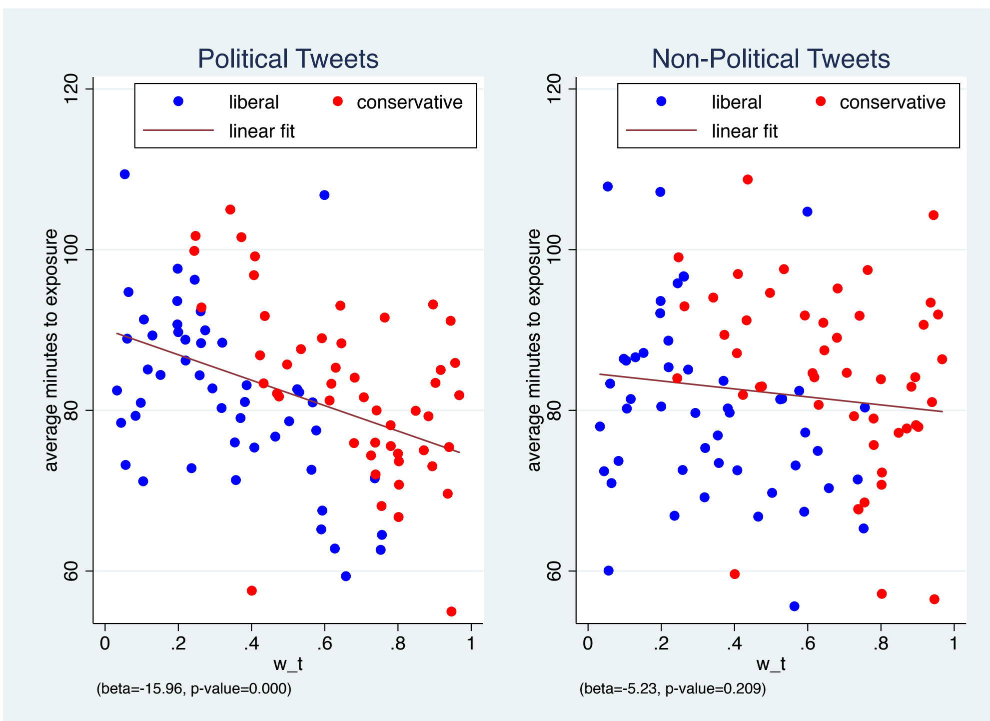


Figure A2: Twitter Voter Ideology and Presidential Vote share



(a) Volume of Exposure and Group Size



(b) Speed of Exposure and Group Size

Figure A3: Volume and Speed of Exposure by Content of Candidate Tweets

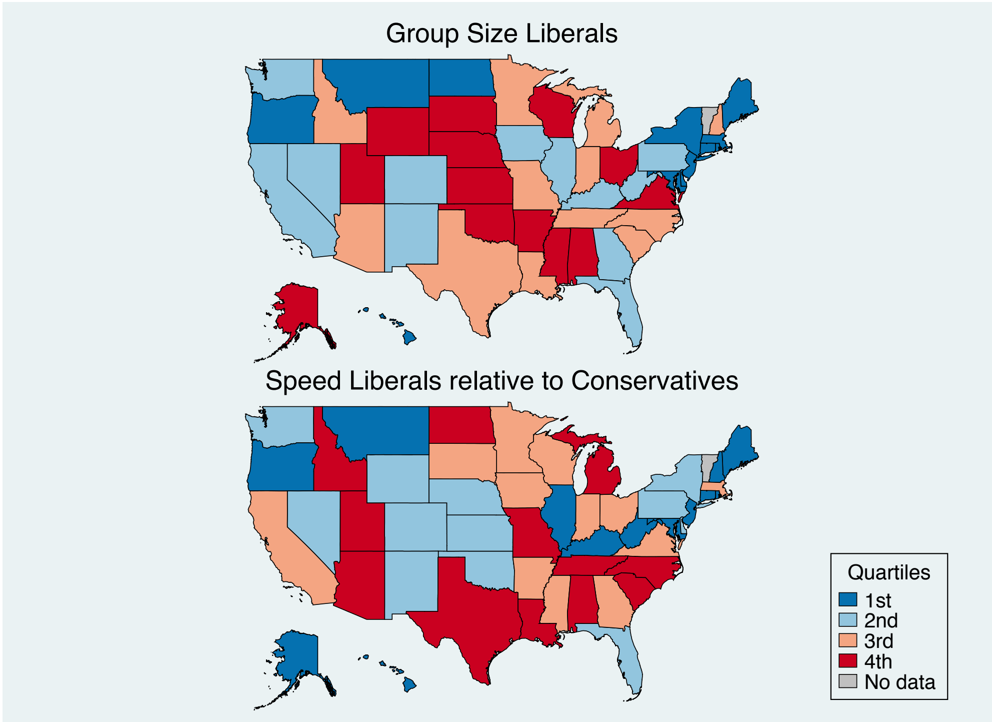


Figure A4: Group Size and Speed of Exposure to Information

Proof of Proposition 2: We have shown in the text that $F_C^1 > F_L^1$, and we now show that $F_C^{\tau-1} > F_L^{\tau-1}$ implies that $F_C^{\tau} > F_L^{\tau}$. Note first that:

$$F_C^{\tau} - F_L^{\tau} = F_C^{\tau-1} - F_L^{\tau-1} + (1 - F_C^{\tau-1}) f_C^{\tau} - (1 - F_L^{\tau-1}) f_L^{\tau}$$

which can be re-written as:

$$F_C^{\tau} - F_L^{\tau} = F_C^{\tau-1} - F_L^{\tau-1} + [(1 - F_L^{\tau-1}) - (F_C^{\tau-1} - F_L^{\tau-1})][f_L^{\tau} + (f_C^{\tau} - f_L^{\tau})] - (1 - F_L^{\tau-1}) f_L^{\tau}$$

expanding the terms in brackets and re-arranging, we have that:

$$F_C^{\tau} - F_L^{\tau} = (F_C^{\tau-1} - F_L^{\tau-1})(1 - f_L^{\tau}) + [1 - F_C^{\tau-1}](f_C^{\tau} - f_L^{\tau})$$

Thus, $F_C^{\tau} - F_L^{\tau}$ is positive if $f_C^{\tau} - f_L^{\tau}$ is positive. This latter difference can be written as:

$$\begin{aligned} f_C^{\tau} - f_L^{\tau} &= q w_C \pi_s F_C^{\tau-1} + q(1 - w_C) \pi_d F_L^{\tau-1} - q(1 - w_C) \pi_s F_L^{\tau-1} - q w_C \pi_d F_C^{\tau-1} \\ &= q F_C^{\tau-1} w_C (\pi_s - \pi_d) + q F_L^{\tau-1} (1 - w_C)(\pi_d - \pi_s) \\ &= q(\pi_s - \pi_d)[F_C^{\tau-1} w_C - F_L^{\tau-1}(1 - w_C)] \end{aligned}$$

This is positive under the maintained assumptions that $F_C^{\tau-1} > F_L^{\tau-1}$, $w_C > 0.5$, and $\pi_s > \pi_d$.

Proof of Proposition 3: Let expected time to exposure for conservatives and liberals, respectively, be given among the first ρ share exposed by $T^C = \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_C} \tau(F_C^\tau - F_C^{\tau-1})$ and $T^L = \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_L} \tau(F_L^\tau - F_L^{\tau-1})$, where $\bar{\tau}_C$ represents the time at which a fraction ρ of conservatives are exposed and likewise for $\bar{\tau}_L$. Using summation by parts, these can be written as $T^C = \bar{\tau}_C - \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_C-1} F_C^\tau$ and $T^L = \bar{\tau}_L - \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_L-1} F_L^\tau$. Taking the difference, we have $T^L - T^C = \bar{\tau}_L - \bar{\tau}_C - \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_C-1}(F_L^\tau - F_C^\tau) - \frac{1}{\rho}\sum_{\tau=\bar{\tau}_C-1}^{\tau=\bar{\tau}_L-1} F_L^\tau$. Using the fact that $\bar{\tau}_L - \bar{\tau}_C = \frac{1}{\rho}\sum_{\tau=\bar{\tau}_C-1}^{\tau=\bar{\tau}_L-1}\rho$, the difference can be written as $T^L - T^C = -\frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_C-1}(F_L^\tau - F_C^\tau) - \frac{1}{\rho}\sum_{\tau=\bar{\tau}_C-1}^{\tau=\bar{\tau}_L-1}(F_L^\tau - \rho)$. Since $F_L^\tau < F_C^\tau$ for all τ and $F_L^\tau < \rho$ for all $\tau < \bar{\tau}_L$, the result is established.

Proof of Proposition 4: Due to the symmetry of the model, it is the case that $C_C^\tau = L_L^\tau$ and that $L_C^\tau = C_L^\tau$ for all τ. Given this, we focus on exposure to conservative information, and, in particular, show that $C_C^\tau > C_L^\tau$ for all τ. Note first that

$$C_C^\tau - C_L^\tau = C_C^{\tau-1} - C_L^{\tau-1} + (1 - C_C^{\tau-1})c_C^\tau - (1 - C_L^{\tau-1})c_L^\tau$$

which can be re-written as:

$$C_C^\tau - C_L^\tau = C_C^{\tau-1} - C_L^{\tau-1} + [1 - C_L^{\tau-1} - (C_C^{\tau-1} - C_L^{\tau-1})][c_L^\tau + (c_C^\tau - c_L^\tau)] - (1 - C_L^{\tau-1})c_L^\tau$$

re-arranging, we have that:

$$C_C^\tau - C_L^\tau = (C_C^{\tau-1} - C_L^{\tau-1})(1 - c_C^\tau) + [1 - C_L^{\tau-1}](c_C^\tau - c_L^\tau)$$

Thus, the sign of $C_C^\tau - C_L^\tau$ involves a comparison of c_C^τ and c_L^τ, which can be written as:

$$\begin{aligned} c_C^\tau - c_L^\tau &= q0.5\pi_s C_C^{\tau-1} + q0.5\pi_d C_L^{\tau-1} - q0.5\pi_s C_L^{\tau-1} - q0.5\pi_d C_C^{\tau-1} \\ &= q0.5(\pi_s - \pi_d)(C_C^{\tau-1} - C_L^{\tau-1}) \end{aligned}$$

This is positive under the maintained assumption that $C_C^{\tau-1} > C_L^{\tau-1}$, and thus $C_C^\tau - C_L^\tau$ is also positive. Finally, we show that $C_C^1 > C_L^1$, which is implied by:

$$\begin{aligned} C_C^1 &= q0.5\pi_s\varepsilon_s + q0.5\pi_d\varepsilon_d - q^2 0.25\pi_s\varepsilon_s\pi_d\varepsilon_d \\ C_L^1 &= q0.5\pi_d\varepsilon_s + q0.5\pi_s\varepsilon_d - q^2 0.25\pi_s\varepsilon_s\pi_d\varepsilon_d \end{aligned}$$

Taking the difference, we have that:

$$C_C^1 - C_L^1 = q0.5(\pi_s - \pi_d)(\varepsilon_s - \varepsilon_d) > 0$$

Proof of Proposition 5: Focusing again on conservative information (without loss of generality), let expected time to exposure for conservatives and for liberals be given, respectively, among the first ρ share exposed by $T^C = \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_C} \tau(C_C^\tau - C_C^{\tau-1})$ and $T^L = \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_L} \tau(C_L^\tau - C_L^{\tau-1})$, where $\bar{\tau}_C$ represents the time at which a fraction ρ of conservatives are exposed and likewise for $\bar{\tau}_L$. Using summation by parts, these can be written as $T^C = \bar{\tau}_C - \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_C-1} C_C^\tau$ and $T^L = \bar{\tau}_L - \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_L-1} C_L^\tau$. Taking the difference, we have $T^L - T^C = \bar{\tau}_L - \bar{\tau}_C - \frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_C-1}(C_L^\tau - C_C^\tau) - \frac{1}{\rho}\sum_{\tau=\bar{\tau}_C-1}^{\tau=\bar{\tau}_L-1} C_L^\tau$. Using the fact that $\bar{\tau}_L - \bar{\tau}_C = \frac{1}{\rho}\sum_{\tau=\bar{\tau}_C-1}^{\tau=\bar{\tau}_L-1} \rho$, the difference can be written as $T^L - T^C = -\frac{1}{\rho}\sum_{\tau=1}^{\tau=\bar{\tau}_C-1}(C_L^\tau - C_C^\tau) - \frac{1}{\rho}\sum_{\tau=\bar{\tau}_C-1}^{\tau=\bar{\tau}_L-1}(C_L^\tau - \rho)$. Since $C_L^\tau < C_C^\tau$ for all τ and $C_L^\tau < \rho$ for all $\tau < \bar{\tau}_L$, the result is established.

Measuring Ideological Segregation: Following Gentzkow and Shapiro (2011), we also compute segregation in media consumption using information on the set of media outlets followed by each voter on Twitter.[1] For comparison purposes, we also compute network isolation. For each voter $j \in J$, let v_{jC} denote the number of conservative followers and v_{jL} the number of liberal followers. We can then define the *share conservative* of voter j as the fraction of his or her followers who are conservative:

$$share\ conservative_j = \frac{v_{jC}}{v_{jC} + v_{jL}}.$$

We can then define conservative exposure for each voter i as follows:

$$conservative\ exposure_i = \frac{1}{\sum_{j \in J} \phi_{ij}} \sum_{j \in J} \phi_{ij} \times share\ conservative_j,$$

where $\phi_{ij} \in \{0,1\}$ as an indicator equal to one if voter i follows voter j. Taking averages across voters within groups, we then have conservative exposure for conservatives and conservative exposure among liberals. With these in hand, the isolation index is given by:

$$isolation = conservative\ exposure_C - conservative\ exposure_L,$$

where $conservative\ exposure_t = \frac{1}{I_t}\sum_{i \in I_t} conservative\ exposure_i$.

This index varies between 0 and 1 and captures the degree to which conservatives, relative to liberals, have a greater tendency to follow voters whose other followers are conservative. As the index increases, both groups become increasingly isolated from each other, as measured by a

[1] This measure has been developed by White (1986) and Cutler et al. (1999), and widely applied to study ethnic and urban segregation.

shrinking share of voters who have both conservative and liberal followers.

Table 8 reports the results from computing these isolation measures. As shown in the first row, conservative exposure among conservatives for the network-based measure is 0.776, and conservative exposure among liberals is 0.372, implying an isolation index of 0.403. Note that this result differs from those in Gentzkow and Shapiro (2011), with a baseline estimate of segregation equal to 0.075. To attempt to reconcile these two sets of findings, high segregation when examining links on Twitter and low segregation when examining consumption of news on the internet, we next examine two differences between these studies. First, it is plausible that our sample, constructed by selecting users who follow politicians, may tend to disproportionately include individuals with strong preferences for linking to like-minded users. To investigate this issue, we focus on Twitter users who follow both parties. Consistent with the view that these voters have weaker preferences for linking to like-minded users, we find that segregation is lower for moderates, when compared to the entire sample. Second, as noted above, we use information on the followers of our sample of media outlets to compute segregation in media consumption on Twitter. As shown in the third row, isolation in media consumption (0.241) for our sample of voters is significantly higher than the measures in Gentzkow and Shapiro (2011) but is significantly lower than our network-based measure of isolation, which equals 0.394 in this sub-sample of voters. Finally, we combine these two approaches by computing isolation in media consumption for moderates. As shown, segregation in media consumption for moderates equals 0.067, which is on par with the measure in Gentzkow and Shapiro (2011), but significantly lower than network segregation for this group, which equals 0.228.

Table A1: Ideological Segregation in Media and Social Networks

	Network Segregation			Segregation in Media Consumption		
	Conservative Exposure			Conservative Exposure		
Followers of	Conservative	Liberals	Isolation	Conservative	Liberal	Isolation
Baseline	0.776	0.372	0.403	n/a	n/a	n/a
Both parties	0.716	0.499	0.217	n/a	n/a	n/a
Media and candidates	0.780	0.387	0.394	0.789	0.547	0.241
Media and both parties	0.717	0.489	0.228	0.723	0.656	0.067

Table A2: Diffusion of Information and Time to Exposure: Continuous Ideology Measure

	Linear Regression		Cox Survival Analysis
	minutes	ln(minutes)	
Liberal voter	20.1629***	0.2512***	-0.2487***
	(0.1002)	(0.0009)	(0.0011)
Ideology mismatch	87.6998***	1.3438***	-1.0901***
	(0.1002)	(0.0009)	(0.0011)
Tweet FE	Yes	Yes	Yes
N	15,629,553	15,629,553	15,629,553
Dependent variable mean	56.78	2.00	56.78

Notes: *** denotes significance at the 99 percent level, ** denotes significance at the 95 percent level, and * denotes significance at the 90 percent level, Sample is based upon the first one percent of each group exposed to tweets that reach at least one percent of each group. Liberal voter here is a continuous measure based upon the fraction of Democratic candidates followed. The dependent variable is minutes to exposure in column 1 and the natural log of minutes to exposure in column 2. Column 3 estimates a Cox survival model, using data on minutes to exposure. In all specifications, the unit of observation is an exposed voter-candidate tweet. Ideology mismatch equals the fraction of Democrats followed for a Republican tweet and the fraction of Republicans followed for a Democratic teweet.

References

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999). The rise and decline of the american ghetto. *Journal of Political Economy 107*(3), 455–506.

Gentzkow, M. and J. M. Shapiro (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics 126*(4), 1799–1839.

White, M. J. (1986). Segregation and diversity measures in population distribution. *Population index*, 198–221.