The Social Structure of Political Echo Chambers:

Variation in Ideological Homophily in Online Networks *

Andrei Boutyline

University of California, Berkeley

Robb Willer

Stanford University

September 7, 2015

Revised and resubmitted to ***Political Psychology***

Keywords: political homophily, ideology, motivated cognition, *Twitter*

[*] This research was supported in by fellowships from National Science Foundation Graduate Research Fellowship Program and Interdisciplinary Graduate Education and Research Traineeship Program. We thank Claude Fischer, Michael Hout, Fabiana Silva, Stephen Vaisey, and participants of the Berkeley Mathematical, Analytical and Experimental Sociology working group for feedback on the paper. Direct all correspondence to Andrei Boutyline at Department of Sociology, 410 Barrows Hall, University of California, Berkeley, CA 94720. Email: boutyline@berkeley.edu

ABSTRACT. We predict that people with different political orientations will exhibit systematically different levels of political homophily, the tendency to associate with others similar to oneself in political ideology. Research on personality differences across the political spectrum finds that both more conservative and more politically extreme individuals tend to exhibit greater orientations towards cognitive stability, clarity, and familiarity. We reason that such a "preference for certainty" may make these individuals more inclined to seek out the company of those who reaffirm, rather than challenge, their views. Since survey studies of political homophily face well-documented methodological challenges, we instead test this proposition on a large sample of politically engaged users of the social networking platform *Twitter*, whose ideologies we infer from the politicians and policy non-profits they follow. As predicted, we find that both more extreme and more conservative individuals tend to be more homophilous than more liberal and more moderate ones.

INTRODUCTION

We draw on research on personality differences across the political spectrum to develop and test the prediction that people with different political orientations will exhibit different levels of political homophily, the tendency to choose to associate with others similar to oneself in political ideology. Ideological groups with greater political homophily possess political networks with more ties among their members, and fewer ties with individuals possessing different ideologies. Thus, greater political homophily is associated with decreased chances of politically diverse interactions and increased rates of interactions with ideologically similar others that tend to reinforce individuals' views and enhance their commitment to their ideological group. These outcomes are in turn likely to increase the polarization of public opinion and promote participation in political collective action.

Since at least John Stuart Mill (1859), political theorists have argued that dialogue across lines of political difference is a key pre-requisite for sustaining a democratic citizenry. Mill held that political disagreement enables individuals to develop skills for critically assessing political claims, and provides the challenge necessary for determining if one's own ideas are justified. Hannah Arendt similarly argued that debate "constitutes the very essence of political life," (Arendt 1961:241), irreplaceable for forming enlightened political opinions that reach beyond the limits of one's own subjectivity to incorporate the standpoints of others. Empirical work on consequences of disagreement has echoed many of these points. Existing research shows that individuals without exposure to such cross-cutting discourse are far less likely to see opposing viewpoints as legitimate, and less able to provide rationales for their own political decisions (Huckfeldt, Mendez, and Osborn 2004; Price, Cappella, and Nir 2002). Such individuals are also more likely to hold extreme attitudes about candidates consisting of entirely positive or negative assessments (Huckfeldt et al. 2004). Moreover, the lack of personal ties to those with different political views is likely to have detrimental effects on political tolerance (Mutz 2002a). Increased political homophily, and decreased cross-cutting contact, is therefore a likely source of polarization and political discord.

Conversely, political homophily creates dense clusters of within group-ties, which prior work shows reinforce behavioral norms and increase social pressure to take part in costly or risky activities (Centola and Macy 2007; Centola 2010). Politically homophilous networks have significant advantages for diffusing political behaviors that require normative pressure or social confirmation—including behaviors like turning out to vote, attending political protests, and engaging in potentially contentious political speech (González-Bailón et al. 2011; Romero, Meeder, and Kleinberg 2011; Kim and Bearman 1997; Knoke 1990). At the same time, political homophily may also insulate individuals from exposure to false or offensive information.

Further, a relative dearth of cross-cutting ties is itself a likely resource for collective action, as exposure to dissent can undermine commitment to the group and the extent to which the group's beliefs are taken as facts. Experimental and observational evidence suggests that heterogeneous ties increase ambiguity, which has a demotivating effect on political participation (Eveland and Hively 2009; Mutz 2002b; Visser and Mirabile 2004)–an effect that has been shown to hold across national settings, and in both online and offline networks (Liu, Dai, and Wu 2013; Mutz 2006; Valenzuela, Kim, and Zúñiga 2012). Campbell summarizes this work by pointing out that strength of preferences, such as identification with a political cause, "does not exist in a vacuum; it is reinforced by a social network of like-minded politicos" (2013:41).

Recently, a number of scholars have sought to qualify this effect by examining variation in consequences of cross-cutting exposure. For example, Jang (2009) found that, while cross-cutting ties are often demotivating, they also motivate participation among the most politically alienated individuals by increasing their understanding of the real differences between competing positions. Klofstad, Sokhey, and McClurg (2013) also found that effects of disagreement vary between kinds of contact and participation, but are overwhelmingly negative. Campbell's (2013) review of literature on networks and participation similarly suggests that, though the effect of cross-cutting ties may not always be negative, it is rarely positive.[1] Thus, while its effects are not monolithic, political homophily on average appears to be an asset for many kinds of collective action.

But how might political homophily vary by individuals' ideology? Two bodies of research show that people at different points in the political spectrum exhibit different levels of desire for clarity, certainty, stability, and familiarity—a cluster of traits we refer to as *preference for certainty*. First, a long line of work from political psychology finds that more conservative individuals exhibit greater preferences for certainty than more liberal ones (Jost et al. 2003a). Second, research on group identity hews that individuals on either ideological extreme possess greater preferences for certainty than more moderate ones (Greenberg and Jonas 2003; Hogg 2007). These findings suggest that more conservative or more extreme individuals may exhibit higher levels of political homophily, as they might be expected to place greater value on encountering concurring opinions and avoiding dissenting ones. As individuals with greater preferences for certainty seek it through social contact, their networks may come to resemble "echo chambers," providing them with reaffirmation and shielding them from disagreement.

[1] Campbell (2013) also highlights research showing that that exposure to disagreement through heterogeneous political *contexts* (as opposed to through cross-cutting *network ties*) may increase motivation through sparking interest and engagement. Nir's (2005) finding that "ambivalent" networks (i.e., those with both homophilous and non-homophilous ties) have a positive effect on motivation similarly demonstrates that other forms of exposure to disagreement may be motivating.

These intuitions are difficult to test with traditional survey data on political networks, which face well-documented methodological challenges, including a substantial pro-homophily bias in respondents' recall of their alters' political orientations, and difficulties establishing "baseline" rates of network homogeneity expected from random mixing (DiPrete et al. 2011; McPherson and Smith-Lovin 1987). Here, we address these problems by using network data from *Twitter*, an online service used by 12% of adult Americans (Smith and Brenner 2012). Employing a recently validated technique for ideological measurement of Twitter users (Golbeck and Hansen 2014), we infer users' political ideology from the ideological positions of members of Congress and policy non-profits they initiate ties with. We then test our hypotheses by examining 238,943 ego networks from across the political spectrum. The Twitter data are not a representative sample of United States voters or any other offline population, which precludes direct statistical generalization of our results to offline phenomena. At the same time, the size and diversity of the Twitter population as well as the observability of Twitter activity bring novel advantages that help overcome long-standing problems common to more traditional data on political networks.

Uncertainty and Threat

In developing our claims about the relationship between ideology and political homophily, we draw upon the substantial literature on personality and political attitudes in social psychology and political science. Beginning with *The Authoritarian Personality* (Adorno et al. 1950), a central argument in this literature has been that individuals' political ideologies and behaviors are partly rooted in chronic personality traits (Jost, Federico, and Napier 2009). Among the most robust results in this work is the finding that more conservative individuals typically exhibit a cluster of traits reflecting greater orientations towards certainty. Classic studies show that, compared to liberals, conservatives have a preference for reasoning that is dichotomous or based on clear categories to qualified or probabilistic reasoning, a greater tendency to experience threat or anxiety when faced with uncertainty, a lower desire for new experiences, and a higher desire to quickly reach firm conclusions quickly (review in Jost et al. 2003a). The *uncertainty-threat hypothesis* (Jost et al. 2003a) proposes that the common thread uniting these findings is differences in responses to unknown, uncertain, or threatening situations, which we refer to as "preference for certainty."

Social-scientific treatments have frequently identified traditionalism and opposition to change as fundamental aspects of conservative ideology (e.g., Huntington 1957; Jost 2006). Both aspects appear related to preferences for a more stable, certain, and familiar world. In contrast, liberalism is associated with a more positive view of change. For this reason, the uncertainty-threat hypothesis predicts that individuals with stronger preferences for certainty should tend towards conservatism over liberalism. This hypothesis has found strong and consistent support across 50 years of research (Jost et al. 2003a).

Uncertainty and Identity

Another line of research suggests that individuals on the ideological extremes, both left and right, show stronger preferences for certainty than more moderate individuals. This view of the political "true believer" (Hoffer 1951) suggests that the motivational needs of managing uncertainty and threat are addressed through rigid adherence to extreme ideologies (Greenberg and Jonas 2003; Hogg 2007). Evidence that certainty preference occurs on either political extreme can be found, for example, in studies of supporters of communism in formerly communist countries (Greenberg and Jonas 2003; Tetlock and Boettger 1989).

According to *uncertainty-identity theory* (Hogg 2007), group identification reduces uncertainty by providing individuals a clear sense of self and prescriptions for behavior based on prototypical group characteristics. Uncertainty was found to increase the strength of party identification among both conservatives and liberals (Hohman, Hogg, and Bligh 2010). Since more extreme groups provide greater contrast between members and non-members and thus clearer behavioral prototypes and membership criteria (Hogg 2004), uncertainty-identity theory predicts that individuals with greater needs for certainty may be drawn to more extreme ideologies. Consistent with this, individuals have been shown to identify with more extreme ideological groups when their level of uncertainty was experimentally increased (Hogg 2004). This is also consistent with the notion that uncertain economic times often coincide with the rise of extreme ideologies. This mechanism could operate at the same time as the one proposed by the uncertainty-threat hypothesis, and a mixed model of the two has found some empirical support (Hogg 2007; Jost et al. 2003b).

From Motivation to Action

We expect that ideological groups whose members hold greater preferences for certainty will exhibit greater levels of homophily. Homophilous contact can confirm worldviews and reinforce ideologies, while heterophilous contact threatens to seed uncertainty and doubt. Thus, it stands to reason that those seeking greater certainty should do so in part via political homophily.

Past research supports this reasoning. Heightened desire for cognitive closure is associated with homophilous preferences such as favoritism for members of one's ethnicity and greater identification with partners in ad-hoc groups (Shah, Kruglanski, and Thompson 1998), and people with higher sensitivity to threat hold more hostile attitudes towards out-groups (Hatemi et al. 2013). The desire for heterophilous contact, on the other hand, is associated with traits typical of low desires for certainty, such as sensation seeking and openness to experience (Mehrabian 1975; Gerber et al. 2012). Past research also confirms that heightened uncertainty leads to a greater affinity for groups of homogenous, similar others (Jetten, Hogg, and Mullin 2000).

Summary of Claims

We argue that individuals higher in preferences for certainty will seek social confirmation and avoid disagreement, making them more likely to form homophilous ties. Drawing on the research reviewed above, we propose two hypotheses:

> **H1**: Ego networks on the ideological right will exhibit greater political homophily than those on the left.
>
> **H2**: Ego networks on the ideological extremes will exhibit greater political homophily than those at the center.

Measuring Political Homophily

Our investigation of political homophily builds on a long research tradition. Early sociometric surveys provided evidence of the political homogeneity of core networks by asking respondents to name and describe their closest contacts (Laumann 1969; Knoke 1990). These “strong-tie” surveys could not speak to the homogeneity of broader ego networks, as stronger ties are markedly more homogeneous than weaker ones (Granovetter 1973). Evidence of political homogeneity in broad acquaintanceship networks came from a recent General Social Survey (2006), which specifically measured both weaker and stronger ties (DiPrete et al. 2011).

However, as DiPrete and colleagues point out, measures derived solely from respondents’ descriptions of alters capture only *perceived* homophily, which may greatly exaggerate its true levels. For example, studies that interviewed both respondents and their alters found that respondents frequently overestimated their political similarity (Goel, Mason, and Watts 2010; Huckfeldt et al. 1995; Laumann 1969). Out of seven respondent-provided alter characteristics that Laumann (1969) verified via interviews with the alter, party identification was the least accurate, with reported and true identification correlated at $r = .51$. Moreover, the rate of mistakes was correlated with ideology, creating a potentially problematic confound.

Another empirical challenge comes from the difficulty of distinguishing between network homogeneity produced by homophilous tendency—“homophily proper” (Wimmer and Lewis 2010) or “choice homophily”—and homogeneity due to other mechanisms. If groups of potential homophilous partners differ in size, random tie creation would lead the majority group to have more homogeneous ties than the minority group, even without any homophilous tendency (Blau 1977; Feld 1982). This kind of "baseline" homophily (McPherson and Smith-Lovin 1987) is difficult to rule out with survey data, as the availability of potential homophilous partners in a social environment is generally unknown. The uneven geographic concentration of Democrats and Republicans suggests that this problem is relevant to political homophily. While studies of complete face-to-face networks within bounded settings can estimate such “baseline” rates with relative ease, their homogeneity and small scale makes observation of political homophily difficult. For example, in a fine-grained study of networks between master’s students in a public policy school, Lazer and colleagues (2010) did not find evidence of significant homophily on the basis of either politics or gender, attributing this lack of political homophily to an artifact of their demographically homogeneous sample.

Thus, measuring political homophily involves three major difficulties. First, to measure discrepancies from baseline levels of homogeneity expected from random mixing, the relative availability of potential homophilous partners must be known. Second, information on alters’ political orientations should be drawn from sources other than the ego’s report. And finally, the network data should cover a broad sample of respondents, and a range of alters beyond the closest "strong-tie" core. To our knowledge, no published work meets all three criteria.

We also know of no work that examines whether rates of political homophily differ across the political spectrum in interpersonal networks. Such difference was, however, noted in an innovative study of political blogs, with the weblink structure between conservative blogs appearing denser than between liberal ones (Adamic and Glance 2005). Though a follow-up re-analysis of the data failed to replicate this finding (Ackland and Shorish 2009), the results still pose a provocative question about possible asymmetries in rates of political homophily. Barberá's (2015) finding that conservative Twitter users forward (or "retweet") messages from other conservative posters at greater rates than liberals retweet messages from other liberals similarly points towards this possibility.

METHOD

To test our claims regarding the relationship between political ideology and levels of political homophily, we examined the Twitter networks of roughly a quarter million politically-engaged Americans. Using a procedure recently validated by Golbeck and Hansen (2014), we located these individuals by identifying the Twitter accounts of major U.S. political actors with previously measured political orientations (159 congresspeople and 33 policy non-profits). We used these as a proxy for the orientations of their followers. We then calculated homophily measures for the ego networks of these followers, and analyzed the resulting dataset via multivariate regression with cluster-adjusted standard errors.

Research Site

Twitter is both a social networking service and media platform. Users post short messages (called "tweets") to their profile. Immediately, everyone subscribing to their account (their "followers") receives copies of those messages. About 90% of all Twitter accounts are public, meaning that anyone can subscribe (or "follow") them, view their posts, or examine their ego networks (Takhteyev, Gruzd, and Wellman 2012)[2]. The entire stream of public tweets can also be searched by keyword, allowing users to locate accounts that interest them. In contrast to offline networks, where the choice of partners is often highly restricted by geography, competent Twitter users who wish to create new homophilous ties can thus do so with ease and on a practically limitless scale. The resulting network is composed of directed and often asymmetric ties of *attention*, and so features high-degree "hub" nodes belonging to major journalists, celebrities, politicians, and other popular content producers. Such hubs form the basis of our sampling strategy.

Between 2010 and 2012, the percentage of adult Americans using Twitter increased from 5% to 12% (Rainie 2010; Smith and Rainie 2010; Smith and Brenner 2012).This broad and quickly growing user base, combined with the unparalleled observability of online social activity, make services like Twitter a valuable resource for social research. However, these data also introduce some important limitations. Like most large complete-network datasets, our dataset is a single cross-sectional snapshot, precluding many approaches to causal inference. Additionally, we lack demographic covariates for our sample. We thus cannot rule out the possibility that homophily on an unobserved trait is responsible for the homogeneity of ties we observe. Furthermore, our sample is not representative of Americans: the Twitter user base is younger, more female, more educated, higher income, and features higher rates of racial and ethnic minorities than the overall population (Smith and Rainie 2010). The higher average education of Twitter users in particular might make them more opinionated and thus more

[2] From this point, we use "ego network" to refer to the set of the ego's Twitter ties, the users those ties point to, and the sets of ties belonging to those users.

politically homophilous than the American public. Our analysis is therefore best viewed as an unusually large and diverse case study rather than a snapshot of the American electorate, leaving open the possibility that the effects we observe are limited to this self-selected, albeit large, population.

On the other hand, Twitter data have important advantages relevant to the methodological challenges detailed above. Since our dataset contains all public Twitter accounts, we can calculate the total number of potential homophilous partners for any given user, which in turn allows us to control for the baseline homophily rates we would observe under random mixing. Equally important, Twitter network data derive from observation rather than self-report, thus avoiding the well-documented pro-homophily bias faced by most survey studies. Finally, while a user's Twitter ego network is by no means the same as their offline ego network, its size and geographical distribution are suggestive of a broad mixture of online and offline contacts as well as stronger and weaker ties (see Takhteyev et al. [2012] on Twitter geography). Thus, while Twitter data bring unfamiliar challenges, they also solve many familiar problems, making them a valuable complement to more traditional data.

While not directly generalizable to offline populations, there are nonetheless good reasons to study Twitter for insight into U.S. political networks. First, Twitter is a significant political communication platform in its own right, as evidenced by the range of major political actors who use it. Twitter use is ubiquitous among U.S. social movement organizations. (Obar, Zube, and Lampe 2012), who use it to disseminate information and mobilize collective action. As of this writing, virtually all congresspeople have Twitter presences (Hemphill, Otterbacher, and Shapiro 2013). Twitter users' attention to these politicians tracks offline behavior: e.g., the volume of Twitter mentions of a congressional candidate predicts her electoral performance, even net of key covariates including incumbency and media coverage (DiGrazia et al. 2013).

Second, studies demonstrate that Twitter networks share many properties and processes with offline phenomena. For example, Dunbar and colleagues show that ego networks on both Twitter and Facebook have strikingly similar distributions of degree and tie strength to offline networks, leading them to conclude that “the structure of online social networks mirrors those in the offline world” (Dunbar et al. 2015:39). Geographic distances, national borders, and frequency of air flights also affect ties in ways that resemble networks offline (Takhteyev et al. 2012). In their study of social movements on Twitter, González-Bailón and colleagues (2011) find evidence that both protest recruitment and informational diffusion occur over Twitter ties. They also find that online political behavior diffusion is consistent with the same “complex contagion” dynamics (Centola and Macy 2007) thought to describe the diffusion of behavior in offline networks, as do Romero, Meeder, and Kleinberg (2011).

The parallels between Twitter and offline electoral politics are also illustrated by Barberá (2015), who shows that the ideological positions of members of the 112th Congress can be estimated solely from Twitter ties among their followers. Barberá treats shared followers similarly to how roll-call ideal-point scaling techniques interpret shared votes. The resulting estimates nearly perfectly recreate roll-call measures of the politicians’ ideological positions ($r > 0.94$), yielding a distribution of ideal points for ordinary users that approximates this distribution offline. They also closely track survey and demographic measures of citizen ideology when agglomerated at state level ($r > 0.87$).

Sample Selection

To create our sample, we first searched Twitter for members of the 111th U.S. Congress, locating 31 active accounts belonging to senators and 128 belonging to representatives (30% of both chambers).[3] For robustness, we also gathered a sample of U.S. policy non-profits. Our search for 50 such organizations most frequently cited in major U.S. media (Groseclose and Milyo 2005) produced 33 accounts, consisting of think tanks such as the RAND Corporation and policy groups such as the Sierra Club.

Research shows that the perceived partisanship of news media has a strong effect on who consumes it, with audiences generally preferring news media that is consistent with their views (Iyengar and Hahn 2009; Stroud 2008). Similarly, we expect that, on average, we can infer the political orientation of a user from the ideological positions of the hubs they follow. Golbeck and Hansen (2014) validated this approach by examining the Twitter postings of users who follow congresspeople, finding that the ideological scores of politicians reliably predicted their Twitter followers' presidential election vote choices and preferences for ideological news. As we discussed above, Barberá (2015) also showed that the Twitter tie structure between followers of congresspeople closely reflects the relative ideological positions of these politicians.[4] Thus, our hypotheses suggest that audiences of more conservative or extreme political hubs may follow one another at greater rates than those of more liberal or moderate ones.

[3] Like Golbeck and Hansen (2014), we excluded John McCain, as his recent candidacy for president gave him a categorically different Twitter presence. We also dropped hubs with less than 100 followers, since they were not prominent enough to be properly considered hubs.

[4] The validity of using ideological positions of legislators to proxy those of their constituents has been the subject of a number of recent critiques, which point out that legislators tend to be more ideologically extreme than members of the general public, and that many non-ideological factors affect electoral outcomes (e.g., Bafumi and Herron 2010; Enns and Koch 2013). However, we note that, while any individual has little control over who represents her in Congress, she can choose whether or not to follow any congressperson on Twitter, and can follow any number of congresspeople she wishes. This greatly increased freedom of choice sets Twitter-based measures validated by these studies apart from those criticized in the literature.

Data

Our primary data come from a publicly available Twitter dataset created by Kwak, Lee, Park and Moon (2010), which contains a complete snapshot of the publicly visible Twitter network from June 2009 (over 40 million nodes and 1.47 billion ties). The dataset consists of only the network structure itself, with no information about the nodes beyond their Twitter account numbers. We linked our hubs to their offline identities via data retrieved from Twitter servers, and calculated all network measures via custom *MySQL* routines.

We use archival data from 2009 because it crucially pre-dates Twitter's "Who to Follow" feature. Since July 2010, this feature has encouraged Twitter users to follow the same accounts as their alters, thus nudging them towards greater homophily. As of May 2013, this feature was responsible for the creation of over a million Twitter ties *per day* (Gupta et al. 2013), rendering Twitter data gathered after 2010 less suitable for studying homophily.

We utilize a number of further datasets for information on political hubs in our sample. For members of Congress, we use the Congressional Committee Assignments dataset (Stewart and Woon 2011) and election results from the CQ Press Voting and Elections Collection (2010). For their constituencies, we use state- and congressional district-level information from the U.S. Census Bureau's American Community Survey (2009), Current Population Survey (2009), and the decennial census (2000). For non-profits, we use their publicly available 2010/2011 tax returns filed with the Internal Revenue Service (IRS Form 990), and background data from *GuideStar* non-profit reports (2015).

Measures

<u>Political Orientation</u>. For our measure of congressperson ideology, we utilize DW-Nominate scores computed from voting rolls for the 111th U.S. Congress (Poole and Rosenthal 2007;

Carrol et al. 2011). The primary dimension of these scores captures most of the variance and closely corresponds to the liberal-conservative dimension in US politics (Poole and Rosenthal 2007), ranging from roughly -1 (liberal) to 1 (conservative). We divide this dimension by its standard deviation.

We use the ideological scores of congresspeople as proxies for the orientations of their followers. Like Golbeck and Hansen (2014), for the 31% of users in this sample who follow two or more hubs, we average these hubs' scores. Since we have separate hypotheses concerning Left-Right and Center-Extremes differences, we decompose the scaled score into its magnitude and direction components:

$$score = |score| * (-1)^{\mathbb{1}(score>0)} = magnitude * (-1)^{direction}$$

We define the *ideological extremity* variable as the magnitude of the ideological score, which is equal to $|score|$. It captures how far a given politician is from the ideological center. We then define the *ideological conservatism* variable as the direction component of the score, which is $\mathbb{1}(score > 0)$. It represents whether the hub is conservative (1) or liberal (0).

For policy non-profits, we use ideological scores which Groseclose and Milyo (2005) computed by counting the number of times each non-profit was cited in floor speeches by members of the 103rd through 107th Congresses, and then averaging the ADA ratings of the citing members. We linearly translate this measure to range from -1 to 1, thus placing it on roughly the same scale as DW-Nominate scores to make results visually comparable across both datasets[5]. We then follow the above procedures to create our conservatism and extremity measures.

[5] The two scores, however, remain distinct enough that we do not combine the populations in any of our regression analyses.

Homophily. Twitter is both a social networking platform and a media service (Kwak et al. 2010). Its network structure consists of ties of attention. Some ties are interpersonal, connecting regular users with one another, and are thus akin to ties of friendship or acquaintanceship. Others connect regular users to hubs belonging to major public figures or organizations, and thus more closely resemble the connections between audience members and celebrities or media sources. This makes the observed Twitter network essentially two-mode. Though the gradient-like nature of Twitter popularity sometimes makes these modes difficult to distinguish in practice, we can imagine setting aside hub nodes and examining the interpersonal (user–user) network apart from the audience (user–hub) network.

Our homophily measure follows this reasoning. For any follower *f* of political hub *H*, we define her "homophily with regard to H" as the percentage of other accounts followed by *f* that in turn also follow *H.* A user's homophily with regard to a political figure they follow is thus equal to the percentage of other users they follow who also follow this figure (see Figure 1). If V is the set of all public nodes on Twitter and $T_{ij} = 1$ if directed tie i→j exists and 0 otherwise, this equals

$$\boldsymbol{o}(f, H) = \frac{\sum_{g \in V} T_{fg} T_{gH}}{\sum_{g \in V} T_{fg} - 1} * 100\%$$

This measure ranges from 0 to 100%. As this measure is undefined for nodes with no outgoing ties except the tie to hub H, we drop such nodes ($N = 2{,}665$) from our sample.

[Figure 1 about here]

This homophily measure has two advantages. The first is conceptual. If hub *H* posts a political message on Twitter, 50% of *f*'s alters would also receive a copy of this message. Since users frequently forward the messages they receive to their own followers (Kwak et al. 2010) this measure captures the *ceteris paribus* chances of *f* receiving additional exposures to *H*'s message, which fits with the conception of homophilous ties as building blocks of echo chambers. Recent analyses of diffusion on Twitter confirm that such multiple exposures promote the spread of political messages, as expected from diffusions which spread through normative influence (Romero et al. 2011). Thus, $\boldsymbol{o}(f, H)$ should capture a property of network structure that directly corresponds to its ability to sustain normative diffusions.

The second advantage of $\boldsymbol{o}(f, H)$ is methodological. We can enumerate all of the followers of hub *H* to determine the exact count of *f*'s *potential* homophilous partners. This gives us the rare opportunity to statistically control for differences in the baseline rates of network homogeneity expected under random mixing, which are different for populations of different sizes.

<u>Reciprocity.</u> The measure $\boldsymbol{o}(f, H)$ has a potential to confound homophily with reciprocity. Imagine the following scenario: followers of hub A are as likely as followers of hub B to initiate new homophilous ties with one another; however, followers of B are more likely than followers of A to reciprocate homophilous ties once they are initiated by another follower. As a result, followers of B have higher average homophily scores than followers of A in spite of not actually being more homophilous in the sense we use here. To deal with this confound, we introduce a reciprocity measure to capture an actor's tendency to reciprocate incoming homophilous ties from other hub followers. For each follower *f* of hub *H*, this measure equals the count of ties received by *f* from other followers of *H* that *f* reciprocates by sending a tie back, divided by the total count of such ties received by *f* (see Figure 2):

$$\boldsymbol{r}(f, H) = \frac{\sum_{g \in V} T_{gH} T_{fg} T_{gf}}{\sum_{g \in V} T_{gH} T_{gf}}$$

We standardize this measure, and set it to 0 for all nodes with no incoming homophilous ties.

[Figure 2 about here]

Characteristics of Congresspeople. Congresspeople differ not only in their ideological orientation, but also in their prominence, characteristics of their constituencies, and various other ways that may influence the rates at which their followers follow one another. We address these potential confounds with several controls. First, we expect that constituents make up a non-negligible portion of each congressperson's followers. Since geographical proximity significantly increases the probability of a social tie between two people both offline (Butts 2003) and on Twitter (Takhteyev, Gruzd, and Wellman 2012), constituencies with denser populations may yield more densely connected followers. Thus, we add covariates for the total population of the constituency, as well as its density and rate of internet usage[6].

Some congresspeople also have greater national profiles than others. While less prominent members may be known largely within their constituencies, more prominent ones may have more diverse and geographically dispersed audiences with correspondingly lower tie densities. To control for these differences in prominence, we add terms for chamber ("senator"), leadership ("chamber/party" and "committee"), and number of years in chamber ("seniority").

Highly motivated or mobilized followers may also follow one another at heightened rates. In June 2009, the emerging Tea Party movement could have had this effect on some conservatives. This mobilization also resulted in the creation of Tea Party caucuses, formed during the 111th Congress in the House and at the start of the 112th Congress in the Senate. Since the Tea Party itself has no clear member rolls or membership criteria, we use memberships in these caucuses to control for Tea Party affiliation in our model. To further control for differences

[6] To estimate district-level internet usage from 2009 state-level data, we use the Census geography files to calculate each district's rural and urban populations, and then use these to take weighted averages of the state's rural and urban internet usage rates.

in mobilization and the density of supporters within the constituency, we also add a covariate for victory margin from the congressperson's last election.

Characteristics of Non-profits. Like members of congress, non-profits vary in ways which may affect ties between their followers. First, since followers of less prominent non-profits may be a more specialized audience those of more prominent non-profits, they may also form ties to one another at higher rates. To control for this confound, we add covariates for total budget and age of the non-profit. Conversely, each non-profit's staff likely forms a portion of its online audience, and this portion may be higher for organizations with larger staffs. Since shared workplaces are a likely basis for social tie formation (Feld 1982), such non-profits may appear to have more homophilous followers than smaller organizations. To address this, we control for each organization's employee and volunteer counts[7].

While some of the non-profits are single independent organizations, others are affiliated with separately-incorporated regional or special-purpose organizations. For example, the National Association for the Advancement of Colored People's (NAACP) name is shared by many regional chapters of the NAACP itself (e.g., Los Angeles NAACP), by a series of day-care centers, and by other specialized affiliates. Organizations with a more dispersed geography and mission could have more diverse and therefore less homophilous audiences. We thus add a dummy term for such affiliates[8]. Conversely, Washington, D.C.'s large and active non-profit sector may grant Washington-based non-profits a densely connected audience, and especially so if they engage in substantial lobbying activity. We therefore also control for whether the non-

[7] To compensate for the substantial positive skew in the budget and volunteer counts, we take the logarithm of both variables.
[8] We searched for such affiliates using the Foundation Center's database of IRS records (990finder.foundationcenter.org).

profit is headquartered in Washington, and for whether its tax status allows unrestricted lobbying activity[9].

RESULTS

We located 31 hubs belonging to senators, 128 hubs belonging to House members and 33 hubs belonging to policy non-profits, yielding a total of 238,943 unique followers (see Table 1). These followers were closely divided between liberal and conservative hubs, with 131,576 unique ego networks connected to liberal and 133,210 connected to conservative hubs. But, while the total number of conservatives and liberals in the sample was similar, they followed different categories of hubs. There were more followers of Republican than Democratic members of Congress, but more followers of liberal than conservative non-profits. Since our substantive findings do not differ between congressional and non-profit hubs, these different distributions do not appear to affect our results.

[Table 1 about here]

Across all the users in both samples, we find an average homophily rate $\boldsymbol{o}(f, H)$ of 11.0%. Therefore roughly 1 out of 9 alters of each user in our sample follows the same political figure as that user. As we demonstrate in Appendix A, this rate is substantially higher than would be expected under random mixing, and is thus consistent with the presence of political homophily in the sample.

[9] All organizations in our sample are 501(c)(3) non-profits, which can take unlimited tax-deductible donations but can only engage in limited lobbying. Roughly half are also separately incorporated as 501(c)(4)/(6) non-profits, which cannot accept such donations but are instead allowed unlimited lobbying. For such dually incorporated non-profits, which file two tax returns, we use the sum of their 501(c)(3) and 501(c)(4)/(6) budgets as their total budget. Though these organizations are separately incorporated, they often share the same physical location and the same employees. Thus, instead of summing the staff counts, we use the greater of the two numbers.

Followers of Members of Congress

We employ multivariate linear regression with two-way clustered standard errors to test our hypotheses about how political homophily varies by ideology. Our main dependent variable is the homophily measure $\boldsymbol{o}(f, H)$. Our primary independent variables are ideological conservatism and ideological extremity. Because the ideologies of congresspeople and non-profits are captured with different measures, we analyze their followers separately.

We control for chamber with the "senator" dummy. We also control for our reciprocity measure $\boldsymbol{r}(f, H)$. To account for the levels of homophily expected from baseline mixing in differently-sized populations, we add the covariate *hub network size*, which is the count of the hubs' followers minus one. Since random mixing produces baseline homophily rates that are a linear function of the number of potential homophilous partners (see Appendix A), we can use this quantity to control for baseline mixing in a regression framework. There are also potential differences between new versus more experienced users of Twitter. Since Twitter account identifiers were assigned sequentially and the growth of Twitter's user base was roughly exponential through 2009, we expect the age of a Twitter account to be roughly proportional to the logged count of accounts with greater identifiers: $age\left(id_f\right) \propto \log\left(max\ id - id_f\right)$. We use this to create controls for *account age* and *hub age*[10]. We also include a covariate for *ego network size*. We standardize all non-dummy controls.

Since 31% of the users in our data follow multiple hubs, and each follower f has a different value of $\boldsymbol{o}(f, H)$ for each hub she follows, our unit of analysis is the user-hub pairing. There are two non-nested levels of non-independence in these data. Some variables are non-

[10] While it is possible to retrieve the exact account creation dates from the Twitter API, the number of nodes in our sample would make that approach extremely costly.

independent for all followers of a hub, and others for all observations of a follower. Without accounting for this non-independence, our standard errors could be significantly underestimated, as the true amount of independent data points for our primary independent variables is closer to the number of hubs ($N = 159$) than it is to the number of observations ($N = 383{,}292$). We employ multi-way clustered standard errors to prevent this underestimation (Petersen 2009; Thompson 2011), clustering by hub and follower. This accounts for non-independence, but also results in significance levels heavily constrained by the number of hubs.

The results of this analysis are in Model 1 of Table 2. They show that ideological conservatism was strongly and positively associated with homophily rate ($\beta = 3.80, p < 0.001$). Thus, with all other variables held constant, conservatives' homophily rates were 3.8 points higher than those of liberals, offering support for our first hypothesis. We also found that ideological extremity has a strong positive association with homophily ($\beta = 4.53, p < 0.001$). indicating that a one standard deviation increase in extremity was associated with a 4.5 point increase in homophily. Since our average observed homophily rate is 11.2%, these differences are both statistically and substantively significant, thus lending support to our hypotheses.

[Table 2 about here]

We next repeated this analysis with the addition of further controls for characteristics of congresspeople and constituencies we described above. Results are in Model 2 of Table 2. Consistent with the above conjectures, we found a marginal positive effect for Tea Party membership ($\beta = 1.85, p < 0.10$) and a weak but significant positive effect for margin of election victory ($\beta = 0.32, p < 0.01$). On the other hand, the previously-significant Senator dummy decreased in magnitude and significance ($\beta = -1.24, n.s.$), indicating that these covariates may partially account for the lower homophily rates among followers of senators.

Both ideological conservatism and ideological extremity decreased slightly in magnitude, but remained highly significant ($\beta = 3.25, p < 0.001$ and $\beta = 4.06, p < 0.001$, respectively), again supporting our hypotheses.

Followers of Non-Profits

The analysis of Twitter followers of members of Congress supports both of our hypotheses. However, the fact that liberal and conservative hubs correspond to two different political parties leaves open the possibility of top-down, organizational explanations of our results: for example, these asymmetries could be driven by differences in the internet strategies pursued by the two parties. To address these concerns, we replicated our analyses with followers of major policy non-profits, which are further removed from the organizational structure of political parties and the social environment of electoral politics.

Results are given in Table 3. Model 1 replicates our initial test of ideological conservatism and extremity effects. Due to the much smaller hub-level population ($N = 33$), cluster-adjusted standard errors are larger than for congressional hubs, reducing the statistical significance of most coefficients. Nonetheless, we find that ideological conservatism is still significantly associated with higher homophily ($\beta = 7.46, p < 0.01$), an effect of greater size than for congressional hubs. Ideological extremity is also still significantly associated with higher homophily at roughly the same magnitude of effect as in the previous sample ($\beta = 4.32, p < 0.01$).

[Table 3 about here]

With additional controls for further characteristics of the non-profits (Model 2 of Table 3), the effect of conservatism decreased slightly in magnitude but increased in significance ($\beta =$

$6.31, p < 0.001$). The effect of extremity also increased in significance ($\beta = 4.33, p < 0.001$). The analysis of followers of non-profits hubs thus provides additional support for our hypothesis.

The network plot in Figure 3 illustrates these homophily differences. It depicts ties between the followers of the Cato Institute (conservative) and Amnesty International (liberal), two well-known non-profits with comparable numbers of followers (10,298 and 10,638). Their mean follower homophily rates are 13.04% and 6.75%, respectively, a difference similar in magnitude to the conservatism effect estimated by our models. The force-directed layout places connected nodes and disconnected ones further apart, leading the cluster of Cato Institute's followers to appear visibly smaller because of its higher homophily.

[Figure 3 about here]

We performed additional analyses with both sample populations to further establish the significance and robustness of our results (see Appendix B). First, we investigated whether the observed homophily differences extend to stronger ties, which may play a more important role than weaker ties for diffusing social influence (Bond et al. 2012; González-Bailón et al. 2011; McAdam and Paulsen 1993). These analyses confirmed that the same substantive results obtain when our analyses are restricted to stronger ties. To ensure the robustness of our findings to modeling assumptions, we also examined these homophily differences using randomization inference and propensity score matching, analyses which also supported both hypotheses. Additionally, we leveraged the known Twitter account age of all individuals in our sample to test for a possible reverse causal relationship between homophily and ideological extremity (see Appendix C). Results indicated that such a relationship is unlikely to explain the results we report here.

DISCUSSION

We drew on research from political psychology in hypothesizing that more conservative or ideologically extreme individuals would be more homophilous than more liberal or moderate ones. To avoid methodological challenges faced by most survey studies of political homophily, we tested these hypotheses with 238,943 ego networks belonging to followers of political figures on Twitter. These analyses yielded consistent and robust evidence for both hypotheses.

If the same homophily differences we observed with these Twitter networks extend to social networks more generally, the higher homophily rates of more conservative and ideologically extreme individuals could have significant consequences for the emergent dynamics of their respective political networks. These rates should, all things being equal, result in networks that embed their members in denser webs of like-minded associations, which could then insulate individuals from the demotivating effects of dissenting views, and may enable political behaviors to spread faster than they would through sparser networks. Our results thus suggest that homophily might provide a structural advantage to the mobilization of right wing or politically extreme social movements relative to left-wing or moderate ones. We would similarly expect the negative effects of network homogeneity on tolerance and understanding to be unevenly distributed.

While the status of non-elite political polarization in the U.S. is still debated (e.g., Fiorina and Abrams 2008), we note that our findings of left-right differences in homophily fit with accounts that find polarization to be asymmetrically tilted towards the political right (e.g., Butler 2009). Moreover, they recall Sunstein’s (2002) warning that social networks may amplify polarization: if homophily is found at higher levels at the extremes of the ideological distribution, the resulting concentration of homogenizing and mobilizing influence could push extreme attitudes ever further away from the center.

REFERENCES

Ackland, Robert, and Jamsheed Shorish. (2009). "Network Formation in the Political Blogosphere: An Application of Agent Based Simulation and E-Research Tools." *Computational Economics* 34(4):383–98.

Adamic, Lada, and Natalie Glance. (2005). "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog." Pp. 36–43 In *Proceedings of the 3rd International Workshop on Link Discovery*, *LinkKDD '05*..

Adorno, Theodor, Else Frenkel-Brunswik, Daniel Levinson, and R. Nevitt Sanford. (1950). *The Authoritarian Personality*. Oxford, England: Harpers.

Arendt, Hannah. (1961). *Between Past and Future Eight Exercises in Political Thought*. Revised edition, 1968. NY: The Viking Press.

Bafumi, Joseph, and Michael Herron. (2010). "Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress." *American Political Science Review* 104(03):519–42.

Barberá, Pablo. (2015). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23(1):76–91.

Blau, Peter. (1977). *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. Free Press.

Bond, Robert M. et al. (2012). "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489 (7415):295–98.

Butler, Daniel M. (2009). "The Effect of the Size of Voting Blocs on Incumbents' Roll-Call Voting and the Asymmetric Polarization of Congress." *Legislative Studies Quarterly* 34(3):297–318.

Butts, Carter. (2003). “Predictability of Large-Scale Spatially Embedded Networks.” In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, edited by Ronald Breiger, Kathleen Carley, and Philippa Pattison. National Academies Press.

Campbell, David. (2013). “Social Networks and Political Participation.” *Annual Review of Political Science* 16(1):33–48.

Carrol, Royce et al. (2011). “DW-NOMINATE Scores With Bootstrapped Standard Errors.” Retrieved May 9, 2011 (http://www.voteview.com/dwnomin.htm).

Centola, Damon. (2010). “The Spread of Behavior in an Online Social Network Experiment.” *Science* 329(5996):1194.

Centola, Damon, and Michael Macy. (2007). “Complex Contagions and the Weakness of Long Ties.” *American Journal of Sociology* 113(3):702–34.

CQ Press. (2010). “CQ Press Voting and Elections Collection.”

DiPrete, Thomas, Andrew Gelman, Tyler McCormick, Julien Teitler, and Tian Zheng. (2011). “Segregation in Social Networks Based on Acquaintanceship and Trust.” *The American Journal of Sociology* 116(4):1234–83.

Enns, Peter, and Julianna Koch. (2013). “Public Opinion in the U.S. States 1956 to 2010.” *State Politics & Policy Quarterly* 13(3):349–72.

Dunbar, R.I.M., Valerio Arnaboldi, Marco Conti, and Andrea Passarella. 2015. “The Structure of Online Social Networks Mirrors Those in the Offline World.” *Social Networks* 43:39–47.

Eveland, William, and Myiah Hutchens Hively. (2009). “Political Discussion Frequency, Network Size, and ‘Heterogeneity’ of Discussion as Predictors of Political Knowledge and Participation.” *Journal of Communication* 59(2):205–24.

Feld, Scott. (1982). “Social Structural Determinants of Similarity among Associates.” *American Sociological Review* 47(6):797–801.

Fiorina, Morris, and Samuel Abrams. (2008). “Political Polarization in the American Public.” *Annual Review of Political Science* 11(1):563–88.

Gerber, Alan, Gregory Huber, David Doherty, and Conor Dowling. (2012). “Disagreement and the Avoidance of Political Discussion: Aggregate Relationships and Differences across Personality Traits.” *American Journal of Political Science* 56(4):849–74.

Goel, Sharad, Winter Mason, and Duncan Watts. (2010). “Real and Perceived Attitude Agreement in Social Networks.” *Journal of Personality and Social Psychology* 99(4):611–21.

Golbeck, Jennifer, and Derek Hansen. (2014). “A Method for Computing Political Preference among Twitter Followers.” *Social Networks* 36:177–84.

González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. (2011). “The Dynamics of Protest Recruitment through an Online Network.” *Scientific Reports* 1.

Granovetter, Mark. (1973). “The Strength of Weak Ties.” *American Journal of Sociology* 78(6):1360–80.

Greenberg, Jeff, and Eva Jonas. (2003). “Psychological Motives and Political Orientation—The Left, the Right, and the Rigid: Comment on Jost et al.(2003).” *Psychological Bulletin* 129(3):376–82.

Groseclose, Tim, and Jeffrey Milyo. (2005). “A Measure of Media Bias.” *The Quarterly Journal of Economics* 120(4):1191–1237.

GuideStar. (2015). “Nonprofit Reports and Forms 990 for Donors, Grantmakers, and Businesses.” Retrieved June 10, 2015 (http://www.guidestar.org/).

Gupta, Pankaj (2013). “WTF: The Who to Follow Service at Twitter.” Pp. 505–14 In *Proceedings of the 22nd International Conference on World Wide Web*. Geneva, Switzerland.

Hatemi, Peter, Rose McDermott, Lindon Eaves, Kenneth Kendler, and Michael Neale. (2013). “Fear as a Disposition and an Emotional State: A Genetic and Environmental Approach to Out-Group Political Preferences.” *American Journal of Political Science* 57(2):279–93.

Hoffer, Eric. (1951). *The True Believer: Thoughts on the Nature of Mass Movements*. HarperCollins.

Hogg, Michael. (2004). “Uncertainty and Extremism: Identification with High Entitativity Groups under Conditions of Uncertainty.” In *The Psychology of Group Perception: Perceived Variability, Entitativity, and Essentialism*. Psychology Press.

———. (2007). “Uncertainty–Identity Theory.” *Advances in Experimental Social Psychology* (39):69-126

Hohman, Zachary, Michael Hogg, and Michelle Bligh. (2010). “Identity and Intergroup Leadership: Asymmetrical Political and National Identification in Response to Uncertainty.” *Self and Identity* 9(2):113–28.

Huckfeldt, Robert, Paul Allen Beck, Russell Dalton, and Jeffrey Levine. (1995). “Political Environments, Cohesive Social Groups, and the Communication of Public Opinion.” *American Journal of Political Science* 39(4):1025–54.

Huckfeldt, Robert, Jeanette Morehouse Mendez, and Tracy Osborn. (2004). "Disagreement, Ambivalence, and Engagement: The Political Consequences of Heterogeneous Networks." *Political Psychology* 25(1):65–95.

Huntington, Samuel. (1957). "Conservatism as an Ideology." *American Political Science Review* 51(2):454–73.

Iyengar, Shanto, and Kyu Hahn. (2009). "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use." *Journal of Communication* 59(1):19–39.

Jang, Seung-Jin. (2009). "Are Diverse Political Networks Always Bad for Participatory Democracy? Indifference, Alienation, and Political Disagreements." *American Politics Research* 37(5):879–98.

Jetten, Jolanda, Michael Hogg, and Barbara-Ann Mullin. (2000). "In-Group Variability and Motivation to Reduce Subjective Uncertainty." *Group Dynamics: Theory, Research, and Practice* 4(2):184–98.

Jost, John, Jack Glaser, Arie Kruglanski, and Frank Sulloway. (2003a). "Political Conservatism as Motivated Social Cognition." *Psychological Bulletin* 129(3):339–75.

———. (2003b). "Exceptions That Prove the Rule—Using a Theory of Motivated Social Cognition to Account for Ideological Incongruities and Political Anomalies: Reply to Greenberg and Jonas (2003)." *Psychological Bulletin* 129(3):383–93.

Jost, John. (2006). "The End of the End of Ideology." *American Psychologist* 61(7):651–70.

Jost, John, Christopher Federico, and Jaime Napier. (2009). "Political Ideology: Its Structure, Functions, and Elective Affinities." *Annual Review of Psychology* 60(1):307–37.

Kim, Hyojoung, and Peter Bearman. (1997). "The Structure and Dynamics of Movement Participation." *American Sociological Review* 62(1):70–93.

Klofstad, Casey, Anand Sokhey, and Scott McClurg. (2013). “Disagreeing about Disagreement: How Conflict in Social Networks Affects Political Behavior.” *American Journal of Political Science* 57(1):120–34.

Knoke, David. (1990). “Networks of Political Action: Toward Theory Construction.” *Social Forces* 68(4):1041–63.

Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. (2010). “What Is Twitter, a Social Network or a News Media?” Pp. 591–600 In *Proceedings of the 19th International Conference on World Wide Web*.

Laumann, Edward (1969). “Friends of Urban Men: An Assessment of Accuracy in Reporting Their Socioeconomic Attributes, Mutual Choice, and Attitude Agreement.” *Sociometry* 32(1):54–69.

Lazer, David, Brian Rubineau, Carol Chetkovich, Nancy Katz, and Michael Neblo. (2010). “The Coevolution of Networks and Political Attitudes.” *Political Communication* 27(3):248–74.

Liu, Tzu-Ping, Shih-chan Dai, and Chung-li Wu. (2013). “Cross-Cutting Networks and Political Participation: Lessons of the 2010 City Mayoral Elections in Taiwan.” *East Asia* 30(2):91–104.

Marsden, Peter, and Karen Campbell. (2012). “Reflections on Conceptualizing and Measuring Tie Strength.” *Social Forces* 91(1):17–23.

McAdam, Doug, and Ronnelle Paulsen. (1993). “Specifying the Relationship Between Social Ties and Activism.” *The American Journal of Sociology* 99(3):640–67.

McPherson, Miller, and Lynn Smith-Lovin. (1987). "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review* 52(3):370–79.

Mehrabian, Albert. (1975). "Affiliation as a Function of Attitude Discrepancy with Another and Arousal-seeking Tendency." *Journal of Personality* 43(4):582–92.

Mill, John Stuart. (1859). *On Liberty*.

Mutz, Diana. (2002a). "Cross-Cutting Social Networks: Testing Democratic Theory in Practice." *The American Political Science Review* 96(1):111–26.

———. (2002b). "The Consequences of Cross-Cutting Networks for Political Participation." *American Journal of Political Science* 46(4):838–55.

———. (2006). *Hearing the Other Side: Deliberative Versus Participatory Democracy*. Cambridge University Press.

Petersen, Mitchell. (2009). "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *Review of Financial Studies* 22(1):435–80.

Poole, Keith, and Howard Rosenthal. (2007). *Ideology and Congress*. Transaction Publishers.

Price, Vincent, Joseph Cappella, and Lilach Nir. (2002). "Does Disagreement Contribute to More Deliberative Opinion?" *Political Communication* 19(1):95–112.

Rainie, Lee. (2010). *Internet, Broadband and Cell Phone Statistics*. Pew Research Center.

Romero, Daniel, Brendan Meeder, and Jon Kleinberg. (2011). "Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter." Pp. 695–704 In *Proceedings of the 20th International Conference on World Wide Web*. New York, NY, USA: ACM.

Shah, James, Arie Kruglanski, and Erik Thompson. (1998). “Membership Has Its (epistemic) Rewards: Need for Closure Effects on in-Group Bias.” *Journal of Personality and Social Psychology* 75(2):383–93.

Smith, Aaron, and Joanna Brenner. (2012). *Twitter Use 2012*. Pew Research Center.

Smith, Aaron, and Lee Rainie. (2010). *8% of Online Americans Use Twitter*. Pew Research Center.

Stewart, Charles III, and Jonathan Woon. (2011). “Congressional Committee Assignments, 103rd to 112th Congresses, 1993--2011.”

Stroud, Natalie. (2008). “Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure.” *Political Behavior* 30(3):341–66.

Sunstein, Cass. (2002). “The Law of Group Polarization.” *Journal of Political Philosophy* 10(2):175–95.

Takhteyev, Yuri, Anatoliy Gruzd, and Barry Wellman. (2012). “Geography of Twitter Networks.” *Social Networks* 34(1):73–81.

Tetlock, Philip, and Richard Boettger. (1989). “Cognitive and Rhetorical Styles of Traditionalist and Reformist Soviet Politicians: A Content Analysis Study.” *Political Psychology* 10(2):209–32.

Thompson, Samuel. (2011). “Simple Formulas for Standard Errors That Cluster by Both Firm and Time.” *Journal of Financial Economics* 99(1):1–10.

United States Census Bureau. (2009). “Table 1155. Household Internet Usage by Type of Internet Connection and State: 2009.” *Current Population Survey*.

United States Census Bureau / American FactFinder. (2000). “P2: Urban and Rural.” *2000 Census*.

———. (2009). "B01003: Total Population." *2009 American Community Survey*.

Valenzuela, Sebastián, Yonghwan Kim, and Homero Gil de Zúñiga. (2012). "Social Networks That Matter: Exploring the Role of Political Discussion for Online Political Participation." *International Journal of Public Opinion Research* 24(2):163–84.

Visser, Penny, and Robert Mirabile. (2004). "Attitudes in the Social Context: The Impact of Social Network Composition on Individual-Level Attitude Strength." *Journal of Personality and Social Psychology* 87(6):779–95.

Wimmer, Andreas, and Kevin Lewis. (2010). "Beyond and Below Racial Homophily. ERG Models of a Friendship Network Documented on Facebook." *American Journal of Sociology* 116(2):583–642.

TABLES

Table 1. Descriptive Statistics

	Liberal	Conservative	Total Unique
Senator hubs	14	17	31
Unique followers of senators	53919	51272	96022
Mean ideological score	-0.856	0.984	0.153
House member hubs	48	80	128
Unique followers of House members	27846	78314	98597
Mean ideological score	-0.697	1.224	0.504
Political non-profit hubs	18	15	33
Unique followers of non-profits	63202	45399	103864
Mean ideological score	-0.902	0.953	-0.059
Total unique followers	131576	133210	238943

Note. The "Total Unique" column does not generally equal the sum of the “Liberal” and “Conservative” columns because some Twitter users follow both liberal and conservative hubs.

Table 2. Followers of Members of Congress: Effect of Ideology on Homophily

	Model 1		Model 2	
Non-independent within hubs (N=159):				
Ideological conservatism	3.802	***	3.251	***
	(0.764)		(0.646)	
Ideological extremity	4.533	***	4.057	***
	(0.701)		(0.556)	
Hub network size	3.725	***	3.138	***
	(0.774)		(0.688)	
Hub age	1.141	***	1.142	***
	(0.301)		(0.283)	
Senator	-2.171	**	-1.244	
	(0.708)		(0.966)	
Number of constituents	—		-0.118	
			(0.352)	
Const. population density	—		-0.203	
			(0.134)	
Const. Internet access	—		-0.252	
			(0.277)	
Seniority	—		-0.454	
			(0.4)	
Committee leadership	—		-0.253	
			(0.944)	
Party leadership	—		2.151	
			(1.433)	
Victory margin	—		0.32	**
			(0.1)	
Tea Party caucus	—		1.85	#
			(1.116)	
Non-independent within followers (N=160,452):				
Account age	0.129		0.234	*
	(0.108)		(0.117)	
Ego network size	-1.193	**	0.145	
	(0.389)		(0.102)	
Reciprocity	0.269	*	-1.176	**
	(0.114)		(0.382)	
(Intercept)	4.768	***	4.762	***
	(1.054)		(0.93)	
R^2	0.161		0.168	

Note. Results from linear regressions with two-way clustered standard errors (in parentheses). All non-dummy predictors have $\sigma = 1$. $\#\ p < 0.10, *\ p < 0.05, **\ p < 0.01, ***\ p < 0.001$ (two-tailed.)

Table 3. Followers of Policy Non-Profits: Effect of Ideology on Homophily

	Model 1	Model 2
Non-independent within hubs (N=33):		
Ideological conservatism	7.456 **	6.305 ***
	(2.416)	(1.597)
Ideological extremity	4.316 **	4.332 ***
	(1.456)	(1.105)
Hub network size	0.629	0.923
	(1.183)	(1.022)
Hub age	1.346 **	0.933 *
	(0.416)	(0.452)
Annual Budget	—	0.976 #
		(0.545)
Number of employees	—	0.408
		(0.724)
Number of volunteers	—	0.258
		(0.697)
Affiliated organizations	—	-7.665 **
		(2.707)
Years since founded	—	-0.329
		(0.633)
DC area	—	3.067 *
		(1.506)
Unrestricted lobbying	—	2.452
		(2.177)
Non-independent within followers (N=103,864):		
Account age	0.29	0.267
	(0.346)	(0.328)
Ego network size	-1.02 *	-1.04 *
	(0.518)	(0.527)
Reciprocity	1.152 **	1.103 *
	(0.388)	(0.43)
(Intercept)	3.303 *	5.188 *
	(1.367)	(2.153)
R^2	0.151	0.185

Note. Results from linear regressions with two-way clustered standard errors (in parentheses). All non-dummy predictors have $\sigma = 1$. $\# \, p < 0.10, * \, p < 0.05, ** \, p < 0.01, *** \, p < 0.001$ (two-tailed.)

FIGURES

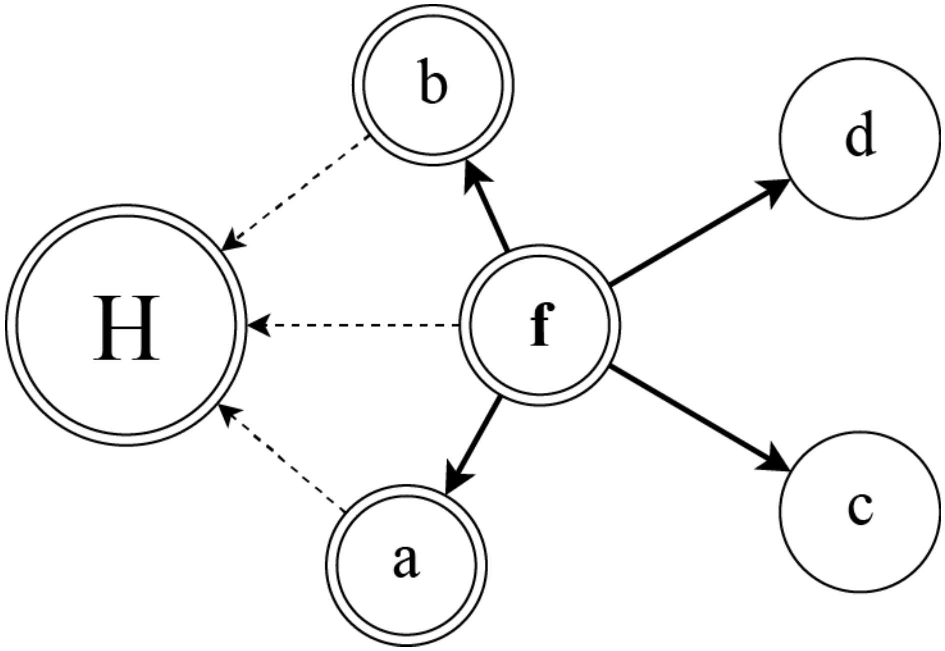

Figure 1: In addition to the tie connecting follower *f* to the hub *H*, *f* has a total of four outgoing ties ($f \rightarrow a, f \rightarrow b, f \rightarrow c, f \rightarrow d$). Since nodes *a* and *b* also follow hub *H*, while *c* and *d* do not, the homophily of *f* with regard to *H* is $\boldsymbol{o}(f, H) = 100\% * 2/4 = 50\%$.

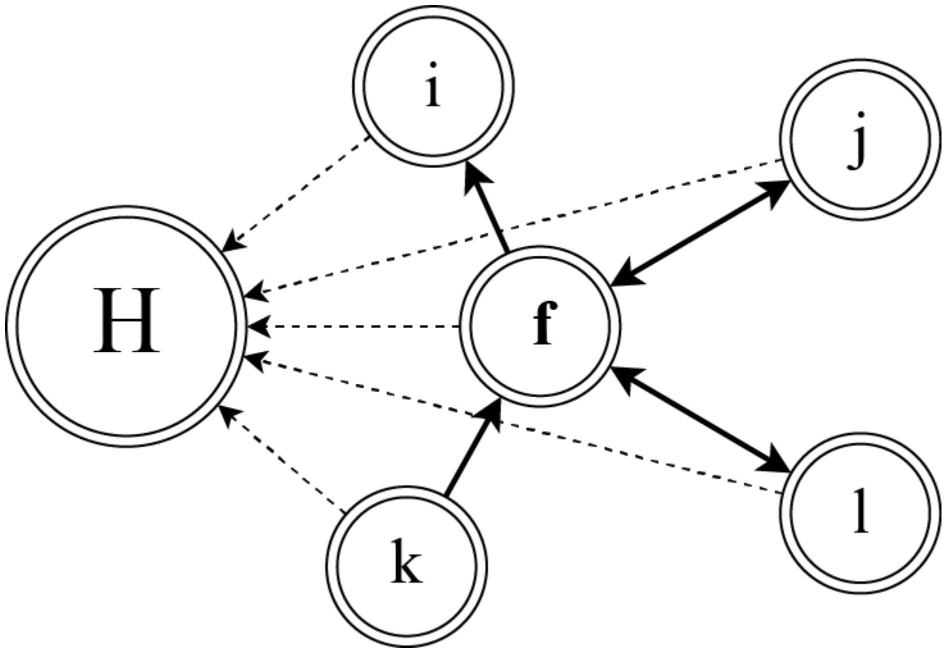

Figure 2: In addition to follower f, hub H has four followers i, j, k and l. Follower f receives homophilous ties from j, k and l, and sends reciprocal homophilous ties to j and l. The reciprocity of f with regard to H is thus $\boldsymbol{r}(f, H) = 2/3 \approx 0.67$.

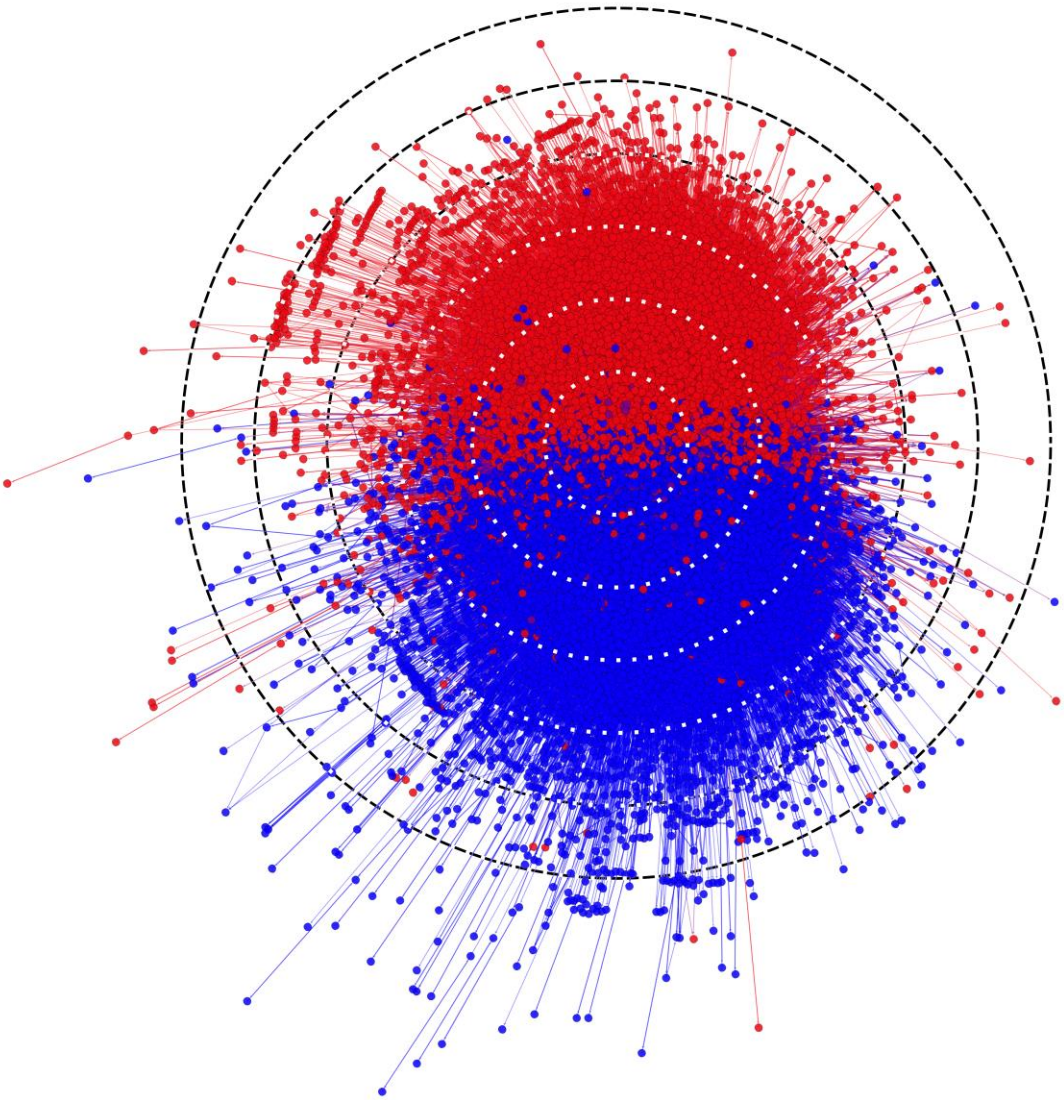

Figure 3: Network of followers of Cato Institute (conservative; red) and Amnesty International (liberal; blue) that receive at least one tie from another follower in the figure, with isolated node pairs omitted. The "spring-based" layout places nodes closer if they are connected by a tie, and further apart if they are not, revealing a clear segregation of the network into two parts as expected under conditions of strong political homophily. Though the follower counts in the two networks differ by only 3%, the Cato Institute network is far denser, resulting in a tighter clustering of nodes and thus a visibly smaller size of the densely connected component (compare edge of component against radial grid).

APPENDIXES

Appendix A: Baseline models of homophily

We use baseline models to examine whether our sample of users who follow congresspeople or policy non-profits contains evidence of homophily beyond the levels that would be expected from random mixing. The null hypothesis of random mixing states that the observed levels of homophily can be accounted for by random tie formation alone—that is, homophily rates are no greater than the proportion of potential homophilous partners in the population. The expected baseline homophily rate of each follower f of hub H is thus simply:

$$baseline(f, H) = 100\% * \frac{count\ of\ potential\ homophilous\ partners}{total\ population}$$

The count of potential homophilous partners is $\sum_{g \in V, g \neq f} T_{gH}$, which is the number of other followers of the same hub. Since the total population of Twitter at the time of data gathering was at least 41.7 million, the baseline homophily rate produced by random mixing in this population would be:

$$baseline_1(f, H) = \frac{\sum_{g \in V, g \neq f} T_{gH}}{4.17 * 10^5}$$

Averaged across all observations in our sample, $baseline_1$ produces an expected homophily rate of 0.02%. In contrast, the average observed rate of $\boldsymbol{o}(f, H)$ was 11.0%. The observed rate is thus substantially higher than the one we would expect from random mixing within this population.

One might be concerned, however, that the null hypothesis of random mixing within the entire Twitter population represents a realistic social process. Why would, for example, two Twitter users who share neither language nor interests form a tie to each other? A more convincing test would therefore account for homogeneity due to random tie formation within more restrictive populations, which would yield higher baseline estimates. Perhaps the most stringent such test comes from the null hypothesis that the observed homophily rates are produced by users in our sample randomly forming ties to only other users in our sample, though without consideration of their political orientation. Since most users in our sample have many ties to those outside the sample, this null hypothesis yields a highly conservative test. Thus, we use the count of unique Twitter users who follow U.S. members of Congress or policy non-profits (0.24 million) for a second baseline:

$$baseline_2(f, H) = \frac{\sum_{g \in V, g \neq f} T_{gH}}{2.4 * 10^3}$$

The mean expected $baseline_2$ rate is 4.24%. While larger than the previous estimated baseline, this number is still significantly lower than the observed rate of 11.0% ($t = 364.9, df = 6.2 * 10^5, p < 0.0001$). Thus, the observed political homophily rates are substantially higher than we would expect from even very conservative estimates of random mixing.

Appendix B: Additional Analyses

<u>Stronger-Tie Homophily.</u> While weaker ties are likely sufficient to diffuse information through the network, prior evidence suggests that stronger ties may play a more important role than weaker ties in political activities such as social movement recruitment and diffusion of organizational skills and other complex information (McAdam and Paulsen 1993; Tindall 2003). In a large experimental study of voter mobilization on Facebook, Bond and colleagues (2012) conclude that online diffusion of voting behavior “works because it primarily spreads through strong-tie networks that probably exist offline but have an online representation” (2012:295). González-Bailón et al. (2011) similarly found that stronger Twitter ties are responsible for diffusing social influence while weaker Twitter ties are responsible for diffusing information. Thus, if the effect of ideology on homophily was restricted to the structure of weak ties only, the

downstream consequences of this relationship might be more limited than if the effect extended to both stronger and weaker ties.

The directional nature of our data allows us to construct a simple symmetry-based measure of tie strength (Marsden and Campbell 2012). Like González-Bailón et al. (2011), we assume that unreciprocated Twitter ties are on average weaker than reciprocated ties. To examine whether the relationship between ideology and homophily persists for stronger ties, we re-calculate our homophily metric ***o****(f,H)* with non-reciprocated homophilous ties excluded. We call the resulting metric ***s****(f,H).* Since hubs do not generally reciprocate ties sent by other users (Kwak et al. 2010), this measure should also exclude most user – hub audience ties (González-Bailón et al. 2011). Thus, it has the added benefit of narrowing our analysis to interpersonal (user – user) ties.

Table B1 gives results of our original model with this revised outcome measure. Account age, which had only a weak positive effect in previous models, has increased substantially in magnitude and become highly significant ($\beta = 1.087, p < 0.001$), and ego network size, which lost its significance and switched sign in model 2 of table 2, now again has a significant but relatively weak negative effect ($\beta = -0.583, p < 0.01$). In contrast, the hub network size and hub age coefficients drop in magnitude and significance. However, the effects of ideological conservatism and ideological extremity on homophily within stronger ties remain positive and significant, much as they were for the overall homophily rates discussed above ($\beta = 2.515, p < 0.001$ and $\beta = 4.227, p < 0.001,$ respectively). These results show that our hypotheses hold when analysis is restricted to only stronger ties.

[Table B1 about here]

In table B2, we repeat our analysis of non-profit followers with the symmetric homophily measure ***s***(*f*,*H*). As compared to model 2 in table 3, the negative relationship between ego network size and homophily drops in magnitude and retains only borderline significance. Account age, on the other hand, acquires a significant positive association with homophily. The effect of ideological conservatism drops in magnitude, but remains highly significant ($\beta = 4.86, p < 0.001$). The effect of ideological extremity also drops in magnitude, but remains significant ($\beta = 2.636, p < 0.05$). These results offer further support for our hypotheses about the relationship between ideological difference and homophily rates.

[Table B2 about here]

<u>Robustness Checks.</u> To ensure the robustness of our findings to modeling assumptions, we also examined homophily differences in these data using two alternate techniques. For these analyses, we dichotomized the ideological extremity variable along its mean. We first repeated our analyses using randomization inference, randomizing both conservatism and extremity for followers of members of Congress and then those of policy non-profits. We then also reexamined our data with propensity score matching. We estimated two sets of propensity scores via logistic regression, using the controls from models 2 in tables 2 and 3 as predictors, and ideological conservatism and extremity as the outcome (treatment) variables. We omitted the Tea Party caucus membership from the ideological conservatism comparison since there were no liberal Tea Party caucus members. Both randomization inference and propensity score matching confirmed that followers of more conservative or more ideologically extreme members of Congress or policy non-profits are more homophilous than followers of less conservative or less ideologically extreme ones ($p < 0.001$ in all eight comparisons).

Appendix C: Reverse Causation

One possible concern with our findings is that, rather than ideology shaping network structures, network structure shaped ideology. While research suggests individuals' ideologies are likely more influential for the composition of their ego network than vice-versa (e.g., Vaisey and Lizardo 2010), it is nonetheless possible that the observed relationship between homophily and ideological extremity is due to a reverse causal process. This alternate mechanism can be stated as two claims: (1) exposure to ideologically homophilous ties leads individuals to become more ideologically extreme, and (2) this effect of homophily on ideology can plausibly account for the observed relationship. Only claim 2 challenges the conclusions we arrive at in this paper.

Though our data are cross-sectional, we sought to leverage the known length of individuals' tenure on Twitter to test this alternate explanation. Many Twitter ties have no previous off-line existence, and exist only within the confines of Twitter. Given previous research on online political networks (e.g., Sunstein 2009), it is reasonable to assume that a non-trivial portion of the politically homophilous ties we observe corresponds to these online-only relationships. The length of a user's tenure on Twitter (*Account age*) provides an upper bound for the duration of exposure to these ties, and can thus be used as a rough proxy for duration of an individual's exposure to his or her observed levels of homophily.

If it were the case that more homophilous networks made individuals more ideologically extreme (claim 1), we would expect an interaction between homophily and account age to have a significant positive effect on ideological extremity. Moreover, if the rates of ideological extremity we observe are due in substantial part to the observed rates of homophily (claim 2), we would expect this term to noticeably diminish the estimated main effect of homophily on extremity. Since account age is an imperfect proxy for total duration of exposure, it would be unlikely to fully account for this main effect. However, we would still expect the addition of the interaction term to result in a significant downward shift of the estimated coefficient.

We thus estimate and compare two regression models. In Model A, we represent ideological extremity as a function of homophily and control variables. In Model B, we add an interaction between ideological extremity and account age. The results are presented in Table C1. As expected, Model A finds a significant positive relationship between homophily and extremity. In Model B, the homophily X account age interaction has a slight positive effect on extremity in both samples ($\beta = 0.014$, $\mathrm{p} < 0.05$ and $\beta = 0.018$, $\mathrm{n.s.}$, respectively). The results thus offer mixed evidence for claim 1. However, the main effect of homophily remains positive and significant for both samples. The coefficient experiences a slight but non-significant drop (0.0932 to 0.0924, $p \approx 0.997$) in the non-profits sample, and a minute *increase* in the congressional sample. Therefore, while we find some evidence that exposure to homophilous ties may increase ideological extremity (claim 1), we conclude that this effect does not challenge our conclusions regarding the relationship between ideology and homophily we observe in this paper (claim 2).[11]

[Table C1 about here]

[11] We also repeated these homophily-by-account-age interaction analyses with the full set of hub-level controls we introduced in Models 2 of Tables 2 and 3 in the main manuscript (e.g., seniority, leadership role and constituency population count for members of congress; Washington headquarters and employee counts for non-profits, etc.) For both populations, the effect of homophily on extremity remained positive and highly significant, and furthermore remained unchanged after the addition of the homophily X account age interaction, thus pointing to the same conclusion as the models we report in Table B1 above.

Table B1. Followers of Members of Congress: Effect of Ideology on Strong-Tie Homophily

	Coefficients	
Non-independent within hubs (N=159):		
Ideological conservatism	2.515	***
	(0.489)	
Ideological extremity	4.227	***
	(0.489)	
Hub network size	1.307	*
	(0.646)	
Hub age	0.668	**
	(0.225)	
Senator	-0.681	
	(0.843)	
Number of constituents	-0.015	
	(0.243)	
Constituency population density	-0.114	
	(0.088)	
Constituency internet access rate	-0.114	
	(0.241)	
Seniority	-0.105	
	(0.34)	
Committee leadership	-0.431	
	(0.797)	
Party leadership	0.561	
	(1.706)	
Victory margin	0.239	**
	(0.088)	
Tea Party caucus membership	1.724	#
	(1.012)	
Non-independent within followers (N=160,452):		
Account age	1.087	***
	(0.152)	
Ego network size	-0.583	**
	(0.191)	
Reciprocity	*N/A*	
(Intercept)	0.373	
	(0.881)	
R^2	0.156	

Note. Results from linear regressions with two-way clustered standard errors (in parentheses). All non-dummy predictors have $\sigma = 1$.
$\# \; p < 0.10, * \; p < 0.05, ** \; p < 0.01, *** \; p < 0.001$ (two-tailed.)

Table B2. Followers of Policy Non-Profits: Effect of Ideology on Strong-Tie Homophily

	Coefficients	
Non-independent within hubs (N=33):		
Ideological conservatism	4.86	***
	(1.049)	
Ideological extremity	2.636	*
	(1.109)	
Hub network size	0.755	
	(0.742)	
Hub age	0.519	
	(0.37)	
Annual budget	0.36	
	(0.329)	
Number of employees	0.183	
	(0.462)	
Number of volunteers	0.036	
	(0.449)	
Affiliated organizations	-5.935	**
	(1.976)	
Years since founded	-0.356	
	(0.408)	
DC area	2.137	*
	(0.909)	
Unrestricted lobbying	2.691	
	(1.715)	
Non-independent within followers (N=103,864):		
Account age	1.191	**
	(0.372)	
Ego network size	-0.553	#
	(0.284)	
Reciprocity	*N/A*	
(Intercept)	3.622	*
	(1.612)	
R^2	0.186	

Note. Results from linear regressions with two-way clustered standard errors (in parentheses). All non-dummy predictors have $\sigma = 1$.
# $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (two-tailed.)

Table C1. Effects of Homophily-by-Account Age Interaction on Ideological Extremity

	Followers of Members of Congress				Followers of Policy Non-Profits			
	Model A		Model B		Model A		Model B	
Non-independent within hubs:	**N = 159**				**N = 33**			
Hub network size	0.008		0.008		0.224	***	0.223	***
	(0.077)		(0.077)		(0.057)		(0.056)	
Hub age	0.049		0.05		-0.082		-0.082	
	(0.029)		(0.029)		(0.042)		(0.042)	
Senator	-0.176	**	-0.176	**	*n/a*		*n/a*	
	(0.063)		(0.062)					
Non-independent within followers:	**N = 160,452**				**N = 103,864**			
Account age	-0.010		-0.010		0.013		0.014	
	(0.006)		(0.007)		(0.015)		(0.017)	
Ego network size	0.017	*	0.017	*	-0.016		-0.015	
	(0.008)		(0.008)		(0.015)		(0.015)	
Homophily	0.087	***	0.087	***	0.093	***	0.092	***
	(0.009)		(0.009)		(0.022)		(0.022)	
Homophily X Account age	—		0.014	*	—		0.018	
			(0.006)				(0.014)	
(Intercept)	1.013	***	1.013	***	0.828	***	0.827	***
	(0.036)		(0.036)		(0.075)		(0.075)	
R^2	0.097		0.098		0.230		0.232	

Note. Results from linear regressions with two-way clustered standard errors (in parentheses). All non-dummy predictors have $\sigma = 1$. $* \, p < 0.05$, $** \, p < 0.01$, $*** \, p < 0.001$ (two-tailed.)