The Dynamics of Message Selection in Online Political Discussion Forums: Self-Segregation or Diverse Exposure?

Hyunjin Song[a], Jaeho Cho[b], and Grace A. Benefield[b]

[a] Department of Communication, University of Vienna, Austria
[b] Department of Communication, UC Davis, USA

Contact Information:

Hyunjin Song
Department of Communication
University of Vienna
Rathausstraße 19/1/9
Vienna, 1010 Austria
hyunjin.song@univie.ac.at

Jaeho Cho
Department of Communication
University of California, Davis
375 Kerr Hall
1 Shields Avenue
Davis, CA 95616 USA
jaecho@ucdavis.edu

Grace A. Benefield
Department of Communication
University of California, Davis
169 Kerr Hall
1 Shields Avenue
Davis, CA 95616 USA
grbenefield@ucdavis.edu

Note: We would like to thank Dr. Jong Hyuk Lee (School of Journalism & Mass Communication, Kyung Hee University, South Korea) and Dr. Yun Jung Choi (Division of Communication and Media Studies, Ewha Womans University, South Korea) for making original data available for this paper. We also thank Dr. Jakob-Moritz Eberl and Dr. Loes Aaldering for their helpful comments on earlier drafts of the manuscript.

Abstract

While the online sphere is believed to expose individuals to a wider array of viewpoints, a worry about self-reinforcing political echo chambers also persists. We join this scholarly debate by focusing on individual motives for political discussion and dyadic- and structural-level mechanisms that can drive one's message selection decision in online discussion settings. Using unobtrusively logged behavioral data matched with panel survey responses, our Temporal Exponential Random Graph Model (TERGM) analysis indicates that message selection in online discussion settings is largely driven by the similarity of one's candidate evaluative criteria and various endogenous structural factors, whereas the impact of overt partisan preference in shaping message selection is much more limited than is often assumed.

Keywords: online political discussion, online discussion forum, message selection, partisan selectivity, cross-cutting exposure, temporal exponential random graph model

The Dynamics of Message Selection in Online Political Discussion Forums: Self-Segregation or Diverse Exposure?

Internet-based citizen communication, particularly in online discussion forums, has a uniquely low amount of contextual constraints, in that individual choices about which information to view and with whom to associate are relatively unconstrained (Dahlgren, 2005). This unprecedented freedom of choice raises the question of whether the enhanced choice leads to communication that crosses ideological divides or homophilic interactions among the politically like-minded. Thus far, empirical endeavors to address this question have produced mixed findings (e.g., Gentzkow & Shapiro, 2011; Messing & Westwood, 2012). Some researchers suggest that, in keeping with the Harbermasian ideal of free and open space for civil society, online communication platforms are a pivotal, cross-cutting space in which citizens with diverse backgrounds and viewpoints can voluntarily interact (Papacharissi, 2004; Stromer-Galley, 2003; Wojcieszak & Mutz, 2009). This ideal space is believed to expose individuals to a wide array of perspectives, fostering quality and richness in citizen deliberation (Dahlgren, 2005). Yet, contrary to the optimistic view, there also is increasing concern that the Internet functions as multiple self-reinforcing political echo chambers (e.g., Sunstein, 2009), which eventually encourages ideological segregation and political enmity between partisan groups (Boutyline & Willer, 2017; Colleoni et al., 2014). Despite the increasing scholarly debate, how exactly the Internet has changed the landscape of everyday cross-cutting exposure is not yet clearly understood.

Within this context, the present study attempts to advance our understanding of the debate by focusing on *message-selection* dynamics in online discussion forums. Although great progress has been made, much of the prior work has been primarily based on participants' retrospective self-reports (e.g., Stromer-Galley, 2003; Wojcieszak & Mutz, 2009) or solely based on what have been posted in online forums (e.g., Himelboim, 2008;

2011; Graham & Wright, 2014), each of which has inherent limitations. Self-reports provide versatile and flexible data as virtually everything can be measured; yet, they often have questionable measurement accuracy, especially when it comes to behavioral constructs (e.g., Prior, 2009). Content posted on discussion forums or social network sites only provides information between visibly connected dyads (i.e., post-reply relations between actors *i* and *j*). Thus, exposure to a message, whether cross-cutting or not, is observed only when actor *i* *replies* after reading actor *j*'s message. This does not fully capture message selection because selecting and reading a message does not necessarily result in replying to the message. In a similar vein, some work has explored selection behavior in online social networks (Boutyline & Willer, 2017; Colleoni et al., 2014); yet, it focuses primarily on "channel" selection (e.g., "following" relations in Twitter), a decision that likely occurs *after* exposure to another's messages (presumably more than one). Thus, there has been a lack of systematic investigation into individuals' *message-selection* decisions – whether one chooses to *read* a given message in a forum – even before they choose to react and reply to the message. This oversight is particularly troubling since key to the debate over the role of the Internet in democracy is whether citizens are indeed exposed to or choose their fellow citizens' messages with viewpoints different from their own.

Recognizing the limitations in extant research, we direct our attention to individuals' message *selection* (i.e., reading) in online discussion forums. Past research suggests that dynamics in communicative interactions "cannot be regressed to mere individuals' predispositions or pure social selection processes based on gender, race, or political viewpoints" (Song, 2015, p. 18; see also Lazer et al., 2010), but rather individual differences (e.g., Mondak, 2010), sociopolitical correlates such as dyadic political agreement and disagreement (e.g., Schmitt-Beck & Lup, 2013), and endogenous dynamics of discussion networks (e.g., Author, 2015) all uniquely explain communicative interactions among citizens. As such, we approach message-selection behaviors from three different layers of analysis: individual motivations, dyadic homophily, and network structural features.

Our goal here is to identify whether, and how, the voluntary reading of one another's messages results from a purposive search for political similarity or is instead mainly propelled by elements that are less likely to pertain to (or be shaped by) one's overt partisan preference, such as motivations (i.e., understanding) and structural features of the discussion network. If the latter is the case, then incidental exposure to cross-cutting messages in online discussion networks is more likely. By identifying the ways in which messages are selected, we also address the flip side – whose messages are more likely to be *read* by others. Understanding these issues also sheds light on how aggregate exposure patterns – as the end-result of someone's message selection – emerge from individuals' message-selection behaviors and what role online citizen communication plays in the democratic process.

In the following discussion, we emphasize two competing explanatory principles – *consistency* and *understanding* – as the two motivational drivers of online political interactions. We further advance our perspective on how such competing principles could operate in a dyadic setting, and ultimately, how online discussion network structures could recursively influence message selection. We then offer an empirical assessment using novel panel survey data matched with behavioral log data collected during a presidential election period from an online forum in which participants voluntarily posted, read, and replied to messages. With detailed information about who selects whose messages and its correlates during a period of heightened attention to politics, our data are aptly suited for disentangling whether online message-selection patterns are primarily structured along partisan lines. Our results from an inferential network-analytic method called temporal exponential random graph model (TERGM) demonstrate that the impact of political preferences in shaping message selection is much more limited than often assumed.

The Two Motivational Drivers of Political Discussion: *Consistency* and *Understanding*

A recurring theme in the study of political communication is how much of citizens' choice about what information to consume is driven by their political beliefs (e.g., Iyengar & Hahn, 2009; Stroud, 2011) and how much of it is explained by other non-partisan considerations (e.g., Messing & Westwood, 2012). Underlying these two possibilities is the distinction between two different, yet co-existing, human motivations: *cognitive consistency* and *understanding* (Holbert, Weeks, & Esralew, 2013, or *directional* vs. *accuracy* goals in Kunda, 1990). The *cognitive consistency* principle suggests that individuals prefer pro-attitudinal messages that lead them to their desired conclusions (e.g., Iyengar & Hahn, 2009; Stroud, 2011). In contrast, the principle of *understanding* posits that people are drawn to messages that they believe help them make sense of the situation at hand and reach accurate conclusions (Holbert et al., 2013). Although these tendencies have been discussed largely in the context of mass media sources and messages, they provide a reasonable explanatory framework for message selection in online discussion settings where interaction is much less spontaneous than face-to-face conversation and discussants often have strong control over message choice (selectivity). Within the context of the current study, we discuss how these two principles of human motivation (i.e., *consistency* vs. *understanding*) shape the dynamics of message selection in an online discussion forum at both individual (i.e., the selection of others' messages and one's own message being selected by others) and dyadic (i.e., actor *i* selecting actor *j*'s message based on a dyadic characteristic shared by the actors) levels.

Principles of Consistency and Understanding at the Individual Level

When it comes to *out-going* message selection (i.e., reading others' message), it is plausible that those who are most motivated by *understanding* would seek out messages that they think are relevant and useful, regardless of whether such messages are pro- or counter-attitudinal. Because having an accurate understanding motivates them, they extend more effort navigating and sorting through available messages, but are less likely to be bounded by their political preferences. Thus, their strong appetite for information translates into a high frequency of reading others' messages. This is in line with previous findings that accuracy motivations (Valentino et al., 2009) and a need for cognition (Cacioppo et al., 1996) each positively predict a host of information-seeking behaviors. On the other hand, those who have a high consistency motivation likely avoid messages with which they disagree; thus, their overall information search is more narrow and selective. In a non-self-selected online forum where both pro- and counter-attitudinal messages are present, people with strong consistency motivation likely screen out messages that are not consistent with their political and partisan beliefs and only consume some "safe" messages (Bennett & Iyengar, 2008), which would lead them to spend less time reading others' messages in general. This is especially likely when the valence of a new message cannot be reliably predicted *before exposure* due to the presence of both pro- and counter-attitudinal messages (e.g., Shook & Fazio, 2009). Recent research, however, suggests that even strong partisans do not *always* avoid attitudinally incongruent information (Garrett & Stroud, 2014; Valentino et al., 2009). This speaks to the possibility that avoidance may not be the only way for those with high consistency motivation to respond to counter-attitudinal information. If this is the case, consistency motivation does not necessarily reduce information seeking and message-selection behaviors.

For patterns of *incoming* message selection (i.e., being read by others), we expect that messages posted by those with higher *consistency* motivation will likely communicate strong, clear partisan perspectives (Ahn, Huckfeldt, & Ryan, 2014). These partisan cues revealed in the posted messages will then attract attention from other participants. Given the widespread preference towards attitudinally congruent information in the general public (Garrett & Stroud, 2014), partisan language likely functions as a trigger for message selection, at least by those on the same partisan side. Also, when a message reflects a psychological desire for attitudinal consistency, the message is more likely to exhibit controversial or conflict-ridden elements. This, in turn, will likely draw attention. In contrast, those with higher *understanding* motivations are less likely to express their partisan viewpoints. They are prone to making more considerate judgments that weigh diverse perspectives of the pros and cons of an issue (Rudolph & Popp, 2007). Since their messages are less likely to contain strong, one-sided partisan information, their messages are less likely to be chosen by others.

Although we expect that an individual's predisposition towards understanding or consistency motivation will influence message-selection patterns in online discussion forums, the rationales behind the relationships still remain speculative. Thus, we propose a research question, rather than hypothesis, as follows:

> **RQ**: In an online discussion forum, how will (a) consistency motivation and (b) accuracy motivation be related to out-going and incoming message-selection patterns, respectively?

Principles of Consistency and Understanding at the Dyadic Level

Above and beyond its impact on message-selection dynamics at a purely individual-level, consistency motivation also plays a role at the dyadic level. Homophily, or the tendency of a given dyad to associate with each other based on their similarities (McPherson et al., 2001), has long been regarded as a powerful determinant of message-selection decisions (Author, 2015; Garrett & Stroud, 2014; Iyengar & Hahn, 2009). Based on either the explicit application of political preferences or a de facto preference for similarity, research has repeatedly suggested that people can selectively construct their own social environment (Kossinets & Watts, 2009; Lazer et al., 2010; McPherson et al., 2001), and especially less likely to be exposed to diverse political viewpoints online (e.g., Bakshy et al., 2015; Himelboim, McCreery & Smith, 2013). Within the present context, this means that the ego ("the focal respondent") and the alter ("the potential discussion partner") are more likely to select each other's messages if they share similar political preferences. Therefore, we posit that:

H1: Participants in an online discussion forum will be more likely to select each other's messages when they share similar political preferences.

The understanding principle, on the other hand, paints a somewhat different picture. We expect that, due to the human desire to reduce the information cost in a decision situation (Downs, 1957; Pietryka, 2016), individuals will be inclined to search for information that is deemed to have high "utility" or "relevance" to the current situation. In the midst of election campaigns, citizens prioritize candidate evaluation and vote choice. These decisions are often based on considerations about various factors, including candidates' personal traits and backgrounds, party affiliations, and/or issue positions. Since each voter has their own criteria for evaluating candidates, they will individually determine which elements of messages provide the most utility and relevance. In line with this expectation, Ahn and colleagues (2014) suggest that voters often actively glean relevant information from their social networks and appear to value political expertise even in the absence of political agreement. Similarly, Hart et al. (2009) shows that disconfirmation bias is substantially reduced for messages with higher informational value. We, therefore, expect that, in a dyad, two discussants with similar candidate evaluation criteria (consisting of how people evaluate candidates' personal traits and backgrounds, party affiliations, and/or issue positions), whether they are like-minded or not, are more likely to select each other's messages when they are considered to be of high utility and relevance for their decision about whom to support. Based on this rationale, our hypothesis is stated as follows:

H2: Participants in an online discussion forum will be more likely to select each other's messages when they share similar candidate evaluation criteria.

The Endogenous Impact of Network Structure

While the aforementioned factors are important aspects of message-selection dynamics in their own right, they do not operate in a social "vacuum." Since a theoretical perspective that ignores substantive interdependencies among actors is inevitably incomplete, we attempt to explicate such interdependencies in message-selection patterns.

Reciprocity

Often in online discussion forums, users not only intentionally seek information but also spontaneously exchange and respond to others' messages. This may take a number of different forms, yet the most simple and frequent form of such "interaction" may manifest as continuous, interactive message-exchange sequences among a set of users. This also implies that such interaction patterns may require a situation in which actor *i* and actor *j* mutually choose to view each other's messages and return their attentions to each other – provided that replying to an original message necessitates the responder to actually click and read that message in the first place. Based on this expectation, we hypothesize that reciprocity (Wasserman & Faust, 1994) will be a significant and positive predictor of online message selection, as follows:

> **H3**: Participants in an online discussion forum will be more likely to reciprocate message selection when another participant has already selected his or her message.

Transitivity, Cyclic Closure, and Local Hierarchy

Transitivity and cyclic closure represent another mechanism of how individuals choose to encounter socially provided messages. Transitivity denotes a situation where node *i* is more likely to create a tie to node *j* when they are both connected to another node *k*. In contrast, cyclic closure denotes a similar but opposite situation where node *j* forms a tie to node *i* when they are connected to another node *k* (Holland & Leinhardt, 1976), as can be seen in Table 1.

While the most common explanation for transitivity is that it reflects the local spread of a social relationship (e.g., "friends of my friends are my friends"), such an explanation is somewhat less likely within the context of *message selection* in an online discussion forum. That is, the spread of a social relationship requires actors to be aware of each other's social relationships in choosing to interact with one another. In online discussion forums, however, information about whether *k* has chosen to view *j*'s messages (which is a prerequisite for a social relationship to spread) is generally not available when *i* choose to view *j*'s messages.

Instead, in light of an understanding-based explanation, we instead argue that a pattern of transitivity may arise from the hierarchical nature of the underlying criteria that people use when they choose each other's messages. Here, individuals are assumed to pursue a tie with others whose messages exhibit higher status (e.g., argument quality, expertise, trustworthiness, etc.) than their own. Thus, actor *i* is expected to seek a tie with a "higher status" actor *j* (i.e., reading *j's* message), given that *i* has an existing relationship with an intermediate-status actor *k* who also has a tie to *j*. In this scenario, *i* does not necessarily have to be aware of *k*'s tie to *j*, which is often invisible in online discussion forums. Rather, because of *j*'s high status, *j*'s messages are sought by many individuals in the network, including *i* and *k*. If *k*'s status is higher than *i* but lower than *j*, *k* will be sought by *i* but not by *j*. When coupled with a negative tendency towards cyclic closure (e.g., *j* is less likely to form a tie with less prestigious actor *i*), positive transitivity pattern can be interpreted as the local status hierarchy in a given network (Lazega et al., 2012). While this does not necessarily imply that, at all times, people only purposively seek out higher status individuals based on message qualities, evidence indicates that people routinely seek and rely on guidance from those who are more politically versed and sophisticated within their social networks (Ahn et al., 2014; Downs, 1957; Huckfeldt, 2001). Consequently, one possible source of such hierarchical network structuring can be an individual's need for political experts and, thus, the choice to view those experts' messages. Therefore, we predict:

H4: Participants in an online discussion forum will be more likely to select each other's message based on a local hierarchy in the discussion forum (i.e., positive transitivity and negative cyclic closure).

Profile Similarity

Another mechanism that helps us understand the nature of message selection in online forums is the concept of profile similarity (DiMaggio, 1986). In addition to the hierarchical nature of message-selection networks, individuals are more likely to view one another's messages when both share the same connections with other actors in the network. For instance, if actors *i* and *j* both choose to view the same set of alters ("activity closure"), or *i* and *j* are *chosen* by the same set of alters ("popularity closure": see Table 1 below), then the same patterns of incoming and out-going connections shared by *i* and *j* signal a common set of properties of the *i-j* dyad (Block & Grund, 2014; Robins et al., 2009). In such a situation, *i* and *j* are more likely to see each other's messages. In line with a consistency-based explanation, this may be viewed as the structural basis of homophily, in that the formation of ties is driven by the similarity in choices with respect to other actors (DiMaggio, 1986). Therefore:

H5: Participants in an online discussion forum will be more likely to select each other's messages when they have selection patterns similar to all other participants (profile similarity).

Preferential Attachment

Many studies indicate that the structure of online social networks tends to follow a power-law distribution, characterized as the skewed distribution of degrees (Barabási & Albert, 1999). While the existence of such a pattern is rather common, it appears that this tendency is also pronounced in online discussion forums. For instance, Himelboim's (2008; 2011) analysis suggests a sharp inequality in the ability to draw attention and elicit further engagement with a given message from a large number of users in online discussion groups. When selecting which messages to click in an online discussion forum, one often pays attention to certain heuristic cues such as the number of "views" and "likes," which signals utility based on the popularity of a message. Therefore, a message that has a large number of engagement cues (such as views or likes) can draw disproportionate selection behaviors through self-reinforcing dynamics, leading to a highly imbalanced message-selection distribution. Therefore, we expect:

H6: Participants in an online discussion forum will be more likely to select messages that have already been selected by a large number of others.

Temporal Dynamics in Message-Selection Criteria

As elections near, it is reasonable to believe that individuals are more mobilized by campaign communication (Author, 2013) and, thus, pay close attention to political messages both online and offline. This is more likely, not only due to a heightened attention to politics, but also because individuals need more information to reduce uncertainties and anxieties about their voting decisions (Downs, 1957). While the literature generally suggests that strong partisans and interested voters arrive at their decisions early in the election campaign cycle (Fournier et al., 2004), the day-to-day dynamics in the campaign environment may prompt them to seek out confirmatory information. Specifically, an increase in uncertainty about the election outcome may induce confirmatory information seeking behavior (Carnahan et al., 2016; Valentino et al., 2009). As changes in the campaign environment (e.g., campaign competitiveness) *over time* induce more anxiety and uncertainty about the election outcome, the effect of preference homophily (i.e., message selection based on similar political preferences) may increase. Therefore:

H7: The effect of preference homophily on message selection increases over time.

Data and Methods

In order to test our predictions, we draw on a unique set of panel data collected during the 2012 South Korean presidential election. The data were collected from an online discussion forum hosted on a research firm's server where participants' posting and viewing activities during a 27-day period until Election Day (from November 23 to December 19, 2012) were unobtrusively logged. A market research firm in Korea, *Embrain*, randomly recruited 400 participants from a national opt-in panel with access to over greater than one million identity-verified individuals that closely matches gender and age distributions of the entire Korean population. The participants were then invited to a custom-created website, named "An Online Forum about the 18th Korean Presidential Election," and were instructed to create their own login ID and freely post and read each other's posts about the upcoming election, as they normally would do in other online forums.[1] Three surveys were each administered in the beginning (4 days after the launching of online discussion), middle (two weeks after the wave 1 survey), and end (right after Election Day) of the study period, respectively. Of the 400 initial participants, a total of 341 participants remained on the online discussion forum for the 27-day study period and completed all three waves of panel surveys. Upon completion of the project, a monetary incentive of approximately US$100 (equivalent

[1] The structure and user interface of the discussion site were adopted from the typical format of most online forums (e.g., Reddit). On the forum site, participants were not only allowed to initiate their own posts but also to read and respond to others' original posts and the subsequent comments. Once logged in, participants were exposed to the main page of the forum where the list of post titles (along with user ID, a timestamp, and the number of views for the post) made by either themselves or other participants were presented, with the latest ones at the top; the title of the post was also accompanied by comment counts (displayed at the right end side of the title in a smaller font size). Actual comments, if any, were posted under an initial post in the order they were posted, forming a thread of discussion and therefore were not readily visible when participants logged in. To summarize, as in many online forums, information available to participants when they choose which messages to read only included the list of thread titles, comment counts, user IDs and timestamp of respective posts, and the number of views for the post. A separate minor section on the main page carried study-related information (e.g., reminder about panel surveys, announcements to encourage forum participation, etc.). No other information was provided that might influence participants' behaviors.

to 100,000 Korean Won) was provided to the participants in return for their participation in the online discussion forum and the three surveys.

The surveys measured the participants' candidate evaluations and criteria, policy preferences, motivations for using online discussion forums, and other key covariates of interest.[2] Activity log data regarding posting and browsing behaviors were later retrieved from the research firm's server and matched with participants' survey responses. We note that this data set was utilized in another publication, for another research purpose, which examined the impact of online political expression on the expressers' political preferences (Author, in press). In wave 1, 22 of the 341 participants neither provided candidate preference nor self-reported political ideology, and additional 7 respondents did not report self-reported political ideology. Therefore, total missing cases were 29 (8.5%) based on *either* candidate preference *or* self-reported political ideology.[3] Since homophily based on these two variables serves as a key predictor in our model, we have excluded such cases (thus, $N = 312$). Yet an identical model including 29 missing cases with multiple imputation (missing data imputation $N = 5$) on candidate preference did not alter the results and conclusion reported herein.

Construction of Networks

Over the period of data collection, participants on average posted approximately 25

---

[2] Although slightly skewed in age (sample median age = 35; population = 38) and sex (sample = 48.3% female; population = 49.67%), our final sample closely matches the general population in demographic profile. In addition, the sample's representativeness is less of a concern because we are taking an inferential network-analytic approach. It is also noteworthy that our sample had enough variability in all of the key covariates (especially for candidate preference), making less likely that our results are biased by the peculiarity of our data. Nonetheless, using an opt-in, not probability-based, panel for recruitment would limit the generalizability of our results and would likely introduce bias associated with participants' motivations for "opting-in" to the panel (Hillygus, Jackson, & Young, 2014).

[3] Among 29 individuals who do not report their ideological self-placement, a total of 7 individuals indeed provided their candidate preference. A separate analysis excluding only those who do not have candidate preference (eligible $N = 319$ instead of 312) yielded identical results reported in the main analyses.

messages and read 102 unique postings made by others, resulting in an average of 547 reading instances per individual.[4] Based on the participants' activity logs, we have derived a "message selection" network as a directed actor-actor binary matrix (312x 312), such that the cell entry X*ij* is defined as 1 when actor *i* chooses to view actor *j*'s message and zero otherwise (in doing so we also retain the direction of ties, such that $Xij \neq Xji$). Since our analytical strategy (i.e., use of TERGM: see *Analysis Strategy* section below for more details) requires all cell entries to be defined as binary rather than integers, we opted to dichotomize numbers of selection instances within the same dyad using the mean number of message selections across all dyadic pairs as a threshold (W1 = 2.5; W2 = 2.9; W3 = 3.1). Therefore, our model only speaks to relatively routine, repeated message selection dynamics in a given network panel rather than all-inclusive message selection dynamics, such as accidental, spontaneous selection behaviors. Yet our additional robustness checks based on models with lower threshold value (0 vs. all other values) for dichotomizing ties, models with daily slices ($t = 26$, instead of three-wave panel network as reported here), as well as based on QAP regression models with equivalent predictors (yet excluding any network-endogenous structural variables) found largely the same results with minor discrepancies in estimated coefficients and significance level.[5]

Based on the dates of the three panel surveys (W1 = Nov 27th to 29th, W2 = Dec 11th

---

[4] Data logs indicate that participants in general were active on the discussion forum. On average, there were 221.81 unique users per each day at the forum (either as a poster or a reader) during the study period (*SD* = 30.99). Data also indicates that participants were especially active as a reader. All of the 312 participants read others' messages at least once, while only few participants did not post at all (*N* = 5, 1.6%) during the study period. Indeed, participants typically read at least 5 or more posts per day (median reading counts divided by total no. days = 5.01) and posted at least 1 message or more every two days (median of posting count divided by total no. of days = .51). Overall, although the distributions of reading and posting frequencies are somewhat skewed (to the right), most participants were engaged to a varying degree in online discussion.

[5] Combined with multiple imputation results, our robustness check suggests that our results and conclusions are reasonably robust against potential methodological issues. All model robustness check results can be found in online Supplemental Information.

to 13th, W3 = Dec 21th to Dec 23th), we created longitudinal panel networks of message selection by partitioning log data from the first two waves and matching it to corresponding survey dates (e.g., log data from Nov 27th to 29th were regarded as the 1st wave of the network panel). Since the 3rd wave of the survey was conducted *after* Election Day (which was Dec 19th) but electronic log data were only collected *until* Election Day, we regard the last three days of log data (Dec 17th to 19th) as the last panel in the network.[6] We consider the log data that were available four days prior to the first survey wave (Nov 27th) as well as that collected *between* each survey waves as lagged observations of the respective network panel. Specifically, log data from Nov 23rd to 26th were considered as lagged observations of the first network (Nov 27th to 29th), data from Nov 30th to Dec 10th as lagged observations of the second network (Dec 11th to 13th), and data from Dec 14th to 16th as lagged observations of the last network (Dec 17th to 19th).

Measures

Consistency and understanding motivations. For consistency motivation (Cronbach's α = .86, M = 4.36, SD = 1.03), respondents were asked six items (based on a 7-point scale from "Not at all" = 1 to "Very much" = 7) about whether they visit online discussion forums (including discussion forums other than in the current study) primarily "to justify my opinion of the issue" or "to confirm that my opinion on the issue is correct." Understanding motivation (α = .81, M = 5.26, SD = .82) was assessed in a similar manner using four items (e.g., "to make an accurate and objective assessment of the issue", "to understand others' opinions", etc.). Since motivations were measured only once at the first wave of the survey, we regard these characteristics as time-invariant covariates in our model.

Preference homophily and evaluative criteria similarity. We define political

[6] Since participants' key characteristics (e.g., candidate evaluations and preferences) are stable across survey waves, we assume participants' characteristics drive the creation of network ties (but not the other way around).

preference homophily (i.e., *consistency* principle) in three different ways: (a) candidate choice, (b) policy preference, and (c) self-reported ideology. Candidate choice homophily (W1: M = .51, SD = .49; W2: M = .55, SD = .49; W3: M = .52, SD = .49) was defined in a way that a given dyad was regarded as homophilous (coded as "1") when they share the same candidate choice. Policy preference homophily (W1: M = .40, SD = .16; W2: M = .38, SD = .16; W3: M = .39, SD = .16, all range = 0 to 1) was operationalized with respondents' preferences towards liberal vs. conservative oriented policy options about economic and North Korea issues. We derived a Euclidean distance, d, of a given dyadic pair in terms of their dissimilarity in policy preferences, which was later converted to similarity by taking 1 / (1 + d) to make a greater value represent preference "homophily." Lastly, ideological dissimilarity (M = 1.46, SD = 1.13 throughout all waves, range = 0 to 6) was measured based on the absolute distance of ideological self-placements between a given dyad, such that a shorter (greater) distance would indicate greater (lesser) degree of homophily.[7]

Next, we define candidate evaluation criteria similarity (M = .48. SD = .15, range = 0 to 1) in a similar manner, using a dyadic Euclidean distance d in terms of relative importance of policy/candidate characteristics (e.g., policy, competence, integrity) versus personal background (e.g., party affiliation, political career, place of origin, etc.) in candidate evaluations. Since candidate evaluative criteria were only measured in the wave 1 survey, we treat this measure as invariant across waves.

Network-endogenous measures. Reciprocity was captured by whether a pair of

[7] Candidate choice (W1: M = .60, SD = .49; W2: M = .66, SD = .47; W3: M = .61, SD = .48) was tapped using a dichotomous measure, where "1" denotes supporting the liberal candidate (Moon Jae-in) vs. "0" denotes supporting the conservative candidate (Park Geun-hye). Policy preferences were measured three times across panel surveys using four-item measures, based on a 7-point scale from "not at all agree" (1) to "very much agree" (7). Ideological self-placement (M = 3.69, SD = 1.31) was measured once in the first wave of the survey (therefore regarded as invariance across waves), based on a 7-point scale from "very liberal" (1) to "very conservative" (7).

actors mutually selected each other's messages. For measures tapping a series of triadic configurations (transitive closure, cyclic closure, activity closure, and popularity closure: see Table 1 for details), we relied on the *directed* version of the geometrically weighted edgewise shared partner (directed GWESP) statistics following the model specifications proposed by Snijders et al. (2006) and Robins et al. (2007). The GWESP term captures higher-order triadic effects in the network, such that the tendency of directly connected ego and alters to have multiple shared third-party discussants (for a detailed discussion of this measure, see Hunter & Handcock, 2006). As described above, our theory suggests that a series of triadic closure patterns would have a substantial effect on message selection dynamics. Similarly, for measuring in- and out-degree distribution effects, geometrically weighted out-degree and in-degree distribution (GWD-out and GWD-in) terms were used where the parameter estimates for GWD terms represent "evenness" of in- and out-degrees based on message selection activities across the network (see Hunter, 2007 for details). We expect these terms to be significant and negative, which would signify differential message selection activities across the network.

Control variables. In order to establish a plausible baseline in our analysis, we control for a host of variables that are known to be related to the extent to which people engage in political discussion. First, we control for participants' socio-demographic factors such as *gender* (1 *being* "female," 48.39%), *age* (in 10-year increment, $M = 3.55$, $SD = .98$), *education* (from "not finished elementary school" = 1 to "currently in post-graduate education or more" = 9, $M = 7.71$, $SD = .97$) and *region of origin* (1 being "Seoul" vs. 0 being "other regions", 40.38% from Seoul). In our analysis, we also controlled for two demographic homophily measures, one based on their gender and the other based on their regional origin (all coded as 1 if a dyad shares the same gender or regional origin), since preference homophily may be confounded with demographic homophily (McPherson et al., 2000). We also controlled for respondents' offline discussion frequency (from "Never" = 1 to "Always" = 7, W1: $M = 4.50$, $SD = 1.04$; W2: $M = 4.62$, $SD = 1.18$; W3: $M = 4.82$, $SD = 1.17$), news use frequency (measured in *hours*, W1: $M = .76$, $SD = .42$; W2: $M = 1.56$, $SD = 1.66$; W3: $M = 1.65$, $SD = 2.32$), internal discussion efficacy (from "Not at all agree" = 1 to "Strongly agree" = 7, $M = 4.72$, $SD = .98$), and hedonic motivation ($\alpha = .75$, $M = 4.47$, $SD = 1.04$) for using online discussion forums. News use frequency was defined as the average exposure in hours to Internet, newspaper and television news about the upcoming election. Internal discussion efficacy was gauged using a four-item composite measure tapping how competent and efficacious an individual is in a typical political discussion setting (e.g., "I am competent at presenting my own opinions in a discussion"). Hedonic motivation was assessed by a three-item measure, all anchored on a 7-point scale, asking whether they participate in online forum based on pleasure-seeking motives (e.g., "it is interesting and fun").

Analysis Strategy

Since we aim to properly capture and explain substantive interdependency dynamics over time, we modeled longitudinally observed message selection networks using a Temporal Exponential Random Graph Model (TERGM), a time-series extension of the ERGM framework with the bootstrapping resampling technique described in Desmarais and Cranmer (2012). It is integral to this approach to model the ties in a given network to be a random variable ("1" for existence of ties, and zero otherwise) to be explained simultaneously by a collection of actor covariates and network-endogenous dependencies (Robins et al., 2007; Snijders et al., 2006), while properly accounting for the non-independence of observations inherent in network data. The ERGM framework is now regarded as the most versatile yet flexible method for evaluating the underlying generative properties of a network, as exemplified in recent applications of the method to various domains (Cranmer et al., 2017).

One important consideration regarding our analytical strategy is that we do *not* directly rely on textual information itself in our current statistical models, although message characteristics might play a non-trivial role in message selection behaviors (e.g., textual similarity between messages of two actors, or whether the message belongs to a certain topic that draws attention from a given reader). However, it is not easy to incorporate textual characteristics directly into our current model because message characteristics should be defined at the tie-level (regarding the strengths, scope, and characteristics of a connection between actor *i* and actor *j* due to a message) while we should also account for the fact that such message characteristics are (at least partially) endogenously determined over time by individual-level message selection behavior itself. Indeed, the current application of the ERGMs in general lacks a proper method of stochastically incorporating topic-based content information (see Kim, Schein, Desmarais, & Wallach, 2017). Readers should bear in mind this limitation of the current approach in evaluating our analyses and results.

In applying a longitudinal inferential network analysis technique, we regarded an observation at a given point in time as depending only on the previous state of the network (i.e., a lagged observation). In capturing temporal dependencies, we include as additional control variables a series of lagged endogenous network statistics that might be relevant in messages selection behaviors, as well as a few additional endogenous network statistics (such as *isolates* and *two-paths*) that are necessary for controlling temporal or lower-order effects when estimating the effect of key parameters. Details of the model specification are provided in online Supplemental Information. Table 1 below summarizes key model terms and their corresponding hypothesis, with their graphical depiction and substantive interpretation.

[Table 1 About Here]

Once models were fitted, we assessed goodness-of-fit (*gof*) to identify the model adequacy by simulating nine hundred new networks (three hundred new networks for each time step) and compared the network characteristics from the observed vs. simulated networks (Hunter et al., 2008). The *gof* results indicate that model specification is satisfactory (see online Supplemental Information for details). All analyses were based on maximum pseudo-likelihood estimation with bootstrapped confidence intervals (Desmarais & Cranmer, 2012), as implemented in the *btergm* package in R (Leifeld et al., 2017).

Results

Using QAP (Quadratic Assignment Procedure) regression models (Dekker, Krackhardt, & Snijders 2007), we first present bivariate, unconditional relationships among our dependent variable (i.e., message selection behaviors) and four dyadic level predictors while controlling for underlying network structures, as can be seen in Table 2 below.[8]

First, if we only focus on the bivariate relationship between message selection behaviors (as captured by our dependent networks) and the first three dyadic predictors for partisan preferences (i.e., *same candidate preference*, *policy preference similarity*, and *ideological dissimilarity*), we observe that partisan preferences alone generally do not predict message selection behaviors well at a bivariate-level. Neither same candidate preference nor ideological dissimilarity were significant in predicting message selection behaviors as reported in Table 2. We do find some indication that policy preference similarity variables become significant at wave 2, yet the pattern indicated that policy preference "similarity" between a given dyad may contribute to "less" message selection behavior for that dyad. In addition, when we predict our dependent variable as a function of evaluative criteria similarity, we see this variable is generally not significant at a bivariate-level as well.

[8] By relying on row-and column-wise reshuffling, the QAP preserve the dependency structure across observations within a dependent variable (i.e., underlying network structure) but remove the associations among predictors and the dependent variable. Since this is essentially a dyadic-level analysis, we only focus on dyadic-level predictors. In addition, since QAP treats dependencies among the observations as a nuisance rather than substantive features that should be modeled, we do not consider any higher-order structures in this analysis. See Cranmer et al. (2017) for a detailed discussion about this issue, in comparison with the ERGM applications.

Moving to a result of the multivariate analyses, Table 3 below reports the key parameter estimates from the final TERGM specifications along with its 95% confidence intervals based on bias-corrected and accelerated CIs using 1000 replications (also graphically reported in Figure S1 and in Table S1 in online Supplemental Information). The leftmost model specification ("Final Model I") in Table 3 includes the effects of motivation and homophily controlled for the hypothesized network structural influence, and the second model specification ("Final Model II") in Table 3 includes dissimilarity based on ideological self-placement (using absolute difference term, which signifies the *opposite* of homophily) instead of candidate preference homophily and policy preference homophily.[9] In addition, a series of interaction models from 3nd to 4th column test whether the effects of preference homophily increase over time. Across all models, coefficients can be interpreted as log odds of a tie conditional on the rest of the network and other model terms. For the remainder of the manuscript, we mainly report the result of the first model specification ("Final Model I" in Table 3), and additionally mention other results where appropriate.

[Table 2 and Table 3 About Here]

Our research question asked how consistency and understanding motivations systematically affect the likelihood of messages *being selected* by other participants, as well as an individual's selection patterns (i.e., selecting others' messages) within the online discussion forum. For the final model specification, we found the effect of consistency motivation to be non-significant in predicting outgoing selection (b = .025, 95% bootstrap CI = [−.044, .077]), so as to understanding motivations predicting incoming selection (b = −.052, [−.080, .022]). In contrast, we found a small but significant tendency for consistency

[9] Given the predominance of ideology in determining specific policy attitudes and candidate preference (Jacoby, 1991; Jost, Federico, & Napier, 2009), we only include ideological placement homophily in our second full model, in order to avoid conceptual and empirical overlap between ideological homophily and the rest of the two preference homophily variables (i.e., homophily based on candidate preference and policy preference).

motivation to predict in-ties (*b* = .034, [.009, .113]) and understanding motivation to predict out-going ties (*b* = .028, [.005, .076]). Empirical patterns indicate that those who strive to understand the outside world, as opposed to those with low understanding motivation, are more likely to select and read others' messages in the online discussion forum. At the same time, on average, people are more likely to select and read messages written by those with higher consistency motivation.

Concerning our dyadic-level homophily variables, neither candidate choice homophily (*b* = −.032, [−.070, .047]) nor policy preference homophily (*b* = −.108, [−.212, .006]) is related to message selection. When we examine the influence of more general "ideological" dissimilarity rather than context-specific candidate preferences or policy preference similarity (as in "Final Model II," Table 3), we still observe that ideological dissimilarity does not predict message selection behaviors (*b* = .024, [-.007; .040]). Thus, H1 is not confirmed. Such null effects indicate that consistency-driven dynamics (i.e., whether a dyad shares a candidate preference, policy preference, or ideological preference) is likely not related to whether people choose to select and view each other's messages. Instead, we find a consistent and substantial effect of similarity in candidate evaluative criteria, such that the more similar a dyad in terms of their candidate evaluative criteria, the more likely they are to expose themselves to another's messages (H2: *b* = .407, [.399, .415]). We return to the implications of this finding in the discussion section.

Our next set of hypotheses concerns the endogenous structural effects of network itself. As shown in Table 3, we have found consistent and robust support for these predictions, such that reciprocity (H3: *b* = .768, [.560, 1.068]), multiple cyclic closure (H4: *b* = −.066, [−.076, −.061]), multiple activity closure (*b* = .035, [.033, .043]) and multiple popularity closure (*b* = .113, [.083, .232], all H5), and preferential attachment (measured as *Popularity spread*, H6: *b* = −4.123, [−5.343, −3.541]) were all strongly supported, controlling for the tendency to not have any ties (*isolates*: $b = 1.003$), open triad (*multiple two-path*: $b = .003$, all CIs straddle zero), temporal dependencies, and other motivation and homophilies.

Among estimated effects, the effect of preferential attachment (or an uneven degree distribution) was the strongest and substantial, as the negative incoming degree distribution parameter indicates (H6: $b = -4.123$). Figure 1 gives a substantive interpretation of the effect, suggesting that, irrespective of time periods, the predicted probability of receiving at least one additional message selection instance from other participants in the forum (excluding who are already connected) sharply increases as a function of the existing in-degree of a node. This suggests that message selection behaviors are largely driven by self-organizing dynamics, consistent with the notion that people are disproportionately drawn to and more likely to expose themselves to *already popular* messages in a forum (Himelboim, 2008).

[Figure 1 About Here]

In addition to the effect of preferential attachment, participants in the online forum were approximately 2 times (conditional odds ratio = 2.15) more likely to browse others' messages based on a reciprocity effect. Likewise, an individual (ego) is approximately 4 to 12 percent more likely to read another participant's (alter) message for every one person increase in the number of other participants to whom the ego and alter are tied based on outgoing (*multiple activity closure*: conditional OR = 1.035) and incoming connection patterns (*multiple popularity closure*: conditional OR = 1.121). This suggests that when message selection patterns signal latent shared characteristics within a dyad, participants are more likely to select each other's message. Participants in our online forum were also slightly less likely to form a closed three-cycle, suggesting the network has a slight tendency against generalized exchange that returns to lower status individuals. The only exception to this pattern was the multiple path closure term (concerning H4: $b = .057$, $[-.053, .094]$), although the direction of the effect was in the expected direction.

Our last hypotheses predicted that as the election approaches, the impact of preference homophily in predicting message selection dynamics would increase. Among tested interaction terms, only candidate choice homophily is found to significantly interact with time trends (Interaction model I: $b_{\text{interaction}}$ = .051, [.038, .071]). Specifically, the effect of candidate choice homophily is found to increase linearly over time, in a way that message selection in a dyad that shares the same candidate choice is more likely later in the election period, as plotted in Figure 3. Panel B of the Figure 2 gives Johnson-Neyman regions of significance as a function of time trends, additionally revealing that there is indeed a preference *towards heterophily* earlier in the election (as indicated in the negative conditional main effect: b = −.135, [−.211, −.111]). But this effect gradually disappears as a preference for the same candidate choice increases. No other interaction terms emerged as significant.

[Figure 2 About Here]

Discussion and Conclusion

Even though prior literature has emphasized the deliberative potential of online discussions (Papacharissi, 2004; Stromer-Galley, 2003), it is not uncommon to find worries about self-reinforcing political echo chambers. Since the debate on whether online settings promote more diverse and balanced exposure to political information is far from resolved, a more comprehensive understanding of the underlying motivational and structural factors that drive citizens' everyday discussion is much needed. Against this background, our study emphasizes *consistency* and *understanding* as the two core explanatory principles of online political discussion at individual- and dyadic-levels and highlights the role of various endogenous structural factors, stemming from the pattern of online discussion itself, as the crucial determinants of message-selection dynamics. This study is among the first to provide direct evidence that can disentangle the various determinants of message-selection decisions in an online discussion forum setting. Our findings suggest that, while there is some modest tendency for message selection to be based on *both* consistency and understanding motivations (especially at the individual level), the impact of *overt* partisan preference, as measured by candidate choice homophily and policy preference homophily, is fairly limited than is often assumed in prior literature. Instead, we have observed a meaningful pattern of message selection driven by a dyadic similarity in candidate evaluative criteria – in other words, the judgmental criteria with which citizens evaluate candidates – as well as robust effects of endogenous structural factors on message selection. These results yield significant new insights and add important nuance to our understanding of how people decide what to read in online discussion settings.

In particular, we have found that those with higher understanding motivation actively seek and expose themselves to others' messages. At the same time, those with high consistency motivation are more likely to be the *target* in message-selection dynamics (i.e., their messages are more likely to be selected by others). Yet, those with higher consistency motivation are not necessarily more likely to seek – presumably confirmatory – social information. Had they been more likely to seek social information, it would have indicated that those with higher consistency motivation seek and are sought by mostly like-minded individuals, providing support for the notion of ideological or partisan selectivity in online discussion settings. However, our results are more in line with Garrett (2009; also see Garrett et al., 2013) and Bakshy et al. (2015), who find that online settings leave substantial room for cross-cutting *exposure*. While our results also show that the preference for opinion-reinforcing information is real (as indicated by the effect of consistency motivation predicting *incoming* ties), this does not necessarily mean that people *only* seek confirmatory information.

More direct evidence supporting this perspective comes from the results of three dyadic-level effects. That is, overt partisan homophily – either based on more specific, concrete candidate/policy choice or based on more abstract ideological identifications – does not play a substantive role in message-selection dynamics. Although we have found that the impact of candidate choice homophily increased in a linear fashion over time, the magnitude of the overall effect was still limited. Instead, the similarity in candidate evaluation criteria has a substantial effect throughout all models. This is a particularly noteworthy finding given that such similarity in *judgmental standards* is not necessarily shaped by ideological or partisan like-mindedness. Rather, consistent with understanding motivation, the result suggests that a utility consideration – that is, the specific information an individual uses in candidate evaluations *irrespective of its potential valence and partisan views* – is one of the crucial factors determining message selection. Such interactions (through message selection) between discussants with similar evaluative criteria are likely to be exposed to diverse opinions and, thus, deliberative outcomes.[10] Therefore, the results at the dyadic level as a whole provide evidence that counters the notion of selectivity in online discussion settings toward ideologically like-minded messages *at the expense of* cross-cutting messages. Also, importantly for our purpose, this conclusion does not change when we use different threshold values as a cut-off in dichotomizing networks, models with daily slices instead of three-wave panel networks, or models with multiple imputation addressing missing data patterns (as reported in online supporting information), which speaks to the robustness of our findings against different methodological choices and analytical decisions.

Across our analyses, preferential attachment emerged as the strongest predictor of message-selection dynamics, corroborating recent studies about online (Himelboim, 2008; 2011) and offline political discussion (Author, 2015). Compared to studies of readily

---

[10] Although the effect of evaluative criteria similarity was more substantial between those who share the same candidate preference ($b_{\text{interaction}}$ = .324, [.039, .466]), the similarity in evaluative criteria had a significant and positive impact on the probability of message selection even among individuals who have different candidate preferences.

"visible" interactions, such as post-reply relationships (Himelboim, 2008; 2011), our behavioral log data show selection behaviors that are not necessarily observable to participants. This suggests that the global-level message-selection dynamics are likely to be, at least partly, driven by aggregate popularity cues (such as the number of "views" or "likes") that enable participants to identify messages of higher social and informational utility. However, these aggregate popularity cues do not necessarily signal whether a given message contains politically congenial messages. Considering that the magnitude of this preferential attachment effect is nearly ten times greater than any of the homophily factors, our results overall suggest that overt partisan considerations played a limited role in message selection dynamics than is often assumed in prior research (e.g., Bakshy et al., 2015; Himelboim et al., 2013). This echoes Messing and Westwood's (2014) finding on selective exposure dynamics on social networking sites.

Of our findings on the structural factors, the results – especially those about triadic configurations – warrant further discussion. Notably, we found significant and positive, yet small, "shared activity" and "shared popularity" effects. The patterns suggest that a pair of participants who viewed the same set of messages, or whose messages were viewed by the same set of people, are likely to see each other's messages. Within triadic settings such as these, particularly within the context of our study, cues indicating similar message-selection patterns between two individuals are not available unless the relationships being studied are also already visible to participants (such as in message – reply relationships). Therefore, our setting – which models "low visibility" message-selection behaviors – makes it particularly unlikely that these effects are driven by characteristics other than actual similarities in the participants' message selection criteria (and/or message writing).[11] At the same time, unlike

[11] The only possible exception would be the situation where other (multiple) third actors leave visible traces (such as comments), which in turn lead a given dyad to select each other's messages. Yet this possibility does not necessarily contradict our conclusions.

our dyadic homophily factors, which assume that participants select messages based on only a single characteristic, the extent of the similarities in *profiles* (i.e., message-selection patterns) does not necessarily exclude the possibility that each discussant has different reasons for selecting messages. Because of this, the notion of profile similarity predicts that those who have a similar "overall" pattern of message selection will eventually have a similar "set" of characteristics. As such, if *i* and *j* choose to see each other's messages based on the similarities of their message-selection patterns to/from all other actors *k*, then it implies that *i* and *j* have a great deal of common attributes (i.e., homophily). This raises the possibility that people may choose to engage with each other based not on just a single characteristic (such as candidate preference) but also on some balance (or a sum) of multiple characteristics ("multidimensional homophily": Block & Grund, 2014). Yet it should be acknowledged that, although this enables many attributes to be simultaneously involved in the consideration of homophily, it remains elusive whether such homophily is driven by overt partisan considerations or by other incidental factors. Even if we interpret these patterns as support for consistency-driven dynamics, the substantive magnitudes of such effects fall short of other understanding-driven factors.

In consideration of this study's findings, we conclude with a few caveats. First, although the coefficient was in the expected direction, we did not find the expected transitive closure effect. While we do not have any definitive explanation for this, it may be that local-level, hierarchy-based dynamics (as measured by a transitive closure effect) become non-significant when there is a strong influence of global-level hierarchies produced by preferential attachment (again, its impact is almost ten times greater than that of transitive closure). In ERGM, since both triadic closure and degree distributions lead to local clustering when they are highly correlated (Levy et al., 2015), strong global hierarchies produced by a degree-related effect may leave almost no room for a weak, local-level triadic closure term in explaining the emergence of a hierarchical network structure.

Second, following our theoretical focus, this study has operationalized “links” among participants as directed message “reading” behaviors. While this is an important addition to the existing literature, which focuses largely on either self-reports (e.g., Stromer-Galley, 2003; Wojcieszak & Mutz, 2009) or written (posted) messages (e.g., Himelboim, 2008; 2011), yet as stated earlier, our model did not consider the characteristics of the messages themselves. Indeed, it is conceivable that individual message-selection behaviors were at least partly driven by textual cues available in thread titles (as the first textual cue that respondents would encounter in selecting other’s messages) or by some interaction between message characteristics and the network dynamics identified here. Future studies would thus be well advised to consider such message (or textual) characteristics within the context of a systematic investigation of individuals’ message-selection decisions. Yet, while the characteristics of the messages themselves are arguably an important avenue for future research, it requires consideration of how latent textual topics and observed message quantities are probabilistically generated and how such factors would further condition the observed network dynamics in a stochastic fashion. To our knowledge, a proper probabilistic model addressing such issues is only now in development (e.g., Kim et al., 2017).

Third, regarding the interaction effect between candidate choice homophily and time trends, we also acknowledge that such patterns may have been driven by participants’ “learning effects,” rather than the effects of campaign competitiveness. That is, based on their continued interactions in the forum, participants could have learned about others’ partisan orientations, making them better able to discern the partisan leanings of messages over time. While our finding could be explained by this alternative explanation, it should be also noted that, ultimately, the impact of candidate choice homophily never exceeded that of other understanding-based effects. As such, the overall results speak to the conclusion that individuals' message-selection (and, thus, exposure) patterns are not necessarily self-segregated along overt partisan lines.

Lastly, we close by recognizing that our single-country, single-election approach may not generalize to other contexts. Yet, given how similar our results are to those of other online (Himelboim, 2008; 2011) and offline (Song, 2015) political discussion studies from considerably different geographical and electoral contexts, we see little reason to expect that the basic underlying mechanisms identified in our study would not be applicable in different times and contexts.

Throughout this paper, we have highlighted the notion that online discussion settings do not necessarily create polarized message-selection patterns because fundamental human motivations – *consistency* vs. *understanding* – play important roles in structuring the way people decide to initiate a communicative interaction by reading others' messages. Consistent with previous evidence (Garrett & Stroud, 2014; Messing & Westwood, 2014), we find that individuals do not organize their message selection based solely on overt partisan considerations. Instead, message-selection patterns in line with understanding goals have been observed. Further, echoing evidence by Lazer et al. (2010) and Song (2015), our results demonstrate that the endogenous structures of an online discussion network, which have less to do with individuals' overt partisan preferences and directional goals, systematically shape individuals' message-selection behaviors. This means that discussants have accidental exposure to cross-cutting political messages. Thus, while it is still possible to isolate oneself from different perspectives online, this study suggests it is not an unavoidable consequence of conscious individual choice.

References

Ahn, T. K., Huckfeldt, R., & Ryan, J. B. (2014). *Experts, activists, and interdependent citizens: Are electorates self-educating?* New York: Cambridge University Press.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*, 1130-1132. doi:10.1126/science.aaa1160

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*, 509-512. doi:10.1126/science.286.5439.509

Block, P., & Grund, T. (2014). Multidimensional homophily in friendship networks. *Network Science, 2*, 189–212. doi:10.1017/nws.2014.17

Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology, 38*, 551-569. doi:10.1111/pops.12337

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*, 197-253. doi:10.1037/0033-2909.119.2.197

Carnahan, D., Garrett, R. K., & Lynch, E. K. (2016). Candidate vulnerability and exposure to counterattitudinal information: Evidence from two US presidential elections. *Human Communication Research*, *42*, 577-598. doi:10.1111/hcre.12088

Cranmer, S. J., Leifeld, P., McClurg, S. D., & Rolfe, M. (2017). Navigating the range of statistical tools for inferential network analysis. *American Journal of Political Science*, *61*, 237-251. doi:10.1111/ajps.12263

Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication, 22*, 147-162. doi:10.1080/10584600590933160

Dekker, D., Krackhardt, D., & Snijders, T. A. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika, 72*, 563-581. doi: 10.1007/s11336-007-9016-1

Desmarais, B. A., & Cranmer, S. J. (2012). Statistical mechanics of networks: Estimation and uncertainty. *Physica A*, *391*, 1865-1876. doi:10.1016/j.physa.2011.10.018

DiMaggio, P. (1986). Structural analysis of organizational fields: A blockmodel approach. *Research in Organizational Behavior, 8*, 335-370.

Downs, A. (1957). *An economic theory of democracy*. New York: Harper.

Fournier, P., Nadeau, R., Blais, A., Gidengil, E., & Nevitte, N. (2004). Time-of-voting decision and susceptibility to campaign effects. *Electoral Studies, 23*, 661-681. doi:10.1016/j.electstud.2003.09.001

Garrett, R. K. (2009). Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication*, *59*, 676-699. doi:10.1111/j.1460-2466.2009.01452.x

Garrett, R. K., & Stroud, N. J. (2014). Partisan paths to exposure diversity: Differences in pro-and counterattitudinal news consumption. *Journal of Communication*, *64*, 680-701. doi:10.1111/j.1460-2466.2009.01452.x

Gentzkow, M., Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics, 126*, 1799-1839. doi:10.1093/qje/qjr044

Graham, T., & Wright, S. (2014). Discursive equality and everyday talk online: The impact of "superparticipants". *Journal of Computer-Mediated Communication*, *19*, 625-642. doi:10.1111/jcc4.12016

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*, 555-588. doi:10.1037/a0015701

Hillygus, D. S., Jackson, N., & Young, M. (2014). Professional respondents in non-probability online panels. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online panel research: A data quality perspective* (pp. 219-237). Chichester, UK: Wiley.

Himelboim, I. (2008). Reply distribution in online discussions: A comparative network analysis of political and health newsgroups. *Journal of Computer-Mediated Communication, 14*, 156-177. doi:10.1111/j.1083-6101.2008.01435.x

Himelboim, I. (2011). Civil society and online political discourse: The network structure of unrestricted discussions. *Communication Research, 38*, 634-659. doi:10.1177/0093650210384853

Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Journal of Computer-Mediated Communication, 18*, 154-174. doi: 10.1111/jcc4.12001

Holbert, R. L., Weeks, B. E., & Esralew, S. (2013). Approaching the 2012 U.S. Presidential election from a diversity of explanatory principles: Understanding, Consistency, and Hedonism. *American Behavioral Scientist, 57*, 1663-1687. doi:10.1177/0002764213490693

Holland, P. W., & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology, 7*, 1-45. doi:10.2307/270703

Huckfeldt, R. (2001). The social communication of political expertise. *American Journal of Political Science*, *45*, 425-438. doi:10.2307/2669350

Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, *15*, 565-583. doi:10.1198/106186006X133069

Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, *59*, 19-39. doi:10.1111/j.1460-2466.2008.01402.x

Jacoby, W. G. (1991). Ideological identification and issue attitudes. *American Journal of Political Science, 35*, 178-205. doi:10.2307/2111443

Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: Its structure, functions, and elective affinities. *Annual Review of Psychology, 60*, 307-337. doi: 10.1146/annurev.psych.60.110707.163600

Kim, B., Schein, A., Desmarais, B. A., & Wallach, H. (2017). *A network model for dynamic textaul communications with appliation to government email corpora*. Paper presented at the 10th Political Networks Conference, The Ohio State University, Columbus, OH.

Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network. *American Journal of Sociology*, *115*, 405-450. doi:10.1086/599247

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480-498. doi:10.1037/0033-2909.108.3.480

Lazega, E., Mounier, L., Snijders, T., & Tubaro, P. (2012). Norms, status and the dynamics of advice networks: A case study. *Social Networks, 34*, 323-332. doi:10.1016/j.socnet.2009.12.001

Lazer, D., Rubineau, B, Chetkovich, C., Katz, N., & Neblo, M. (2010). The coevolution of networks and political attitudes. *Political Communication, 27*, 248–274. doi:10.1080/10584609.2010.500187

Leifeld, P., Cranmer, S. J., & Desmarais, B. A. (2017). *btergm*. Temporal Exponential Random Graph Models by Bootstrapped Pseudolikelihood. R package version 1.9.0.

Levy, M., Lubell, M., Leifeld, P., & Cranmer, S. (2016). *Interpretation of geometrically* *weighted degree estimates in exponential random graph models*. Paper presented at the 9th Political Networks Conference, Washington University, Saint Louis, MO.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415-444. doi:10.1146/annurev.soc.27.1.415

Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, *41*, 1042-1063. doi:10.1177/0093650212466406

Mondak, J. J. (2010). *Personality and the foundations of political behavior*. New York: Cambridge University Press.

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*, 259-283. doi:10.1177/1461444804041444

Pietryka, M. T. (2016). Accuracy motivations, predispositions, and social information in political discussion networks. *Political Psychology, 37*, 367-386. doi:10.1111/pops.12255

Prior, M. (2009). The immensely inflated news audience: Assessing bias in self-reported news exposure. *Public Opinion Quarterly, 73*, 130-143. doi:10.1093/poq/nfp002

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (*p**) models for social networks. *Social Networks, 29*, 173–191. doi:10.1016/j.socnet.2006.08.002

Rudolph, T. J., & Popp, E. (2007). An information processing theory of ambivalence. *Political Psychology, 28*, 563-585. doi:10.1111/j.1467-9221.2007.00590.x

Schmitt-Beck, R., & Lup, O. (2013). Seeking the soul of democracy: A review of recent research into citizens' political talk culture. *Swiss Political Science Review, 19*, 513-

538. doi:10.1111/spsr.12051

Shook, N. J., & Fazio, R. H. (2009). Political ideology, exploration of novel stimuli, and attitude formation. *Journal of Experimental Social Psychology, 45*, 995-998. doi:10.1016/j.jesp.2009.04.003

Snijders, T. A., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology, 36*, 99-153. doi:10.1111/j.1467-9531.2006.00176.x

Song, H. (2015). Uncovering the structural underpinnings of political discussion networks: Evidence from an exponential random graph model. *Journal of Communication, 65*, 146-169. doi:10.1111/jcom.12140

Stromer-Galley, J. (2003). Diversity of political conversation on the Internet: Users' perspectives. *Journal of Computer-Mediated Communication*, *8*. doi:10.1111/j.1083-6101.2003.tb00215.x

Stroud, N. J. (2011). *Niche news: The politics of news choice.* Oxford University Press.

Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton University Press.

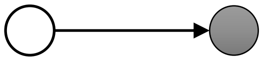
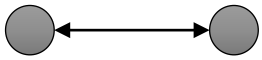
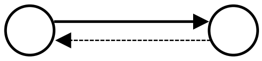
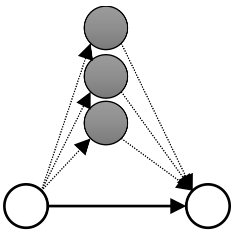
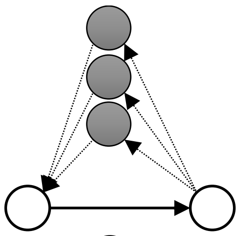
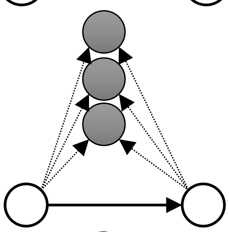
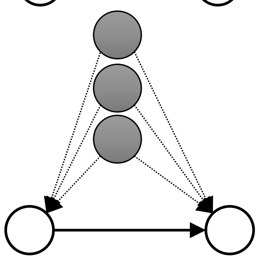
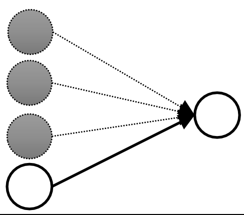
Valentino, N. A., Banks, A. J., Hutchings, V. L., & Davis, A. K. (2009). Selective exposure in the Internet age: The interaction between anxiety and information utility. *Political Psychology*, *30*, 591-613. doi:10.1111/j.1467-9221.2009.00716.x

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Boston, MA: Cambridge University Press.

Wojcieszak, M. E., & Mutz, D. C. (2009). Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement? *Journal of Communication*, *59*, 40-56. doi:10.1111/j.1460-2466.2008.01403.x

Table 1

Key TERGM Parameters, Associated Configurations, and Their Interpretations

Hypothesis	Configuration	Interpretation
RQ: Motivation		A select B's message (B's message is selected by A) based on nodal attributes
H1 & H2: Homophily		A and B select each other's message based on their shared characteristics
H3: Reciprocity		A select B's message when B also select A's message
H4: Multiple path closure (GWESP-OTP)		A select B's message when A has multiple intermediary actors that also leads to B (implies status differentials)
H4: Multiple cyclic closure (GWESP-ITP)		A select B's message when B has multiple intermediary actors that also leads to A (implies lack of status differential)
H5: Multiple activity closure (GWESP-OSP)		A select B's message when they have similar patterns of message selection patterns (implies similarity in latent attributes)
H5: Multiple popularity closure (GWESP-ISP)		A select B's message when their messages are similarly selected by others (implies similarity in latent attributes)
H6: Preferential attachment (GWD-in)		A select B's message when many others also selected B's message

Note. Preferential attachment is measured using geometrically weighted in-degree distribution statistics, which measures *unevenness* of in-degree distribution. Therefore, *negative* GWD-in statistic means *positive* preferential attachment pattern (Levy et al., 2015).

Table 2.

Predicting Dependent Network as a Function of Dyadic-level Predictors, Bivariate QAP-Logit Regression Results

	DV: Message selection network (X_{ij} = 0 vs. 1)							
	IV: Same candidate preference		**IV:** Policy preference similarity		**IV:** Ideological placement dissimilarity		**IV:** Evaluation criteria similarity	
	b	***Pr* ≥ (|*b*|)**	***b***	***Pr* ≥ (|*b*|)**	***b***	***Pr* ≥ (|*b*|)**	***b***	***Pr* ≥ (|*b*|)**
T = 1	.1234	.222	-0.5281	.412	-.0308	.756	.9635	.180
T = 2	.0716	.691	-2.0456	**.005**	.0384	.690	1.0024	.191
T = 3	.0934	.507	-0.3777	.629	.0785	.474	1.3693	.103

Note: *b* = unstandardized regression coefficient, where models include only intercept and a respective predictor variable. We used the double semi-partialing permutation with 1,000 replications for deriving probabilities of observed regression coefficient (*b*) exceeding the either lower or upper tails of the simulated null distribution based on double semi-partialing permutation at .05 level (denoted as *Pr* ≥ (|*b*|) above).

Table 3

Bootstrapped TERGM Estimates (95% Bias-Corrected and Accelerated Confidence Intervals Within Brackets)

	Final Model I	Final Model II	Interaction I	Interaction II
Edges (Intercept)	**-1.890** [-2.932; -1.392]*	**-2.259** [-2.958; -1.732]*	**-1.819** [-2.732; -.304]*	**-2.150** [-2.938; -1.630]*
<u>*Motivation and homophily*</u>				
Consistency, in-ties (RQ)	**.034** [.009; .113]*	**.062** [.047; .144]*	.037 [-.004; .113]	**.065** [.021; .144]*
Consistency, out-ties (RQ)	.025 [-.044; .077]	**.046** [.007; .046]*	.019 [-.112; .071]	**.039** [.039; .039]*
Understanding, in-ties (RQ)	-.052 [-.080; .022]	**-.078** [-.118; -.011]*	-.049 [-.103; .022]	**-.075** [-.094; -.011]*
Understanding, out-ties (RQ)	**.028** [.005; .076]*	**.056** [.047; .092]*	**.036** [.012; .075]*	**.064** [.049; .065]*
Same candidate preference (H1)	-.032 [-.070; .047]		**-.135** [-.211; -.111]*	
Similar policy preference (H1)	-.108 [-.212; .006]		-.091 [-.225; .042]	
Dissimilar ideological placement (H1)		.024 [-.007; .040]		-.013 [-.134; .028]
Similar evaluative criteria (H2)	**.407** [.399; .415]*	**.482** [.398; .482]*	**.385** [.260; .404]*	**.465** [.350; .481]*
<u>*Interaction (H7)*</u>				
Time trends (linear)			.079 [-.059; .262]	**.074** [.016; .133]*
x Same candidate preference			**.051** [.038; .071]*	
x Dissimilar ideological placement				.019 [-.016; .063]
<u>*Endogenous structural effects*</u>				
Reciprocity (H3)	**.768** [.560; 1.068]*	**.720** [.504; 1.069]*	**.768** [.559; 1.068]*	**.719** [.501; 1.069]*
Path closure (gwesp-OTP: H4)	.057 [-.053; .094]	.040 [-.078; .067]	.057 [-.053; .125]	.039 [-.078; .066]
Cyclic closure (gwesp-ITP: H4)	**-.066** [-.076; -.061]*	**-.058** [-.077; -.047]*	**-.066** [-.076; -.061]*	**-.057** [-.077; -.048]*
Activity closure (gwesp-OSP: H5)	**.035** [.033; .043]*	**.027** [.026; .033]*	**.035** [.033; .041]*	**.027** [.027; .032]*
Popularity closure (gwesp-ITP: H5)	**.113** [.083; .232]*	**.129** [.099; .286]*	**.113** [.083; .232]*	**.128** [.100; .286]*
Popularity spread (gw-indegree: H6)	**-4.123** [-5.342; -3.541]*	**-4.828** [-5.429; -3.889]*	**-4.120** [-5.342; -3.537]*	**-4.898** [-5.435; -3.889]*

Note. * = zero outside the 95% bias-corrected and accelerated confidence interval using 1000 replications. Significant results are bolded. All models control for age, gender (including homophily), education, regional origins (including homophily), offline talk frequency, media use frequency, internal discussion efficacy, candidate preference, hedonic motivations, Activity spread (gw-outdegree), being isolate, multiple two-paths (gwdsp), and lagged versions of network-endogenous statistics (previous communication, delayed reciprocity, delayed transitivity, delayed cyclic closure, delayed activity closure, delayed popularity closure, and number of in- and out ties of a given nodes at previous time point). Full results, including interaction models, are reported in the Online Supporting Information.

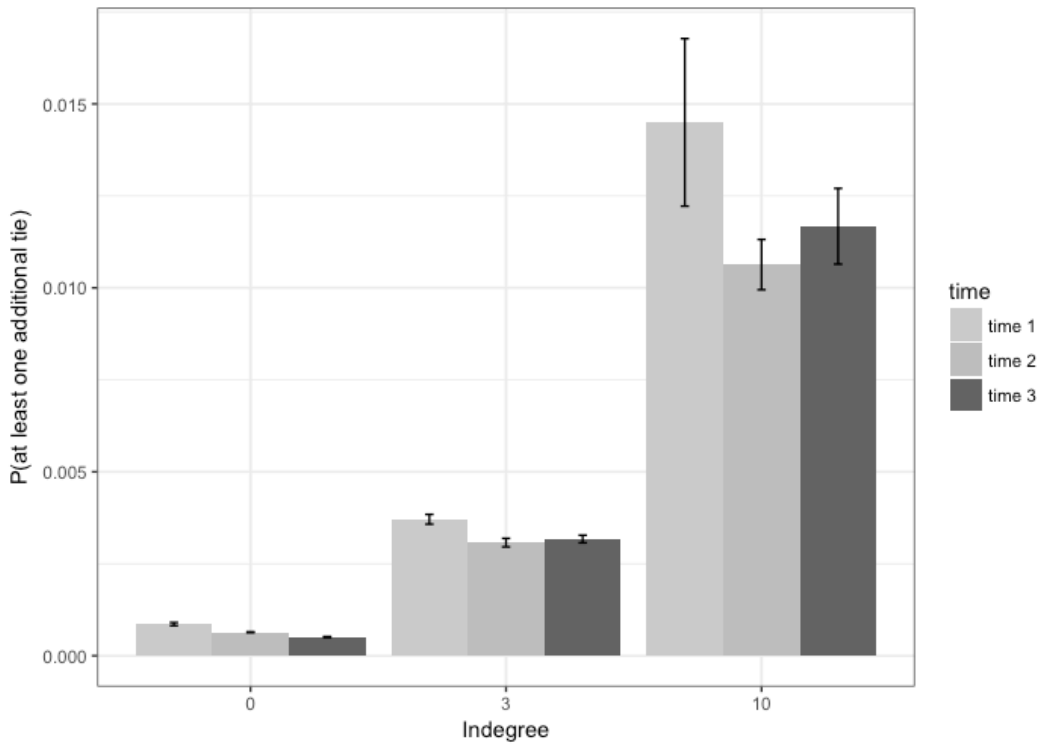


Figure 1. Mean predicted probabilities of *receiving* at least one additional tie (i.e., message being selected by others) as a function of existing incoming ties at 10% (= zero), 50% (= three), and 90% (= ten existing ties) percentile of the in-degree distribution. For each receiver node, we derived the mean edge probabilities of all other nodes (excluding any nodes that are already connected) sending a tie to the target node conditional on the rest of the network and on the model specification.

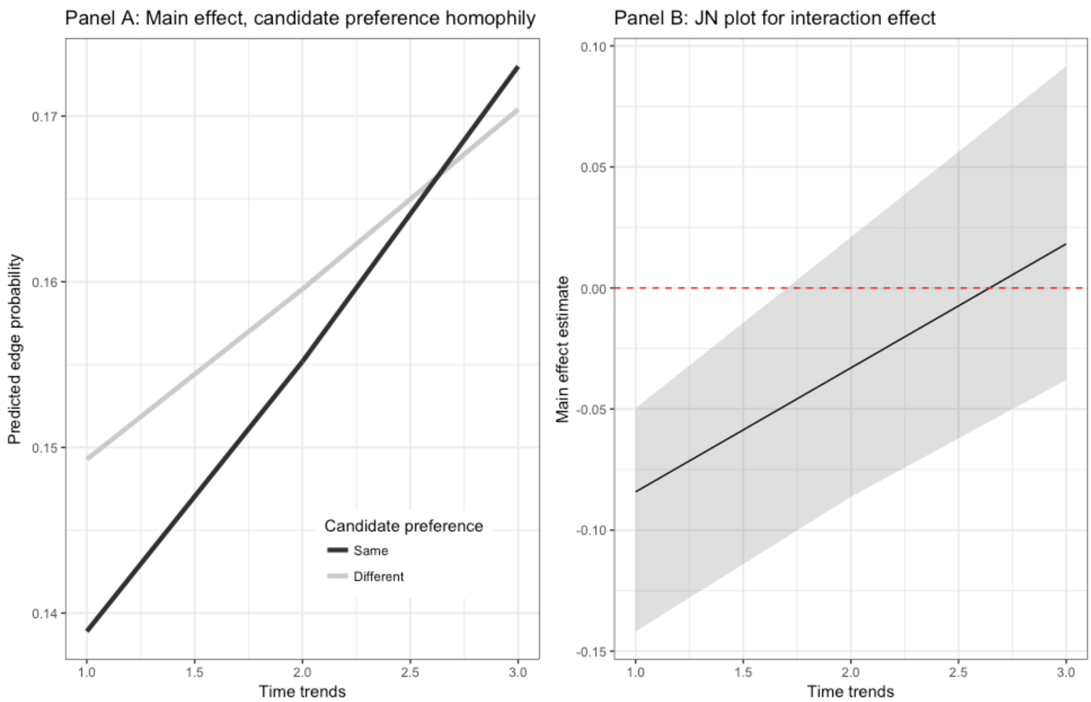


Figure 2. Interaction effects between time trends and candidate preference homophily. Panel A depicts conditional main effects of candidate preference homophily at each time point, and Panel B depicts Johnson-Neyman regions of significance as a function of time.

Online Supporting Information to

"The Dynamics of Message Selection in Online Political Discussion Forums"

Contents

Inclusion of the control terms in the TERGM model

All models reported in Table 3 of the main document control for age, gender (including homophily), education, regional origins (including homophily), offline talk frequency, media use frequency, internal discussion efficacy, candidate preference, hedonic motivations, activity spread (gw-outdegree), isolate, and multiple two-paths (gwdsp), as well as lagged versions of network-endogenous statistics (previous communication, delayed reciprocity, delayed transitivity, delayed cyclic closure, delayed activity closure, delayed popularity closure, and number of in- and out ties of a given nodes at previous time point). Here, we control for (a) all possible temporal dependencies in a form of lagged structural variables (which closely resemble the concurrent structural terms), and (b) other covariates that help control lower-order effects in estimating higher-order effect (e.g., GWDSP), and (c) control for basic degree effects and densities (edge and isolate parameter together, given non-negligible number of isolates in each time period).

Table S1. Full TERGM results including model building procedures

	Control only	Control + Structural	Final Model I	Final Model II
Edges (Intercept)	**-4.977** [-6.749; -6.749]*	**-1.127** [-2.206; -2.206]*	**-1.890** [-2.932; -1.392]*	**-2.259** [-2.958; -1.732]*
Motivation and Homophily				
Consistency motivation (in-ties)			**.034** [.009; .113]*	**.062** [.047; .144]*
Consistency motivation (out-ties)			.025 [-.044; .077]	**.046** [.007; .046]*
Understanding motivation (in-ties)			-.052 [-.080; .022]	**-.078** [-.118; -.011]*
Understanding motivation (out-ties)			**.028** [.005; .076]*	**.056** [.047; .092]*
Hedonic motivation (in-ties)			-.012 [-.029; .001]	**-.026** [-.039; -.021]*
Hedonic motivation (out-ties)			**.102** [.087; .133]*	.043 [-.009; .114]
Same candidate pref			-.032 [-.070; .047]	
Similar policy pref			-.108 [-.212; .006]	
Dissimilar ideology (*absdiff*)				.024 [-.007; .040]
Similar evaluative criteria			**.407** [.399; .415]*	**.482** [.398; .482]*
Endogenous structural effects				
Isolates		**1.021** [.797; .797]*	**1.019** [.908; 1.264]*	**.825** [.598; 1.224]*
Reciprocity		**.765** [.497; .497]*	**.769** [.564; 1.068]*	**.720** [.504; 1.069]*
Multiple path closure (GWESP-OTP)		**.058** [-.056; -.056]*	.058 [-.053; .125]	.040 [-.078; .067]
Multiple cyclic closure (GWESP-ITP)		**-.068** [-.082; -.082]*	**-.066** [-.080; -.060]*	**-.058** [-.077; -.047]*
Multiple activity closure (GWESP-OSP)		**.035** [.030; .030]*	**.036** [.033; .045]*	**.027** [.026; .033]*
Multiple popularity closure (GWESP-ISP)		**.117** [.083; .083]*	**.115** [.093; .232]*	**.129** [.099; .286]*
Multiple two-paths (GWDSP)		**.003** [-.005; -.005]*	.003 [-.007; .007]	**.008** [.002; .016]*
Activity spread (GW-outdegree)		**-4.399** [-4.669; -4.669]*	**-4.350** [-4.557; -4.157]*	**-3.923** [-4.124; -3.623]*
Popularity spread (GW-indegree)		**-4.056** [-5.343; -5.343]*	**-4.049** [-5.342; -3.259]*	**-4.828** [-5.429; -3.889]*
Lagged structural effects				
Previous communication		**.214** [.182; .182]*	**.222** [.192; .253]*	**.531** [.494; .549]*
Delayed reciprocity		**.082** [-.067; -.067]*	.074 [-.073; .194]	.002 [-.197; .139]
Delayed transitivity closure		**.034** [.018; .018]*	**.034** [.020; .055]*	**.032** [.026; .037]*
Delayed cyclic closure		**.037** [.010; .010]*	**.034** [.008; .057]*	**.032** [.003; .035]*
Delayed activity closure		**-.058** [-.068; -.068]*	**-.056** [-.067; -.046]*	**-.061** [-.071; -.037]*
Delayed popularity closure		**-.060** [-.089; -.089]*	**-.059** [-.110; -.043]*	**-.062** [-.124; -.046]*

Persistent sender (out-tie)		**.019** [.009; .009]*	**.019** [.010; .029]*	**.510** [.176; .724]*
Persistent receiver (in-ties)		**.023** [.019; .019]*	**.023** [.018; .038]*	**.116** [.020; .165]*
Controls				
Age (in-ties)	**.101** [-.012; -.012]*	**.003** [-.017; -.017]*	.001 [-.020; .022]	**-.025** [-.033; -.006]*
Age (out-ties)	**.218** [-.097; -.097]*	**.031** [-.224; -.224]*	.052 [-.105; .093]	.101 [-.046; .141]
Female (in-ties)	**-.204** [-.245; -.245]*	**-.001** [-.038; -.038]*	.005 [-.036; .041]	-.023 [-.057; .030]
Female (out-ties)	**-.169** [-.446; -.446]*	**.075** [-.308; -.308]*	.014 [-.348; .254]	.151 [-.266; .389]
Gender homophily	**.010** [-.032; -.032]*	**.051** [.018; .018]*	**.044** [.023; .086]*	**.045** [.021; .100]*
Education (in-ties)	**-.114** [-.182; -.182]*	**-.008** [-.042; -.042]*	-.011 [-.039; .019]	-.015 [-.047; .014]
Education (out-ties)	**-.132** [-.239; -.239]*	**.028** [-.010; -.010]*	.016 [-.015; .091]	-.044 [-.096; .029]
Regional origin = Seoul (in-ties)	**-.418** [-.501; -.501]*	**-.077** [-.116; -.116]*	-.084 [-.130; .044]	**-.109** [-.140; -.030]*
Regional origin = Seoul (out-ties)	**-.192** [-.383; -.383]*	**-.143** [-.635; -.635]*	-.125 [-.438; .350]	-.135 [-.435; .324]
Regional homophily (Seoul)	**-.021** [-.047; -.047]*	**.013** [-.020; -.020]*	.017 [-.014; .080]	.025 [-.002; .049]
Talk freq (in-ties)	**.129** [-.120; -.120]*	**.045** [.021; .021]*	**.046** [.024; .049]*	**.051** [.048; .051]*
Talk freq (out-ties)	**.025** [-.428; -.428]*	**.034** [-.173; -.173]*	.014 [-.099; .161]	.010 [-.111; .206]
Media use (in-ties)	**-.061** [-.108; -.108]*	**-.011** [-.021; -.021]*	**-.011** [-.019; -.003]*	-.006 [-.012; .004]
Media use (out-ties)	**-.070** [-.104; -.104]*	**.040** [.004; .004]*	.033 [-.017; .071]	.043 [-.006; .078]
Internal efficacy (in-ties)	**.051** [-.045; -.045]*	**-.013** [-.040; -.040]*	-.013 [-.058; .055]	**-.030** [-.072; -.006]*
Internal efficacy (out-ties)	**.187** [.132; .132]*	**-.018** [-.098; -.098]*	.024 [-.102; .128]	**.115** [.051; .163]*
Candidate pref = Moon (in-ties)	**.174** [.057; .057]*	**-.018** [-.063; -.063]*	.003 [-.008; .092]	-.003 [-.056; .085]
Candidate pref = Moon (out-ties)	**.315** [.216; .216]*	**-.010** [-.100; -.100]*	.013 [-.123; .066]	.080 [-.004; .100]
Num. obs.	291096	291096	291096	291096

* = zero outside the 95% bias-corrected and accelerated confidence interval based on 1000 replications.
Note: Decay (alpha) values for each geometrically weighted term are: GWESP-OTP = 3, GWESP-ITP = 3, GWESP-OSP = 3, GWESP-ISP = 2, GWDSP = 1, GW-outdegree = 2, and GW-indegree = 3. “Final model” denotes the final model reported in the Table 3 in the main manuscript.

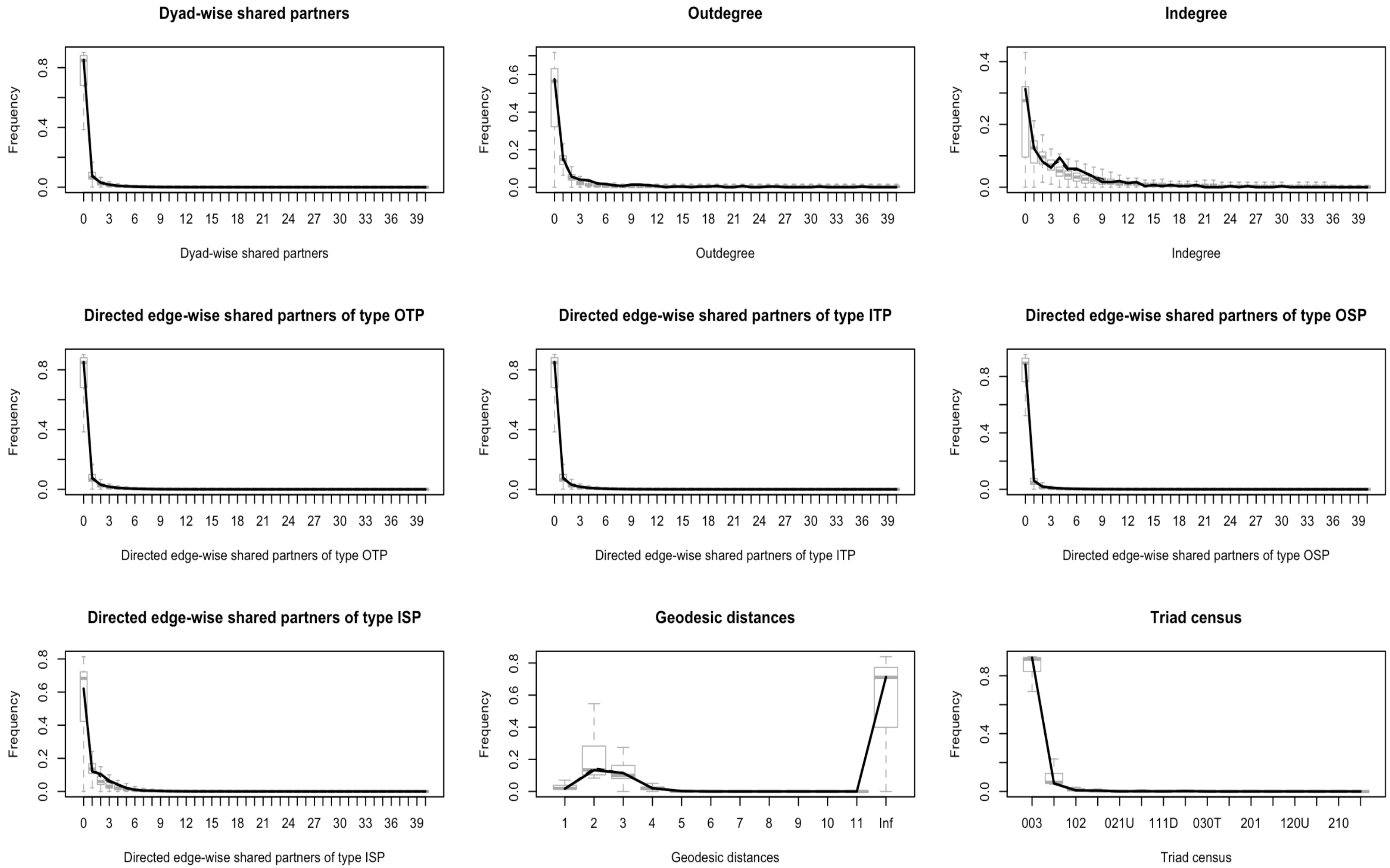
Table S2. Full TERGM results including interactions

	Interaction I (candidate preference)	(evaluation criteria)	(policy preference)	Interaction II (ideology)
Edges (Intercept)	**-1.819** [-2.732; -.304]*	**-1.823** [-2.807; -1.169]*	**-1.936** [-2.937; -1.098]*	**-2.150** [-2.938; -1.630]*
Motivation and Homophily				
Consistency motivation (in-ties)	.037 [-.004; .113]	**.037** [.010; .113]*	**.037** [.010; .113]*	**.065** [.021; .144]*
Consistency motivation (out-ties)	.019 [-.112; .071]	.019 [-.112; .071]	.019 [-.043; .071]	**.039** [.039; .039]*
Understanding motivation (in-ties)	-.049 [-.103; .022]	-.049 [-.103; .022]	-.049 [-.078; .022]	**-.075** [-.094; -.011]*
Understanding motivation (out-ties)	**.036** [.012; .075]*	**.035** [.011; .087]*	**.035** [.011; .075]*	**.064** [.049; .065]*
Hedonic motivation (in-ties)	-.012 [-.038; .001]	-.013 [-.032; .001]	-.013 [-.038; .001]	**-.026** [-.049; -.022]*
Hedonic motivation (out-ties)	**.102** [.094; .130]*	**.102** [.096; .130]*	**.102** [.094; .105]*	.042 [-.005; .110]
Same candidate pref	**-.135** [-.211; -.111]*	-.033 [-.079; .047]	-.032 [-.079; .047]	
Similar policy pref	-.091 [-.225; .042]	-.090 [-.230; .042]	.094 [-.764; .272]	
Dissimilar ideology (*absdiff*)				-.013 [-.134; .028]
Similar evaluative criteria	**.385** [.260; .404]*	.295 [-.359; .639]	**.389** [.255; .405]*	**.465** [.350; .481]*
Interactions				
time trends (linear)	.079 [-.059; .262]	**.083** [.021; .171]*	**.144** [.063; .235]*	**.074** [.016; .133]*
time X same candidate preference	**.051** [.038; .071]*			
time X evaluative criteria similarity		.046 [-.176; .242]		
time X policy preference similarity			-.095 [-.253; .214]	
time X dissimilar ideology				.019 [-.016; .063]
Endogenous structural effects				
Isolates	**1.003** [.793; 1.264]*	**1.005** [.793; 1.152]*	**1.005** [.895; 1.264]*	**.811** [.518; 1.030]*
Reciprocity	**.768** [.560; 1.068]*	**.768** [.559; 1.068]*	**.768** [.507; 1.068]*	**.719** [.501; 1.069]*
Multiple two-paths (GWDSP, 1)	.003 [-.007; .007]	.003 [-.007; .007]	.003 [-.007; .009]	.007 [-.001; .013]
Multiple path closure (GWESP-OTP)	.057 [-.053; .094]	.057 [-.053; .125]	**.057** [.025; .125]*	.039 [-.078; .066]
Multiple cyclic closure (GWESP-ITP)	**-.066** [-.076; -.061]*	**-.066** [-.076; -.061]*	**-.066** [-.080; -.061]*	**-.057** [-.077; -.048]*
Multiple activity closure (GWESP-OSP)	**.035** [.033; .043]*	**.035** [.033; .041]*	**.035** [.033; .043]*	**.027** [.027; .032]*
Multiple popularity closure (GWESP-ISP)	**.113** [.083; .232]*	**.113** [.083; .232]*	**.113** [.098; .232]*	**.128** [.100; .286]*
Activity spread (GW-outdegree)	**-4.395** [-4.557; -4.153]*	**-4.392** [-4.557; -4.152]*	**-4.392** [-4.557; -3.994]*	**-3.968** [-4.165; -3.627]*
Popularity spread (GW-indegree)	**-4.123** [-5.342; -3.541]*	**-4.120** [-5.342; -3.537]*	**-4.121** [-4.810; -3.259]*	**-4.898** [-5.435; -3.889]*
Lagged structural effects				

Previous communication	**.220** [.184; .250]*	**.220** [.184; .250]*	**.219** [.185; .250]*	**.528** [.507; .549]*
Delayed reciprocity	.076 [-.073; .289]	.075 [-.073; .257]	.076 [-.073; .257]	.001 [-.197; .139]
Delayed transitivity closure	**.033** [.019; .051]*	**.033** [.019; .051]*	**.033** [.019; .051]*	**.031** [.025; .036]*
Delayed cyclic closure	**.032** [.008; .041]*	**.032** [.008; .057]*	**.032** [.008; .043]*	**.031** [.003; .035]*
Delayed activity closure	**-.055** [-.060; -.035]*	**-.055** [-.065; -.035]*	**-.055** [-.065; -.035]*	**-.061** [-.072; -.037]*
Delayed popularity closure	**-.058** [-.081; -.034]*	**-.058** [-.110; -.043]*	**-.058** [-.081; -.034]*	**-.060** [-.124; -.046]*
Persistent sender (out-tie)	**.019** [.010; .029]*	**.019** [.010; .025]*	**.019** [.010; .025]*	**.506** [.176; .724]*
Persistent receiver (in-ties)	**.023** [.018; .038]*	**.023** [.018; .038]*	**.023** [.021; .038]*	**.112** [.059; .165]*
Controls				
Age (in-ties)	-.003 [-.023; .020]	-.003 [-.022; .035]	-.003 [-.022; .020]	**-.028** [-.034; -.017]*
Age (out-ties)	.040 [-.192; .091]	.040 [-.112; .090]	.040 [-.113; .090]	.089 [-.049; .134]
Female (in-ties)	.009 [-.037; .043]	.009 [-.036; .071]	.009 [-.036; .071]	-.021 [-.054; .012]
Female (out-ties)	.029 [-.348; .268]	.029 [-.348; .268]	.029 [-.348; .335]	.162 [-.266; .397]
Gender homophily	**.044** [.015; .070]*	**.044** [.015; .086]*	**.044** [.022; .086]*	**.044** [.019; .100]*
Education (in-ties)	-.010 [-.029; .019]	-.010 [-.029; .019]	-.010 [-.029; .018]	-.014 [-.046; .014]
Education (out-ties)	.015 [-.016; .073]	.015 [-.016; .072]	.015 [-.016; .071]	-.046 [-.096; .029]
Regional origin = Seoul (in-ties)	-.083 [-.157; .044]	-.084 [-.131; .044]	**-.084** [-.157; -.031]*	**-.109** [-.128; -.030]*
Regional origin = Seoul (out-ties)	-.143 [-.598; .350]	-.142 [-.450; .350]	-.143 [-.449; .350]	-.152 [-.675; .324]
Regional homophily (Seoul)	.015 [-.014; .048]	.015 [-.014; .080]	.015 [-.014; .080]	.023 [-.002; .048]
Talk freq (in-ties)	**.030** [.018; .037]*	**.030** [.018; .036]*	**.030** [.002; .037]*	**.036** [.034; .036]*
Talk freq (out-ties)	-.005 [-.097; .161]	-.006 [-.130; .161]	-.006 [-.143; .110]	-.009 [-.110; .206]
Media use (in-ties)	**-.018** [-.024; -.002]*	**-.018** [-.024; -.002]*	-.018 [-.024; .000]	**-.013** [-.021; -.008]*
Media use (out-ties)	**.024** [.001; .287]*	.024 [-.017; .075]	.024 [-.017; .074]	.034 [-.006; .063]
Internal efficacy (in-ties)	-.012 [-.058; .055]	-.012 [-.058; .055]	-.012 [-.042; .055]	**-.029** [-.052; -.001]*
Internal efficacy (out-ties)	.030 [-.102; .128]	.031 [-.064; .128]	.031 [-.102; .128]	**.122** [.051; .140]*
Candidate pref = Moon (in-ties)	.006 [-.008; .049]	.004 [-.008; .092]	.003 [-.008; .092]	-.004 [-.056; .055]
Candidate pref = Moon (out-ties)	.017 [-.123; .070]	.017 [-.123; .070]	.016 [-.063; .131]	**.085** [.039; .121]*
Num. obs.	291096	291096	291096	291096

* = zero outside the 95% bias-corrected and accelerated confidence interval based on 1000 replications.
Note: Decay (alpha) values for each geometrically weighted term are: GWESP-OTP = 3, GWESP-ITP = 3, GWESP-OSP = 3, GWESP-ISP = 2, GWDSP = 1, GW-outdegree = 2, and GW-indegree = 3. "Interaction I" and "Interaction II" denote the final interaction models reported in the Table 3 in the main manuscript.

Figure S1. The Goodness-of-fit (*gof*) assessment of final model specification



Note: The distribution of network statistics from the simulated networks (N = 900) does not significantly deviate from that of the observed statistic (bold line), suggesting that model fit is acceptable and adequate.

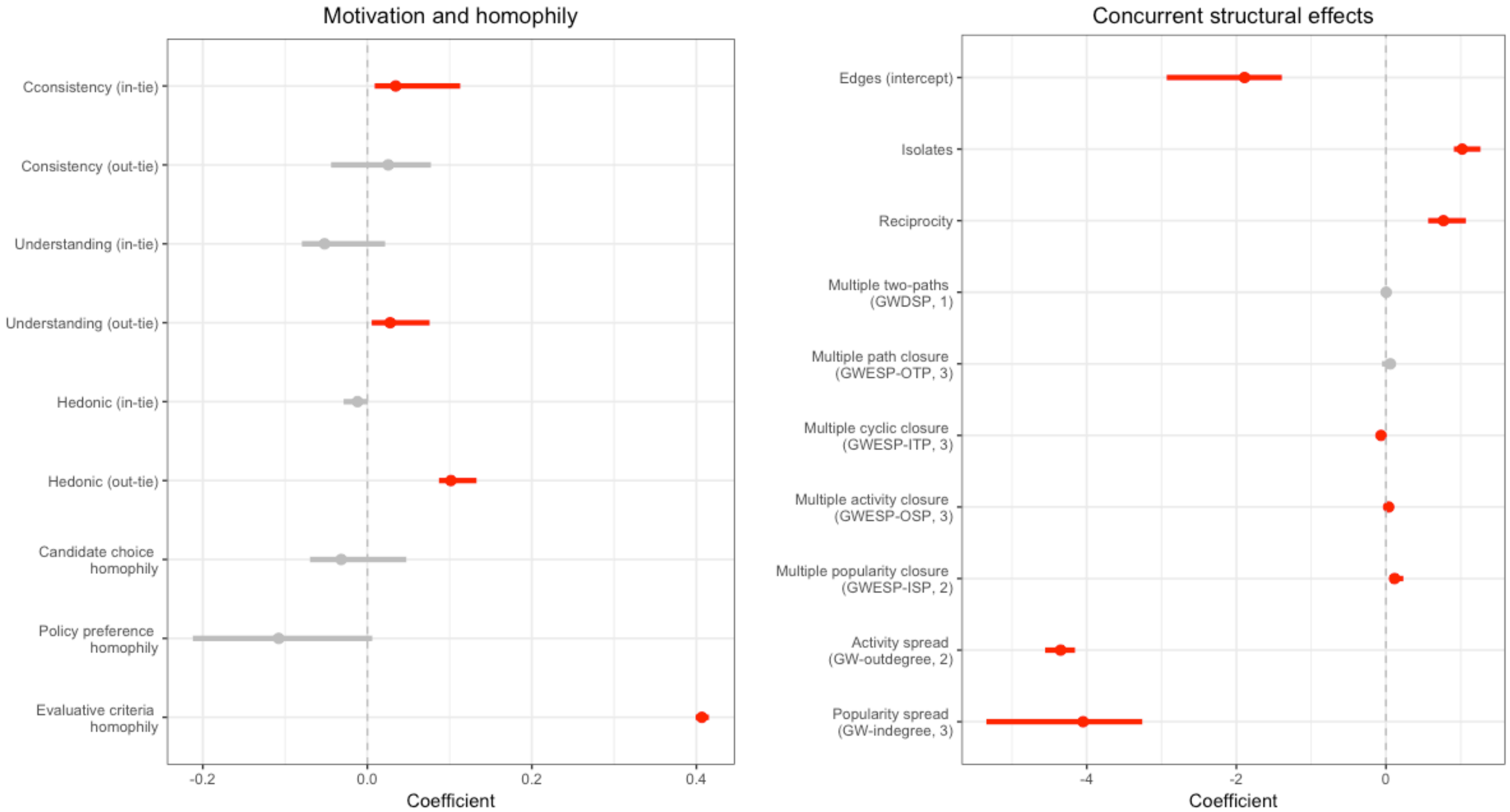
Table S3. Model robustness checks

	Final model I	Multiple Imputation	Daily	No Threshold	MRQAP
Edges (Intercept)	**-1.89** [-2.93; -1.39]*	**-1.19** [-3.29; -.58]*	**-1.24** [-1.80; -.64]*	**-.29** [-.37; -.25]*	1.209
Motivations and homophily					
Consistency (in-ties)	**.034** [.009; .113]*	**.026** [.023; .036]*	**.017** [.002; .036]*	**.026** [.019; .059]*	.052
Consistency (out-ties)	.025 [-.044; .077]	.028 [-.032; .090]	.001 [-.031; .030]	-.027 [-.050; -.006]*	-.030
Understanding (in-ties)	-.052 [-.080; .022]	**-.058** [-.087; -.023]*	**-.027** [-.046; -.009]*	**-.061** [-.085; -.017]*	**-.197***
Understanding (out-ties)	**.028** [.005; .076]*	.026 [-.002; .055]	.022 [-.016; .063]	**.036** [.034; .041]*	**.127****
Hedonic (in-ties)	-.012 [-.029; .001]	-.003 [-.015; .007]	-.006 [-.016; .004]	.007 [-.010; .028]	.079
Hedonic (out-ties)	**.102** [.087; .133]*	**.076** [.040; .112]*	**-.025** [-.048; -.006]*	-.002 [-.007; .030]	**-.300*****
Same candidate preference	-.032 [-.070; .047]	-.039 [-.081; .039]	**.040** [.020; .057]*	**.072** [.059; .094]*	.013
Similar policy preference	-.108 [-.212; .006]	.028 [-.105; .239]	.071 [-.012; .143]	.057 [-.033; .103]	-.092
Similar evaluative criteria	**.407** [.399; .415]*	**.461** [.445; .484]*	**.094** [.017; .176]*	**.053** [.012; .058]*	**.587***
Endogenous structural effects					
Isolates	**1.019** [.908; 1.264]*	**1.243** [.931; 1.402]*	**1.311** [1.051; 1.564]*	**1.470** [.967; 2.285]*	
Reciprocity	**.769** [.564; 1.068]*	**1.027** [.550; 1.298]*	**.848** [.759; .974]*	**.903** [.754; 1.008]*	**.416*****
Multiple two-paths (GWDSP)	.003 [-.007; .007]	.002 [-.006; .005]	.002 [-.001; .005]	-.001 [-.003; .001]	
Path closure (GWESP-OTP)	.058 [-.053; .125]	**.048** [.021; .151]*	**.083** [.066; .101]*	**.021** [.021; .024]*	
Cyclic closure (GWESP-ITP)	**-.066** [-.080; -.060]*	**-.053** [-.063; -.047]*	**-.060** [-.067; -.053]*	-.008 [-.015; -.003]*	
Activity closure (GWESP-OSP)	**.036** [.033; .045]*	.013 [-.013; .044]	**.017** [.008; .026]*	**.011** [.007; .014]*	
Popularity closure (GWESP-ISP)	**.115** [.093; .232]*	**.057** [.011; .073]*	**.081** [.059; .107]*	.010 [-.000; .021]	
Activity spread (GW-outdegree)	**-4.35** [-4.56; -4.16]*	**-4.78** [-5.27; -3.43]*	**-2.88** [-3.23; -2.60]*	**-4.10** [-4.43; -3.79]*	
Popularity spread (GW-indegree)	**-4.05** [-5.34; -3.26]*	**-4.59** [-5.05; -2.99]*	**-3.79** [-4.13; -3.46]*	**-4.63** [-4.84; -4.47]*	
Lagged structural effects					
Previous communication	**.222** [.192; .253]*	**.218** [.115; .262]*	**.221** [.171; .273]*	**.239** [.207; .257]*	**2.336*****
Delayed reciprocity	.074 [-.073; .194]	**.080** [.036; .248]*	-.025 [-.080; .025]	.001 [-.044; .064]	**-.227*****
Delayed transitivity closure	**.034** [.020; .055]*	**.010** [.004; .040]*	**.031** [.020; .041]*	**-.006** [-.008; -.001]*	
Delayed cyclic closure	**.034** [.008; .057]*	**.020** [.013; .037]*	**.002** [-.004; .009]	-.002 [-.005; .003]	
Delayed activity closure	**-.056** [-.067; -.046]*	**-.030** [-.058; -.009]*	**-.028** [-.035; -.018]*	**-.002** [-.004; -.001]*	
Delayed popularity closure	**-.059** [-.110; -.043]*	**-.033** [-.091; -.017]*	**-.009** [-.016; -.001]*	**-.013** [-.015; -.011]*	
Persistent sender (out-tie)	**.019** [.010; .029]*	**.019** [.009; .025]*	**.017** [.014; .019]*	**.009** [.005; .011]*	

Persistent receiver (in-ties)	**.023** [.018; .038]*	**.019** [.009; .027]*	**.002** [-.000; .003]	**.010** [.007; .012]*	
Controls					
Age (in-ties)	.001 [-.020; .022]	.002 [-.003; .026]	**-.022** [-.036; -.009]*	**-.015** [-.041; -.004]*	.049
Age (out-ties)	.052 [-.105; .093]	.069 [-.193; .120]	.029 [-.002; .064]	**.038** [.008; .052]*	**.307*****
Female (in-ties)	.005 [-.036; .041]	.022 [-.005; .070]	**-.037** [-.062; -.001]*	.009 [-.017; .042]	-.103
Female (out-ties)	.014 [-.348; .254]	.055 [-.269; .282]	**-.043** [-.105; -.003]*	-.005 [-.084; .042]	.037
Gender homophily	**.044** [.023; .086]*	**.069** [.043; .101]*	.018 [-.003; .041]	.016 [-.011; .034]	.055
Education (in-ties)	-.011 [-.039; .019]	-.007 [-.032; .017]	**-.019** [-.038; -.000]*	-.005 [-.018; .008]	-.092
Education (out-ties)	.016 [-.015; .091]	.006 [-.049; .130]	-.023 [-.060; .004]	-.027 [-.043; .018]	**-.252*****
Regional origin = Seoul (in-ties)	-.084 [-.130; .044]	-.049 [-.114; .017]	**-.077** [-.107; -.058]*	**-.071** [-.131; -.013]*	**-.377***
Regional origin = Seoul (out-ties)	-.125 [-.438; .350]	-.109 [-.463; .097]	**.098** [.032; .156]*	**.046** [.011; .088]*	**.433*****
Regional homophily (Seoul)	.017 [-.014; .080]	**.035** [.024; .049]*	.015 [-.012; .042]	.017 [-.010; .074]	-.032
Talk freq (in-ties)	**.046** [.024; .049]*	**.038** [.019; .045]*	**.026** [.015; .037]*	**.050** [.015; .070]*	**.268****
Talk freq (out-ties)	.014 [-.099; .161]	.003 [-.169; .076]	-.013 [-.037; .012]	.012 [-.043; .024]	-.016
Media use (in-ties)	**-.011** [-.019; -.003]*	-.015 [-.067; .003]	-.008 [-.014; .001]	-.010 [-.037; .002]	-.076
Media use (out-ties)	.033 [-.017; .071]	.007 [-.036; .075]	-.004 [-.013; .006]	.001 [-.012; .020]	**-.130****
Internal efficacy (in-ties)	-.013 [-.058; .055]	-.014 [-.033; .008]	**-.013** [-.021; -.005]*	-.010 [-.017; -.003]*	-.134
Internal efficacy (out-ties)	.024 [-.102; .128]	.015 [-.143; .113]	**.049** [.027; .078]*	**.065** [.028; .092]*	**.226*****
Candidate pref = Moon (in-ties)	.003 [-.008; .092]	-.005 [-.038; .044]	**-.034** [-.053; -.004]*	-.025 [-.105; .007]	-.019
Candidate pref = Moon (out-ties)	.013 [-.123; .066]	.024 [-.025; .074]	-.015 [-.055; .035]	-.023 [-.059; .021]	**.278*****
Num. obs.	291096	347820	2522832	291096	
F-statistic					1294.354
df1					29
df2					76787
Multiple R-squared					0.328
Adjusted R-squared					0.328

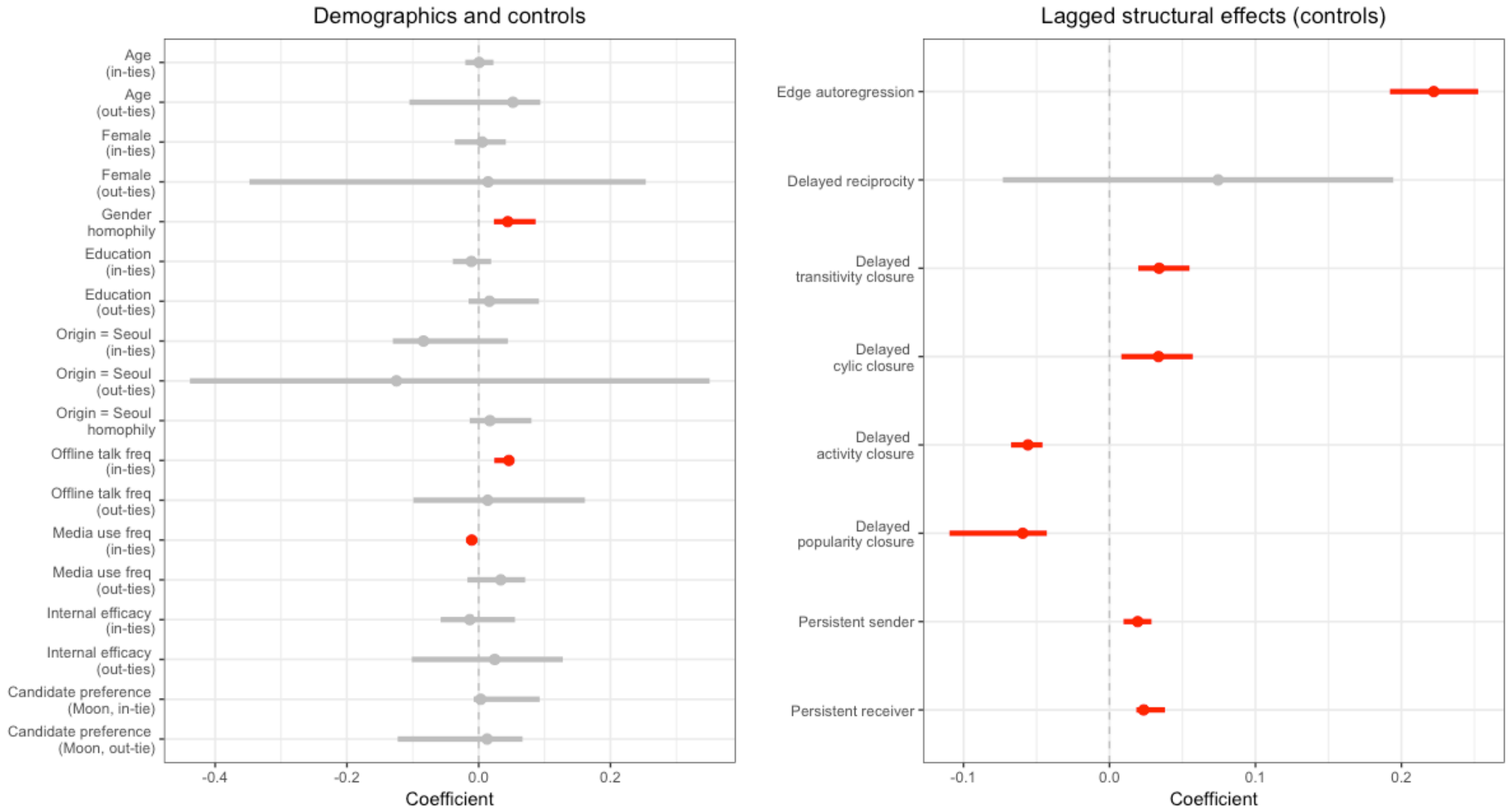
Note: ***p < 0.001, **p < 0.01, *p < 0.05 (or 0 outside the 95% bias-corrected and accelerated confidence interval based on 1000 replications). Significant results in bold. Final Model I = the leftmost final result reported in Table 3 of the manuscript. Multiple Imputation = multiple imputation for missing candidate preference at wave 1. Daily = Daily slice model with t = 26. No Threshold = no threshold model, such that ties defined as 0 vs. all other values. MRQAP = Multiple Regression using double semi-partialing Quadratic Assignment Procedure on the single aggregated network. For "no threshold" and "MRQAP" model, we did not dichotomize the original valued matrix. All other models use dichotomized matrix based on mean number of selection instances within each time slice. MRQAP model does not report traditional CIs or standard errors.

Figure S2. Parameter estimates and 95% confidence intervals from the final model



Note: Coefficients for key predictor variables (as reported in Table 2 in the main document) and their 95% CIs are reported (significant model terms are denoted in red).

Figure S2. Parameter estimates and 95% confidence intervals from the final model (con'd)



Note: Coefficients for control variables and their 95% CIs are reported (significant model terms are denoted in red).

Table S4. Additional robustness checks

Predicting "change statistics" of key network-endogenous variables as a function of partisan homophily variables (i.e., *same candidate preference*, *similar policy preference*, and *dissimilar ideological self-placement*) using *netlm* (QAP-regression) and *netlogit* (QAP-logit regression)

DV:	**Popularity spread** (gw-indegree)		**Activity spread** (gw-outdegree)		**Reciprocity** (mutual)	
IV: Same candidate preference						
Time point	***b***	***Pr* ≥ (|*b*|)**	***b***	***Pr* ≥ (|*b*|)**	***b***	***Pr* ≥ (|*b*|)**
T = 1	-.0056	.257	-.0122	.089	.1234	.225
T = 2	-.0024	.750	.0023	.828	.0716	.662
T = 3	-.0053	.286	-.0068	.401	.0934	.530
IV: Policy preference similarity						
Time point	***b***	***Pr* ≥ (|*b*|)**	***b***	***Pr* ≥ (|*b*|)**	***b***	***Pr* ≥ (|*b*|)**
T = 1	.0059	.847	.0804	.010	-.5281	.405
T = 2	.0564	**.036**	.1355	**.001**	-2.0456	**.005**
T = 3	.0275	.312	.0394	.347	.3777	.593
IV: Dissimilar ideological self-placement						
Time point	***b***	***Pr* ≥ (|*b*|)**	***b***	***Pr* ≥ (|*b*|)**	***b***	***Pr* ≥ (|*b*|)**
T = 1	-.0004	.933	-.0011	.868	-.0308	.724
T = 2	-.0049	.241	-.0091	.160	.0384	.679
T = 3	.0000	.999	-.0069	.251	.0785	.475

Note: *b* = unstandardized regression coefficients, where models include only intercept and a respective predictor variable. We used the double semi-partialing permutation with 1,000 replications for deriving probabilities of observed value (*b*) exceeding the either lower or upper tails of the simulated distribution at .05 level (denoted as *Pr* ≥ (|*b*|) above). Since change statistics for reciprocity is constrained to be zero and one, we used *netlogit* (logit regression QAP) instead of *netlm* (QAP regression). For all other dependent change statistics, we used *netlm* (QAP regression).