Selection Homophily in Political Discussion Networks:
Evidence from Formal Dynamic Models of the Selectivity Function

Matthew D. Sweitzer

Presented at the 2019 Annual Convention of the National Communication Association
Baltimore, Maryland

Abstract

Partisan homophily in political discussion networks is a profoundly impactful phenomenon. Much of the existing research has focused on the effects of homophilous network structures on the consequences of such structures, such as political participation and information flow. However, the antecedent conditions to homophily are not as well understood. In particular, the mechanisms of selection, structural features, and social influence are difficult to disentangle using traditional social scientific methods. This article presents a formal theory of the role social selection, or the active choice on the part of an individual to interact with more or less similar others, plays in the formation of a discussion network over time. I then test how varying degrees of selectivity result in more or less homophilous structures using agent-based modeling. The results indicate that selection of similar discussants leads to homophily in the network overall, and that a high degree of selectivity is necessary to produce levels of homophily that are found in observational studies of such networks. This suggests that a modest reduction in the selectivity function may have substantial effects on the degree of homophily in real-world discussion networks. The method employed here demonstrates how computational tools can be used to further communication theory where conventional research practices have difficulty isolating mechanisms.

Keywords: discussion networks, homophily, selectivity, issue publics, agent-based modeling

Selection Homophily in Political Discussion Networks:

Evidence from Formal Dynamic Models of the Selectivity Function

Homophily, or the tendency for people to share network connections with similar others (McPherson, Smith-Lovin, & Cook, 2001), is a powerful and well-studied phenomenon in political discussion contexts. In observational studies, political homophily has been found in neighborhoods (Gimpel & Hui, 2015; Lang & Person-Merkowitz, 2015), social media contexts (Colleoni, Rozza, & Arvidsson, 2014; Himelboim, Smith, & Shneiderman, 2013; Bond & Sweitzer, 2018), and romantic relationships (Huber & Malhotra, 2017). Other research has focused on the effects of homophily on normative political outcomes, including participation (Mutz, 2002; Nir, 2011), knowledge (Eveland & Hively, 2009), and information flow (Bakshy, Messing, & Adamic, 2015). However, comparatively little attention has been paid to the antecedent communicative processes which result in homophily at the network-level.

Three systematic mechanisms are theorized to produce homophily above chance (Cowan & Baldassarri, 2018): Interpersonal influence, structural features, and selection (sometimes referred to as “choice”). Distinguishing between these mechanisms can be challenging, particularly when networks are dynamic and researchers often only have access to one or a few time points after the network has already been formed (Aral, Muchnik, & Sundararajan, 2009). Interpersonal influence occurs when two people who share a network connection become more similar due to one or both people adopting the attitudes or behaviors of the other individual. For example, Nickerson (2008) and Bond et al. (2012) studied the propagation of voting behavior in pre-existing networks. By focusing on networks which have already been subject to structural and selection processes, coupled with random assignment to experimental conditions, the authors are able to control for the other mechanisms.

Structural features come into play when the context of an individual’s associations which are not ostensibly linked to political discussion is spuriously associated with ideology. For example, supposed that Sam and Nasrin are both well-educated and wealthy. They may, by virtue of these features, live in similar neighborhoods, work at the same organization, or travel in social circles with a great deal of overlap. They may, therefore, be more likely to engage in any discussion than if they did not share these characteristics. To the extent that wealth and education correlate with values and ideologies, their connection in the political discussion network, even if political discussion arises coincidentally, would therefore increase the observed level of homophily. In this way, politically homophilous interactions may be a byproduct of selection decisions that are made in domains which are not expressly political, but which are correlated with political ideology. The breadth of potential characteristics which may drive structural foundations of political homophily and their variable effects makes this mechanism very difficult to quantify in observational studies.

The final mechanism, selection homophily, concerns an individual's decision – conscious or otherwise – to engage in political discussion with another person on the basis of shared attitudes. There may be many factors involved in this decision, including structural features like an existing friendship. For this reason, some have theorized that selection of social network alters does not often rely on information about the political opinions of that alter (Huckfeldt, Johnson, & Sprague, 2004; Sinclair, 2012). However, this assertion stands in stark contrast to the findings in the minimal group paradigm (Billig & Tajfel, 1973; Diehl, 1990) which suggest that people's group identity and selection into connections with similar others can be instigated in previously unacquainted groups and with very trivial identity cues. Further, if one distinguishes between a connection in a social network writ-large and a connection in a political discussion network more specifically – wherein a tie represents active engagement in discussion of political matters – then it stands to reason that decisions concerning a connection in the latter would include expressly political considerations. To the extent that one's decisions about their alters in their political discussion network relies on evaluations of the ideological similarity between themselves and the alter, the accumulation of these individual decisions may result in homophily in the network as a whole.

The study presented here offers a formalization of the process of selection homophily. Specifically, I advance the concept of selectivity, or the degree to which one's decisions about the status of a tie in the political discussion network favors alters who the ego perceives to be similar to themselves (Lazarsfeld & Merton, 1954). While selection is a dyadic process, the agglomeration of these tie decisions among all dyads in a network produces homophily to the extent that similarity preferences are normative. The perception of similarity and the consequent tie decision are informed by two theories of political interactions: selective exposure and issue publics.

The contributions of this study are two-fold: To establish a formal model of the selectivity function, and – perhaps more importantly – to investigate how varying levels of selectivity affect the resulting levels of homophily in the social network. In the following pages, I discuss selective exposure and issue publics, paying specific attention to their application in interpersonal interactions. I then develop mathematical expressions of discursive interactions and the selectivity function. Finally, I test three such functions using agent-based modeling (ABM) of dynamic discussion networks to examine the role of an individual's selectivity in creating a network that is homophilous. Such tests offer new insight into the role of selection homophily in network formation.

Selective Exposure

Though typically utilized in research to address an individual's selection of news media sources, selective exposure (Sears & Freedman, 1967) was formulated to apply to all types of information sources. Selective exposure to interpersonal communication sources applies in the context of political discussion networks, because an individual's tie decisions concern whether they talk about political matters with another person (for a review of early interpersonal applications, see Zillman & Bryant, 1985). Recent applications of selective exposure theory in interpersonal contexts, for example, examine decisions to "follow" a like-partisan on social media (e.g., Himelboim et al., 2013). The theory argues that people are discerning when it comes to sources of political information. Specifically, they prefer to engage with those sources of information which they expect will comport with their pre-existing attitudes. Just as in mediated settings, where Democrats tend to watch MSNBC and Republicans tend to watch Fox News, in interpersonal contexts, partisans discuss political matters with other, ideologically-similar people. This constrains the information they may be exposed to in much the same way as mediated selective exposure (Feezell, 2016; Halberstam & Knight, 2016). The perception that an information source will reaffirm one's ideology stems from, among other things, previous interactions with the source.

Implied in this theory, and confirmed in later studies (e.g., Garrett, 2009), are two complementary, but distinct processes: a tendency to engage with attitude-reaffirming sources, and a tendency to avoid attitude-discrepant sources. Applying these processes in a political discussion network context, one can see how these processes might result in homophily. When a person avoids discussing political matters with someone who they think will disagree with them, the tie is either severed or ignored in the network.[1] Likewise, when a person chooses to engage with people who are similar to themselves, the tie is either maintained or formed in the network, possibly reinforcing homophilous relationships. If these processes are normative across individuals in the network, then the overall network structure will evince homophily.

Issue Publics

The theory of issue publics (Krosnick, 1990) concerns people's tendency to act on those political issues which are most salient or important to themselves. The theory stems from the psychological concept of "cognitive misers." This perspective argues that people have a limit to their cognitive capacities, and as such, will constrain the amount of effort they invest in understanding and acting on related, but comparatively undervalued concepts. For example, an American conservative may be more invested in either social or fiscal issues, and disproportionately so in the subsumed issues, such as gun ownership, abortion policy, or law enforcement. The conservative who considers gun ownership to be the most important issue may then be more likely to sign a petition concerning gun rights or to join the National Rifle Association than they are to canvas in support of abstinence education.

[1] Again, it is helpful to distinguish between social and political discussion networks. If a tie is severed in the latter, it does not necessarily mean that the same tie is severed in the broader social context. Rather, the alters no longer engage in discussion about political topics. It is not uncommon for people to avoid political discussion while retaining some other form of social relationship (Settle & Carlson, 2019).

More recent research integrating issue publics and selective exposure theories finds that partisans are more selective when the information pertains to an issue that is of high importance to them (Kim, 2009). In other words, perceived discrepencies between one's opinion and the viewpoint of the information source are *magnified* when the issue is important to them, but *lessened* when the issue is of lower importance. This augmented perception influences selection of an information source. This comes into play in political discussion networks when individuals are evaluating the level of similarity they have with another person and how it affects their decision to form or dissolve a tie. That is, people will be more likely to act on information that concerns the topics that are most important to them, while the less salient topics are not considered as deeply. In conjunction with selective exposure, the theory of issue publics provides a way in which differences of opinion may be magnified or diminished depending upon the concerns of the individual making the tie choice.

Mathematical Expression of Selectivity

Using these theoretical frameworks, I advance the concept of selectivity, or degree to which one's tie decision favor connections with high-similarity alters. To express this mathematically, it helps to envision a simple form of political discussion: The sharing of answers to a public opinion questionnaire. In simple statements of the extent to which one agrees or disagrees with a position, the alter evaluating this information can heuristically interpret the difference in the opinions expressed as the numerical distance between the points on the scale. Using selective exposure, one can surmise that the probability that a person would engage in political discussion with another person is inversely related to the difference in these expressed opinions. The difference can be calculated by taking the absolute value of the distance between the attitude of ego, i, and alter, j, on topic, t:

$$|A_{it} - A_{jt}| \tag{1}$$

Next, the theory of issue publics argues that people's political behaviors are tied to the subjective salience of the issue which their behavior concerns. In the example conversation, the behavior is a person's decision regarding their tie with an alter, and the issue is the topic currently being discussed. Combining these theories, I argue that issue salience acts as a "weight" on the perception of ego's political difference with alter such that the differences are perceived as greater when the issue is highly important to ego. In this way, the perceived differences of opinion may differ between ego and alter, even though they discussed the same issues, as a different set of issues may be more important to either the ego or the alter. In this way, the perception of difference with alter can be described as the interaction of the mathematical distance in the expressed attitudes with the salience, S, of the topic, t, to ego, i:

$$|A_{it} - A_{jt}| * S_{it} \tag{2}$$

Finally, discussions of political matters do not often concern just one topic, and so the perception of political difference with alter must update to include information about all of the topics previously discussed. Though there may be individual variations in the way people remember and amalgamate the sum total of all of the information one has about their discussion partner, here I use an average of all of the t topics discussed to simplify matters. Thus, I arrive at the final measure of one's *weighted perceived difference with alter* score:

$$\frac{\sum_{t=1}^{k}(|A_{it} - A_{jt}| * S_{it})}{k} \tag{3}$$

where k represents the number of topics discussed over time by the dyad. Using this measure, one could add algebraic functions to derive a probability distribution which corresponds to the likelihood that a specific decision regarding the status of a tie is selected. In other words, I can adjust the extent to which agents in the ABM are selective by assigning different probabilities of decision-making to certain values of the perceived difference measure. This probabilistic distribution is the *selectivity function* which I will manipulate in the ABMs that follow.

To the extent that low values of the weighted perceived difference with alter measure corresponds to a higher probability of either maintaining an existing relationship or adding a new connection in the political discussion network – and to the extent to which this probability decreases with increases in the perceived difference measure – an individual performing these tie decisions is engaging in the processes theorized concerning selection homophily. If this particular behavior is normative across individuals in the network, then the product of these decisions will affect the network as a whole. The dynamics of these multi-level relationships are presented in the theoretical model in Figure 1. From this logic, the core hypothesis of this study can be drawn:

H1: Higher degrees of selectivity for individuals results in higher levels of network-level homophily.

While the basics of the selectivity function and its relation to network-level homophily are straight-forward, the specific intricacies of this relationship can be quite complex. The distribution of the independent variable updates with each interaction over time. Individual decisions about breaking a tie may not only affect the composition of their ego network, but may result in meaningful changes in the topology of the overall network. It is thus unknown how sensitive the homophilous structure of the network may be to changes in an individual's behavior. I therefore present the following as a research question:

RQ1: What amount of selectivity is required to generate networks which are identifiably homophilous?

To examine these processes in greater detail, I rely on three ABMs of dynamic discussion networks. These models impose global strategies of the selectivity function which agents use to update their local network (i.e., ego network) structure. The structure of the overall network then serves as the dependent variable in these analyses.

Method

Agent-based Modeling

Agent-based modeling is a computational tool that can be used to simulate complex processes. In the social sciences, it can be applied to phenomena which are otherwise difficult to observe (de Marchi & Page, 2009) and for which a closed-form analysis is difficult (e.g., for problems that are NP-hard). In one of the earliest examples of this technique, Schelling (1969) developed a model of racial neighborhood segregation, moving "black" and "white" agents until they had a sufficient number of similar neighbors to satisfy a "rule" Schelling himself established. The rule, or strategy employed by agents in this case, represent happiness or a sense of security found in having neighbors who are of a similar race. While Schelling knew that the outcome of segregated groups was represented in the real world, the processes that led to segregation –

particularly racial preferences – were difficult to ascertain with human subjects due to social desirability biases. Simulation, in this sense, enables researchers to investigate questions for which the processes are difficult to study using more conventional social scientific methods.

Using an agent-based model in the same vein as Schelling, I model individual agent behaviors by making assumptions about how agents may make decisions in the model, and observe the global-level conditions of the social network that result from these dyad-level decisions. In doing so, the selection strategies I assign to agents serve as the independent variable and network properties after the model has run serve as the dependent variable. If agents can transition from a network with random connections to one where the connections definitively show evidence of homophily, then the strategy that produces such a result can be said to be influential in the formation of the network's structure.

Agent Characteristics

The models presented here are programmed in the Breadboard software for social network experimentation (McKnight & Christakis, 2018) and simulate political discussion over several "rounds", or time points, each corresponding to a different political topic. Using a dynamic model which permits network restructuring over time allows for the examination of emergent processes like homophilous selection (e.g., Macy, Kitts, Flache, & Benard, 2003). Each agent begins with a political party identity score that is assigned to them by drawing a random integer from the standard 1-7 partisan identification (strong Democrat to strong Republican) scale. Next, each agent is assigned ten attitudes, each corresponding to a different topic. Because political attitudes are strongly correlated with one's party identification (Jacoby, 1988), partisan identification scores are used to inform attitude scores. Specifically, I add a number drawn randomly from a Gaussian distribution ($M = 0$, $SD = 1$) to the partisan identification score and round to the nearest integer on a 1-7 scale for each of ten attitude scores. Using this technique, agents have attitudes which are strongly correlated with their partisan identification, but not perfectly so ($r = .79$). Moreover, because the scales are rounded more often beyond the floor and ceiling for those agents whose partisan identification score is situated at the ends of the scale, stronger partisans are more likely to have higher consistency in their attitudes (see Peterson, Slothuus, & Togeby, 2010) compared to independents and party leaners. Finally, for each of the ten political topics, agents are assigned issue importance scores by drawing random integers between 1 and 7.

In each model, $N_v = 50$ agents are placed into an Erdős-Rényi (1959) random graph in which approximately 20% of the possible edges are realized (the specific percentage of ties realized varies across simulations, but in all cases it is very close to 20%). Using a random graph model as the starting condition ensures that homophily is not a feature of the network at the outset. At the beginning of each round, agents share their attitude scores for the corresponding topic (in round 1, the first topic is discussed; in round 2, the second, etc.) with each of their connected alters. Next, 40% of all of the possible ties – both present and absent in the given round – are selected at random for possible updating. By restricting the proportion of ties which can update in a given round, one can “model in” the viscous nature of social networks (Rand, Arbesman, & Christakis, 2015). One of the two agents in the dyad is selected at random to make a decision regarding the status of the tie. If the tie is not present in the given round – that is, if the agents were not in each other’s discussion ego network in the previous round – then the agents’ attitudes regarding the present topic are shared with one another.

The decision an agent faces varies depending upon the status of the tie in the given round. If the tie is present, then the choice is between maintaining the tie, or dissolving it. If the tie is not present, then the choice is between adding the new tie, or ignoring it. In both cases, the former results in the tie being present in the subsequent round, whereas the latter results in the tie being absent in the subsequent round. Hereafter, I refer to the tie-present choice as the associative tie choice, and to the tie-absent choice as the dissociative tie choice. In order to probabilistically link these tie decisions to perceived attitudinal differences between ego and alter, I use three different distributions of the selectivity function, described in the next section.

Selectivity Functions

Selectivity, or the degree to which people actively choose to form relationships with other, like-minded people, is the key process behind selection homophily (Lazarsfeld & Merton, 1954).

The selectivity function acts as a sort of “gradient heuristic” (Page, 2007, pp. 62–64), or a strategy used by the agents in the model to assess the attitudinal distance between themselves and their networked alters and to make decisions regarding the status of a tie based on that assessed distance. Depending on the degree of selectivity, agents will form network ties with more or less similar alters. If one were to think about this selectivity function as a distribution of the probability of performing associative tie choices across values of the perceived difference with alter measure, then selection homophily should occur when the distribution is right-skewed. That is, when people perform associative tie choices more often on alters with whom they perceive few differences of opinion, and dissociative tie choices on alters with whom they perceive larger (augmented by issue salience) differences of opinion.

To test the role of selectivity in driving network-level homophily, I test three distributions of the selectivity function. Importantly, each of these distributions was selected such that the same number of associative and dissociative tie choices were selected across all three ABMs. This means that the global density, or the proportion of all potential ties in the network that are actualized, should be roughly equivalent across all three models throughout all ten discussion rounds. This is important, as homophily is characterized by increased local density, or the proportion of potential ties among each node’s alters that are actualized (McPherson et al., 2001). Globally-dense networks also tend to be locally-dense, which may obscure substantive differences in the network structure between models. Thus, changes in local density – keeping global density constant – will be used to compare the resulting structure in each model. What differs between models is the extent to which the associative tie choices favor connections with similar alters, and so if one’s alters are also selective, I expect local density to increase.

The first distribution (hereafter, “Model 1”) represents a null model in which tie choices are decoupled from weighted perceived differences with alter. This can be formalized with the function:

$$P(y_{ij}) = 0.251 \tag{4}$$

where $P(y_{ij})$ refers to the probability that agent i selects the associative tie choice concerning their relationship with agent j. In other words, each time an agent faces a decision, 25.10% of the time they will select the associative tie choice, regardless of the attitudes expressed by the alter or ego's subjective issue importance of the issues discussed.

The next distribution I tested ("Model 2") is one of moderate selectivity. This distribution is given by the function:

$$P(y_{ij}) = \frac{-0.348}{42} * \frac{\sum_{t=1}^{k}(|A_{it} - A_{jt}| * S_{it})}{k} + 0.348 \tag{5}$$

Here, I incorporate the weighted perceived difference with alter measure developed above. Because attitude and issue importance scores both range on a 1-7 scale, this measure can range from 0, representing no difference in the expressed opinions between ego and alter, to 42, representing maximally different opinions on maximally important issues to ego. When this measure equals 0, the ego makes the associative tie choice 34.80% of the time. This probability then decreases linearly until it reaches 0.00% when the weighted perceived difference score equals 42.

The final distribution ("Model 3") represents a high degree of selectivity and is expressed by the function:

$$P(y_{ij}) = \frac{110}{120 + (\frac{\sum_{t=1}^{k}(|A_{it} - A_{jt}| * S_{it})}{k})^3} \tag{6}$$

This distribution follows an inverse-sigmoidal curve. When the weighted perceived difference with alter is 0, the probability that ego selects the associative tie choice is 91.67%. However, that quickly drops to 50% when the weighted perceived difference is just 4.64 and approaches 0% for the remainder of the scale. Such a strong right-skew demonstrates both high preference for low-difference alters, and avoidance of moderate-to-high difference alters. If selectivity results in network-level homophily (H1), this relationship should be strongest in Model 3.

Analytic Strategy

First, analysis of both the networks and key descriptive statistics of the ABM across each of the eleven time points should help elucidate important differences between models. In the networks, homophily would be evidenced by clear segregation of the nodes by the party identification scores. That is, “Democrat” and “Republican” nodes should be grouped together with other like-partisans, and “independents” are likely to occupy the space between groups, as opposed to direct connections between nodes of different party affiliations. Because each of the selectivity functions were designed to have similar overall levels of associative and dissociative tie choices, then the overall graph density – the proportion of all possible ties for which the tie is present – should be roughly equivalent across models. Two other statistics should differ between models: transitivity and assortativity. Transitivity is a measure of how locally-dense a network is, or the extent to which nodes tend to form tightly-knit groups. This measure is taken as the proportion of all 2-edge triads (i is connected to j and j is connected to k) for which the third edge is present as well (i is connected to k, or a closed triangle). Assortativity is a measure ranging from -1 (perfect dissimilarity) to 1 (perfect similarity) of the Pearson correlation of some numeric node characteristic – in this case, party identification – across the network. These two measures, taken together, indicate the extent to which agents group together with other like-partisans.

For hypothesis testing, I use a bootstrapped temporal exponential random graph model, or BTERGM (Leifeld, Cranmer, & Desmarais, 2018, 2019). Like other models in the exponential random graph family, BTERGMs are capable of elucidating the generative features of a network. That is, the coefficients represent the log-likelihood that a tie in the network was formed given some particular process and controlling for the other processes specified in the model. What sets the BTERGM apart, though, is its ability to handle temporal networks, conditional on the status of a tie in the previous time period. Moreover, the bootstrapping uses a maximum pseudolikelihood function, using $10,000$ bootstraps in the case of the models outlined here, to generate 95% confidence intervals of estimates and to better account for non-normality in the features specified in the model. For the purposes of this study, the model is specified by including the absolute difference of the party id[2] variable and edge memory to ascertain how partisan homophily contributes to tie formation in round t above and beyond the tie structure in round $t-1$. Because the absolute value of the difference in scores represents greater political distance, a significant

[2]Note that this is not the same as using the attitude scores as the dependent variable, because party identification scores are not shared with other agents.

negative value for this coefficient is expected when selection homophily is a generative feature of the network.

Further, the same model can be run without the homophily parameter in the specification. Then, comparisons of the model fit indices between the models with and without homophily included as an explanatory variable can serve as a robustness check – that homophily substantially contributes to the process of network formation and without it, the BTERGM fails to match key features of the observed network (for a review of parameter specification and model fit, see Cranmer & Desmarais, 2011). In other words, if homophily is key to the observed network's formation, then the simulated networks of the BTERGM without homophily specified in the model should have very different characteristics from the observed network.

Results

Model 1

Model 1 of the ABM imposed tie decision strategies on the agents that did not account for their perceptions of political similarity with their alters. In other words, this model serves as a test of the null condition. If homophily is a significant generative process in this network, absent a strategic selective process, then perhaps the selection mechanism may not function as presently theorized. As shown in the network graphs of this model (see Figure 2), the network structure does not appear to change much from round to round. Time_0 in the figure represents the network when agents are assigned discussion partners at random. After ten discussion rounds (Time_{10} in the figure), the network appears just as randomly structured as the starting condition.

The interpretation of the graphs is reinforced when examining the descriptive statistics of the network at each time period (see Table 1). Though the graph density increases modestly over time, the increase is in line with the other two models. Transitivity shows a modest increase as well (from 0.19 at Time_0 to 0.28 at Time_{10}). The increase in closed triads is reflective of the greater global density, or the increased probability that any tie is actualized in the network, and is likely not the result of any systematic clustering of the nodes. Observation of the assortativity statistic further emphasizes the lack of systematic clustering by partisans. At Time_0, the value is

-0.06, indicating a slight preference for heterophily in the network.[3] Throughout the discussion rounds, this statistic never becomes positive and seems to fluctuate randomly near zero. The selectivity strategy imposed in Model 1 does not produce high correlations in the party identification scores of connected agents.

Finally, the BTERGM estimation results (see Table 2) presented in Model 1 indicate that homophily, measured as the absolute difference in party identification scores of connected alters, did not significantly predict the presence of a tie. Though the confidence interval includes zero and the estimate is not significant, the theta is positive, which is suggestive of a preference for heterogenous ties. Taken together, these results suggest that an individual strategy of no selectivity does not produce network-level homophily. This makes sense, though, as this model tests the absence of selection homophily.

Model 2

The second ABM tests a moderate level of selectivity. The network graphs (see Figure 3) show that, once again, random assignment at Time_0 successfully imposes no specific homophilous structure on the network at the outset. Unlike in Model 1, though, the graphs of Model 2 do show the network structure transitioning toward segregation. This shift is clearest in the graph of Time_3, where the red “Republican” nodes occupy the right hemisphere and the “Democrats” occupy the left hemisphere. However, these graphs appear to show quite a bit of noise which makes the signal of homophily difficult to discern.

In the descriptive statistics of Model 2 (see Table 1), the signal is no clearer. Once again, the graph density is in line with the slight increases shown in Model 1, indicating successful construction of the selectivity function to induce the same proportion of associative to dissociative tie choices. Also like Model 1, the transitivity, or clustering of the network, tends to increase slightly over time. The assortativity measure of the correlation of party identification scores does increase and remain positive for discussion rounds 1 through 10. However, there does not appear to be any consistency to the way similarity among connected agents is maintained, as

[3]This is to be expected, given that the number of potentially different alters with whom an agent could be connected in the start of the model is slightly greater than the number of potentially similar alters.

the score reaches its highest level at Time$_3$ (0.17) and then fluctuates below that for the remainder of the model.[4] This suggests that the network reaches its "equilibrium" for the degree of partisan clustering relatively quickly, and that level is not very high – nowhere near levels found in observational studies.

However, the BTERGM estimation (see Table 2) suggests that the moderate selectivity strategy means that homophily is a generative process in the networks; $\theta = -0.08$, 95% CI $= [-0.11, -0.05]$. In other words, the odds that a tie is formed between two individuals with each unit increase in the absolute difference between party identification scores is 92.21% as likely as the preceding unit. Though significant, this effect seems to be fairly modest considering that the maximum difference in party identification scores is only 6. Whether this relationship between selectivity and network-level homophily is robust will be addressed later on. Overall, Model 2 does show some evidence of a systematic process that produces partisan homophily. However, the degree of clustering among like-partisans does not approach levels that are typically found in observational studies. This may suggest that the degree of individual selectivity is likely higher in real-world settings.

Model 3

The Model 3 ABM demonstrates how high selectivity operates in a dynamic network. Looking at these graphs (see Figure 4), it is very clear that this increase in agents' selectivity produces a very different network structure compared to Model 2. Already by Time$_3$ the graphs show that "Republicans" and "Democrats" occupy distinct spaces in the network. Through later discussion rounds, the network becomes more elongated in shape, indicating both greater distance between the nodes on the periphery of the network and greater cohesiveness among the partisans occupying those positions. By Time$_{10}$, not only do "Democrats" and "Republicans" look segregated, connected only by "independents" in the center of the graph, the nodes also appear to be stratified by their strength of partisanship as well. In other words, extreme partisans occupy the very ends of the network, connected only to each other and to other partisans, while the party

[4]For reference, an observational study by Bond and Sweitzer (2018) found that users of online forums designated for political discussion exhibited assortativity between 0.4 and 0.8

leaners and true independents occupy the cohesive parts of the network between partisan groups.

The descriptive statistics (see Table 1) further evince the systematic segregation and clustering of like-partisans. Just as before, the graph density remains similar to the previous two ABMs, suggesting that the number of associative and dissociative tie choices does not change between these models. Rather, it is specifically how these choices are allocated in Model 3 that makes a substantial difference in the resulting network structure as it pertains to partisanship. Both transitivity and assortativity increase steadily from round to round. Only in the last two rounds does the transitivity score fail to increase, though this may be indicative of the model having reached an equilibrium. Unlike in Model 2, the high selectivity ABM shows levels of assortativity that are on par with observational research of political discussion networks (e.g., Bond & Sweitzer, 2018).

The BTERGM results (see Table 2) also confirm these findings. The absolute difference in party identification scores significantly and negatively affects tie formation; $\theta = -0.35$, 95% CI $= [-0.44, -0.29]$. For each unit increase in party id differences, the odds of a tie forming are just 70.13% of the odds when agents are one unit closer in party id. Across party id differences from 0 to 6, these odds indicate relatively strong disparities in the tie choices among similar alters compared to dissimilar ones. These results, taken together, conclusively support hypothesis 1 regarding the role of selectivity in forming homophilous networks. Exactly how much selectivity is required before homophily is discernable in the network remains uncertain, though, as the BTERGMs found significance in the absolute difference scores for both Model 2 and Model 3. In the following section, I examine the robustness of these models.

Robustness Checks

In general, one method of evaluating the results of a BTERGM is to remove the key parameter from the model specification and compare the resulting model fit against the original model (Cranmer & Desmarais, 2011). This enables the researcher to examine the impact that a key explanatory variable has on the generative process for the network. I apply that to this study by excluding the homophily term – absolute difference of party identity – from the BTERGM's described above and examining the resulting models for their fit to the observed networks. If the simulated networks in the new BTERGM are characteristically different from the observed network, then homophily is critical to that model specification, and thus to the formation of the observed network. However, if the new model does fit well, then homophily was not a crucial component of the generative processes of the network. The fit indices include a variety of ego-level (degree), dyad-level (dyad- and edge-wise shared partners), and network-level (geodesic distance and modularity) statistics.

Figure 5 shows the differences in the model fit indices between the model with homophily (top panel) and without homophily (bottom panel) in the model specification of the moderate selectivity ABM (Model 2). Here, solid black lines – or red in the case of the modularity statistic – indicate the distribution of the statistics in the ABM network. Meanwhile, the box plots in the background – or the solid black line for the modularity statistic – show the frequency with which each statistic was found at each value across all of the $10,000$ simulations of the network. If the observed findings fall within – or at least close to – the boxes, then the model fit the observed network well for that value of the given statistic. As can be seen in this figure, both versions of the BTERGM specification fit the ABM well. In fact, there is little if any noticeable shift in the statistics, let alone an appreciable one. This indicates the homophily is not a critical generative process of the network when agents only employ a moderate level of selectivity.

Figure 6 details the model fit indices for the BTERGMs estimated with (top panel) and without (bottom panel) homophily for the high selectivity ABM. Unlike with Model 2, these graphs show drastic differences in the fit indices between models. For each of the node- and dyad-level statistics, the observed findings now fall at or outside the edges of the boxes for the simulated statistics. Furthermore, the network-level statistics show even starker differences – and potentially problems with model degeneracy. Though the observed levels of geodesic distances still falls within the bounds of the boxes from the simulated networks, the box plots show a much wider and much more skewed set of distributions of the simulated values relative to the prior specification. Moreover, the distribution of the simulated modularity scores is now wildly different from the observed distribution. This suggests that the high selectivity model shows a great deal more evidence of division into tight-knit groups, and without specifying homophily in the model, the BTERGM fails to replicate this structure.

While the ABM of moderate selectivity failed to consistently show evidence of homophily, the model of highly selective agents definitively demonstrates homophilous network structure. Moreover, the degree of homophily shown after ten discussions in this model closely matches the correlation of ideologies observed in real world settings (Bond & Sweitzer, 2018). While the answer to the research question of where the threshold of selectivity exists which will conclusively result in partisan homophily remains elusive, it could be surmised that the threshold selectivity function lies somewhere between the moderate and high selectivity distributions tested here. That said, given that the highly selective model demonstrates homophily near the levels seen in actual political discussion networks, it is likely the case that the real-world average distribution of the selectivity function is somewhat closer to this model. The implications of this conclusion are wide-ranging and could be cause for concern.

Discussion

This research examined whether and to what extent processes related to an individual's selection of political discussion partners can culminate in a network structure that is homophilous. I derived the selectivity function as the connection between one's perception of the degree of political difference they have with an alter and their decision concerning the status of their tie with that alter. Through the application of dynamic agent-based models wherein agents were assigned different selection strategies, I was able to determine that only a high degree of selectivity results in homophily that is easily identified by network analytic techniques. Moreover, only the high selectivity model resulted in a level of homophily that is on par with the findings of observational research (e.g., Bond & Sweitzer, 2018; Bakshy et al., 2015; Colleoni et al., 2014).

This finding is somewhat surprising. A similar study by Baldassarri and Bearman (2007) was conducted which used ABMs to identify the role of the influence mechanism in producing homophilous networks. In their model, agents were placed into a static network, though the attitudes were permitted to shift such that they could become congruent with the attitudes of their networked alters. Their findings suggest that even a modest influence effect results in substantial coalescence around particular points of view – that is, the networks become homophilous by virtue of the convergence of attitudes (see also Abelson, 1979). Here, though, I find that the unique contribution of the selection mechanism must be strong to produce similar segregation in a dynamic network.

On the one hand, this finding might be cause for concern, considering that the social identities tied to political parties are so strong (Greene, 2004) and that group differentiation has been observed with much more trivial identity cues in zero-history groups (Billig & Tajfel, 1973; Diehl, 1990). On the other hand, the findings related to Model 2 suggest that even a modest shift in an individual's selectivity could drastically reduce network-level homophily.

This may prove to be an invaluable insight for researchers hoping to reduce homophily in political discussion networks, as interventions could be designed which affect such a small shift in selectivity. Further research is needed to identify precisely which selectivity function serves as the threshold above which homophily is identifiable in the network. By identifying such a threshold, this could help to ascertain the objective of such interventions. Also, human subjects research is needed to both verify the veracity of the simulations presented here and to establish what the selectivity function looks like in real-world political discussion networks.

Although explication of the selection mechanism and simulations of its unique contribution to network-level homophily are substantial contributions to the literature on political discussion networks, this study is not without its drawbacks. By using random assignment in zero-history networks and constraining "discussion" to political topics, I essentially ruled out the effect of structural features on homophily. In a similar vein, I nullified the impact of interpersonal influence by keeping attitudes static. However, both of these processes play a part in the construction and maintenance of homophilous networks. While allowing each of these to vary in a model would make distinguishing their unique effects difficult (Aral et al., 2009), doing so would make for a more ecologically valid – albeit much more complex – model.

Similarly, I made a number of assumptions in programming the model which, although done so in the name of simplicity, also limit the generalizability of these findings beyond the simulations. For example, the simple averaging of the discussed topics does not take into consideration the potential for primacy/recency effects over the course of discussion, nor does it allow for variations in individuals' memory for the discussed opinions. Though accounting for each of these would make the model much more complex, such a model would provide an interesting connection to the literature on social information processing.

Likewise, allowing for individual differences in strategies or dynamic strategies could make for an interesting, though vastly more complex model. Recent research on the topic of selective exposure has differentiated between selection into engagement with attitude-consistent information sources and avoidance of attitude-discrepant sources (Garrett, 2009). Although the selectivity function allows both of these processes to occur simultaneously, I have been unable to ascertain which is the more predominant function, as doing so would require closer examination of the frequencies of the weighted perceived difference score for each of the $4,900$ decisions in each model. It may also be likely that the selectivity function varies with the history of the relationship such that people are more lenient when it comes to perceived differences with well-established discussion partners. Incorporating variations in both of these phenomena could prove challenging but worthwhile.

The models shown here indicate the importance of individual-level processes to the development of group-level outcomes. Selectivity has far-reaching effects beyond one's own ego network. Though homophily derived from the choices concerning the status of ties is possible, it takes a great deal of selectivity to instigate and sustain homophily. This leaves the door open for exogenous effects to mitigate selectivity, and thus homophily. Continued research on this subject is needed to evaluate the potential of the reduction of selectivity in achieving normative democratic outcomes.

References

Abelson, R. P. (1979). Social clusters and opinion clusters. In P. W. Holland & S. Leinhardt (Eds.), *Perspectives on social network research* (p. 236–256). New York, NY: Academic Press.

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*, 21544–21549. doi: 10.1073/pnas.0908800106

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

Baldassarri, D., & Bearman, P. (2007). Dynamics of political polarization. *American Sociological Review*, *72*, 784–811. doi: 10.1177/000312240707200507

Billig, M., & Tajfel, H. (1973). Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, *3*, 27–52. doi: 10.1002/ejsp.2420030103

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*, 295–298. doi: 10.1038/nature11421

Bond, R. M., & Sweitzer, M. D. (2018). Political homophily in a large-scale online communication network. *Communication Research*. (Advance online publication) doi: 10.1177/0093650218813655

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chambers or public sphere? predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, *64*, 317–332. doi: 10.1111/jcom.12084

Cowan, S. K., & Baldassarri, D. (2018). ”it could turn ugly”: Selective disclosure of attitudes in political discussion networks. *Social Networks*, *52*, 1–17. doi: 10.1016/j.socnet.2017.04.002

Cranmer, S. J., & Desmarais, B. A. (2011). Inferential network analysis with exponential random graph models. *Political Analysis*, *19*, 66–86. doi: 10.1093/pan/mpq037 de Marchi, S., & Page, S. E. (2009). Agent-based modeling. In J. M. Box-Steffensmeier, H. E. Brady, & D. Collier (Eds.), *The Oxford handbook of political methodology* (p. 71–95). New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780199286546.003.0004

Diehl, M. (1990). The minimal group paradigm: Theoretical explanations and empirical findings. *European Review of Social Psychology*, *1*, 263–292. doi: 10.1080/14792779108401864

Erdős, P., & Rényi, A. (1959). On random graphs, I. *Publicationes Mathematicae*, *6*, 290–297.

Eveland, W. P., Jr., & Hively, M. H. (2009). Political discussion frequency, network size, and ”heterogeneity” of discussion as predictors of political knowledge and participation. *Journal of Communication*, *59*, 205–224. doi: 10.1111/j.1460-2466.2009.01412.x

Feezell, J. T. (2016). Predicting online political participation: The importance of selection bias and selective exposure in the online setting. *Political Research Quarterly*, *69*, 495–509. doi: 10.1177/1065912916652503

Garrett, R. K. (2009). Echo chambers online? politically motivated selective exposure among internet news users. *Journal of Computer Mediated Communication*, *14*, 265–285. doi: 10.1111/j.1083-6101.2009.01440.x

Gimpel, J. G., & Hui, I. S. (2015). Seeking political compatible neighbors? the role of neighborhood partisan composition in residential sorting. *Political Geography*, *48*, 130–142. doi: 10.1016/j.polgeo.2014.11.003

Greene, S. (2004). Social identity theory and party identification. *Social Science Quarterly*, *85*, 136–153. doi: 10.1111/j.0038-4941.2004.08501010.x

Halberstam, Y., & Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, *143*, 73–88. doi: 10.1016/j.jpubeco.2016.08.011

Himelboim, I., Smith, M., & Shneiderman, B. (2013). Tweeting apart: Applying network analysis to detect selective exposure clusters in Twitter. *Communication Methods and Measures*, *7*, 195–223. doi: 10.1080/19312458.2013.813922

Huber, G. A., & Malhotra, N. (2017). Political homophily in social relationships: Evidence from online dating behavior. *The Journal of Politics*, *79*, 269–283. doi: 10.1086/687533

Huckfeldt, R., Johnson, P. E., & Sprague, J. (2004). *Political disagreement: The survival of diverse opinions within communication networks*. New York, NY: Cambridge University Press.

Jacoby, W. G. (1988). The impact of party identification on issue attitudes. *American Journal of Political Science*, *32*, 643–661. doi: 10.2307/2111240

Kim, Y. M. (2009). Issue publics in the new information environment: Selectivity, domain specificity, and extremity. *Communication Research*, *36*, 254–284. doi: 10.1177/0093650208330253

Krosnick, J. A. (1990). Government policy and citizen passion: A study of issue publics in contemporary America. *Political Behavior*, *12*, 59–92. doi: 10.1007/BF00992332

Lang, C., & Person-Merkowitz, S. (2015). Partisan sorting in the United States, 1972–2012: New evidence from a dynamic analysis. *Political Geography*, *48*, 119–129. doi: 10.1016/j.polgeo.2014.09.015

Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. In M. Berger, T. Abel, & C. H. Page (Eds.), *Freedom and control in modern society* (p. 18–66). New York, NY: Van Nostrand.

Leifeld, P., Cranmer, S. J., & Desmarais, B. A. (2018). Temporal exponential random graph models with btergm: Estimation and bootstrap confidence intervals. *Journal of Statistical Software*, *83*, 1–36. doi: 10.18637/jss.v083.i06

Leifeld, P., Cranmer, S. J., & Desmarais, B. A. (2019). btergm: Extensions of exponential random graph models [Computer software manual]. Retrieved from `https://cran.r-project.org/package=btergm`

Macy, M. W., Kitts, J. A., Flache, A., & Benard, S. (2003). Polarization in dynamic networks: A Hopfield model of emergent structure. In R. Brieger, K. Carley, & P. Pattison (Eds.), *Dynamic social network modeling and analysis: Workshop summary and papers* (p. 162–173). Washington, D. C.: National Academies Press.

McKnight, M. E., & Christakis, N. A. (2018). Breadboard: Software for online social experiments [Computer software manual]. Retrieved from `http://breadboard.yale.edu`

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

Mutz, D. C. (2002). The consequences of cross-cutting networks for political participation. *American Journal of Political Science*, *46*, 838–855. doi: 10.2307/3088437

Nickerson, D. W. (2008). Is voting contagious? evidence from two field experiments. *American Political Science Review*, *102*, 49–57. doi: 10.1017/S0003055408080039

Nir, L. (2011). Disagreement and opposition in social networks: Does disagreement discourage turnout? *Political Studies*, *59*, 674–692. doi: 10.1111/j.1467-9248.2010.00873.x

Page, S. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Peterson, M. B., Slothuus, R., & Togeby, L. (2010). Political parties and value consistency in public opinion formation. *Public Opinion Quarterly*, *74*, 530–550. doi: 10.1093/poq/nfq005

Rand, D. G., Arbesman, S., & Christakis, N. A. (2015). Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences*, *108*, 19193–19198. doi: 10.1073/pnas.1108243108

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, *59*, 488–493.

Sears, D. O., & Freedman, J. L. (1967). Selective exposure to information: A critical review. *Public Opinion Quarterly*, *31*, 194–213. doi: 10.1086/267513

Settle, J. E., & Carlson, T. N. (2019). Opting out of political discussions. *Political Communication*, *36*, 476–496. doi: 10.1080/10584609.2018.1561563

Sinclair, B. (2012). *The social citizen: Peer networks and political behavior*. Chicago, IL: Chicago University Press.

Zillman, D., & Bryant, J. (1985). *Selective exposure to communication*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Table 1

Descriptive Statistics of the Three ABM Networks Across Time

	Model 1 – No Selectivity			Model 2 – Moderate Selectivity			Model 3 – High Selectivity		
	Density	Transitivity	Assortativity	Density	Transitivity	Assortativity	Density	Transitivity	Assortativity
$Time_0$	0.20	0.19	-0.06	0.21	0.20	-0.08	0.19	0.19	-0.03
$Time_1$	0.22	0.21	-0.15	0.24	0.25	0.12	0.24	0.26	0.10
$Time_2$	0.24	0.25	-0.11	0.26	0.26	0.17	0.26	0.29	0.27
$Time_3$	0.25	0.25	-0.09	0.27	0.27	0.14	0.27	0.31	0.38
$Time_4$	0.26	0.26	-0.09	0.28	0.29	0.13	0.26	0.31	0.45
$Time_5$	0.27	0.27	-0.12	0.27	0.28	0.13	0.27	0.35	0.54
$Time_6$	0.26	0.27	-0.10	0.26	0.25	0.14	0.27	0.36	0.59
$Time_7$	0.28	0.28	-0.04	0.27	0.26	0.11	0.26	0.37	0.63
$Time_8$	0.29	0.33	-0.05	0.26	0.26	0.11	0.26	0.38	0.63
$Time_9$	0.28	0.29	-0.06	0.26	0.24	0.06	0.25	0.34	0.65
$Time_{10}$	0.28	0.28	-0.06	0.26	0.25	0.06	0.26	0.33	0.68

Note. Graph density represents the proportion of all possible ties in the network for which a tie is present. The selectivity functions were designed to keep this statistic roughly equal between models. Transitivity measures the amount of "clustering" in the network, or the proportion of all 2-edge triads (e.g., i is connected to j, j is connected to k) for which the third edge exists as well (i is connected to k). Assortativity measures the degree to which dyads are similar on the 7-point party id scales (higher scores mean greater homophily), which was not shared between agents during the discussion rounds.

Table 2

BTERGM Results

	Model 1 – No Selectivity		Model 2 – Moderate Selectivity		Model 3 – High Selectivity	
	θ	95% CI	θ	95% CI	θ	95% CI
Homophily†	0.02	$[-0.01, 0.04]$	-0.08	$[-0.11, -0.05]$	-0.35	$[-0.44, -0.29]$
Edge Memory	1.62	$[1.59, 1.65]$	1.63	$[1.59, 1.68]$	1.79	$[1.73, 1.85]$
2-stars‡	< 0.01	$[-0.05, 0.06]$	-0.06	$[-0.11, -0.02]$	-0.02	$[-0.06, 0.01]$
triangles‡	0.04	$[-0.03, 0.11]$	0.01	$[-0.03, 0.07]$	0.10	$[0.02, 0.16]$
4-cycles‡	< 0.01	$[-0.01, 0.01]$	0.01	$[0.01, 0.02]$	< 0.01	$[-0.01, 0.01]$

Note. CI = confidence interval. †Homophily measured as the absolute difference of party id scores; negative estimates indicate homophily while positive estimates indicate heterophily. ‡Control included to improve model fit.

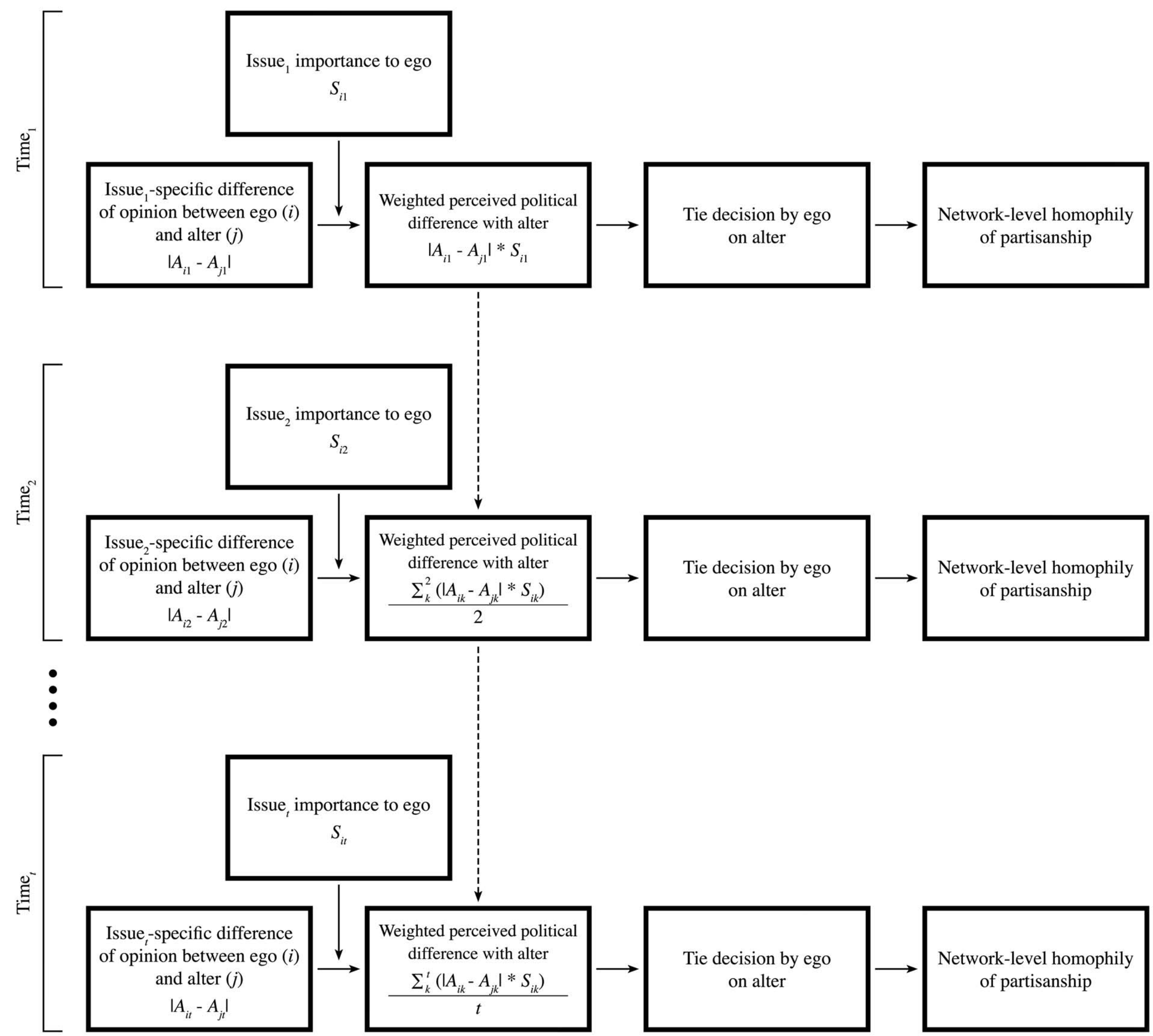


Figure 1. Theoretical path model of the dynamic selectivity function and its relation to network-level homophily. Vertical arrows between different time points indicate how the weighted perceived similarity function updates with each additional topic discussed.

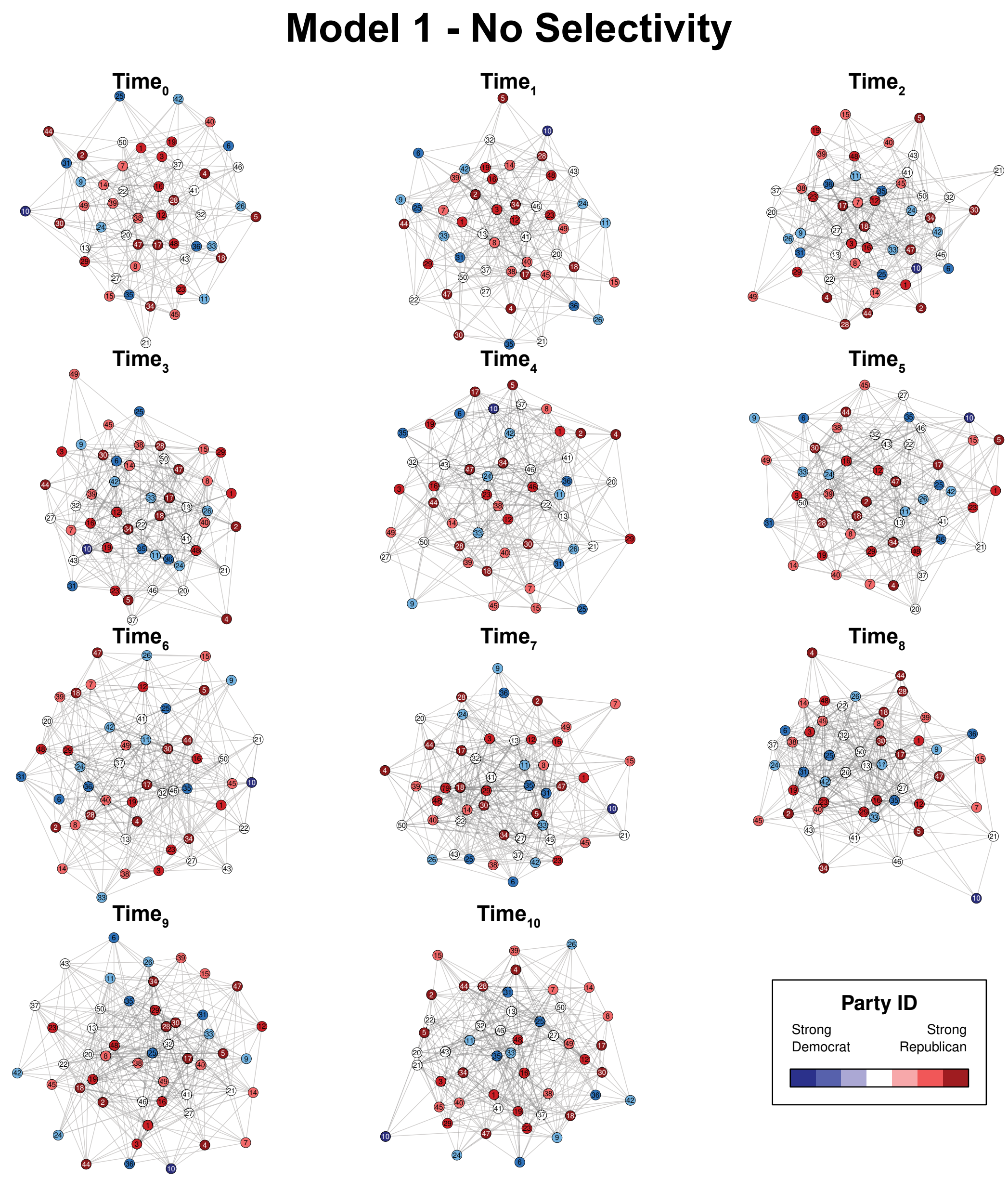


Figure 2. Network graphs of the ABM with no selectivity from Time_0 through Time_{10}. Nodes represent individual agents with a party id score indicated by the node color. Homophily would be indicated by stark segregation of these nodes by colors.

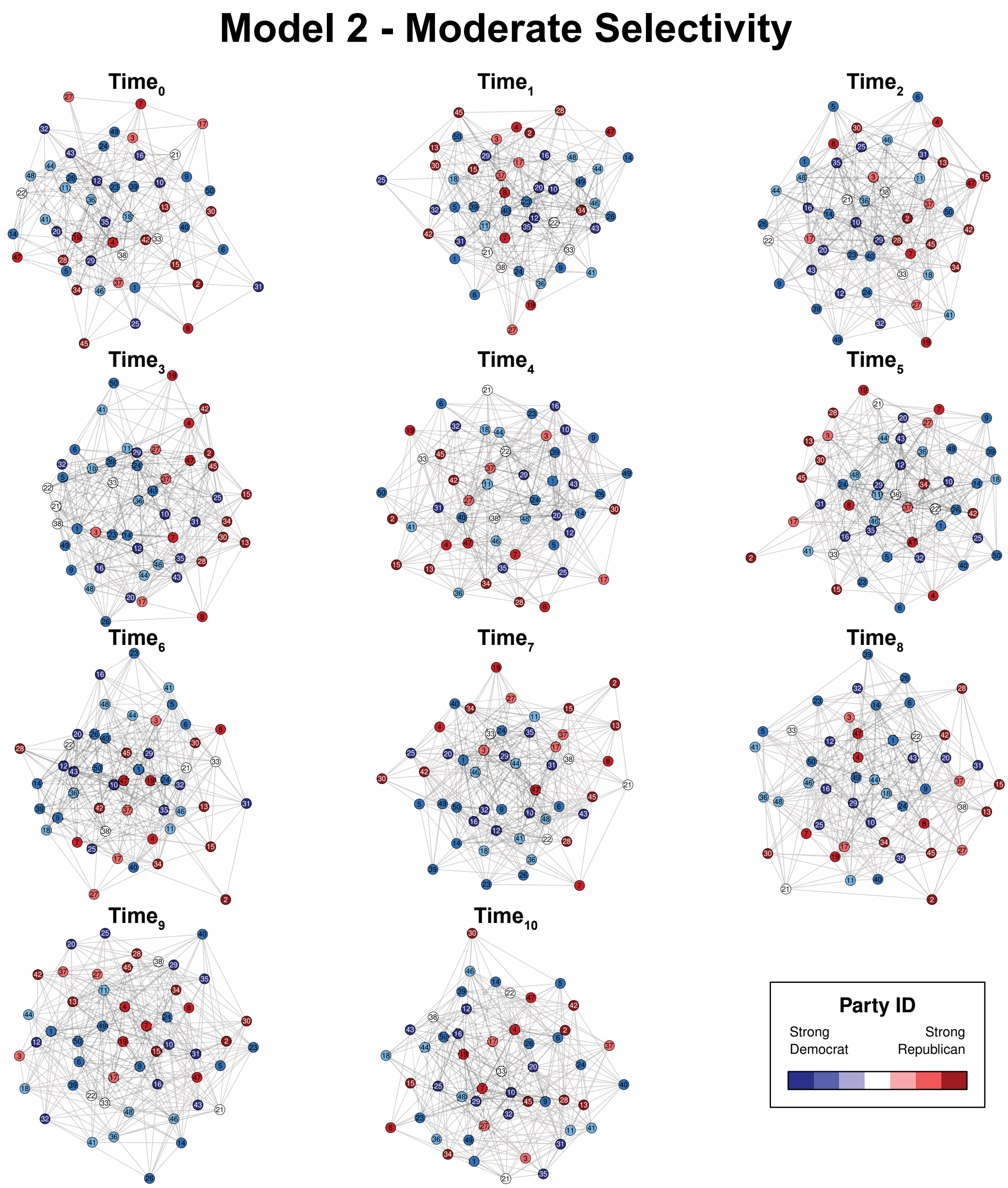


Figure 3. Network graphs of the ABM with moderate selectivity from Time$_0$ through Time$_{10}$. Nodes represent individual agents with a party id score indicated by the node color. Homophily would be indicated by stark segregation of these nodes by colors.

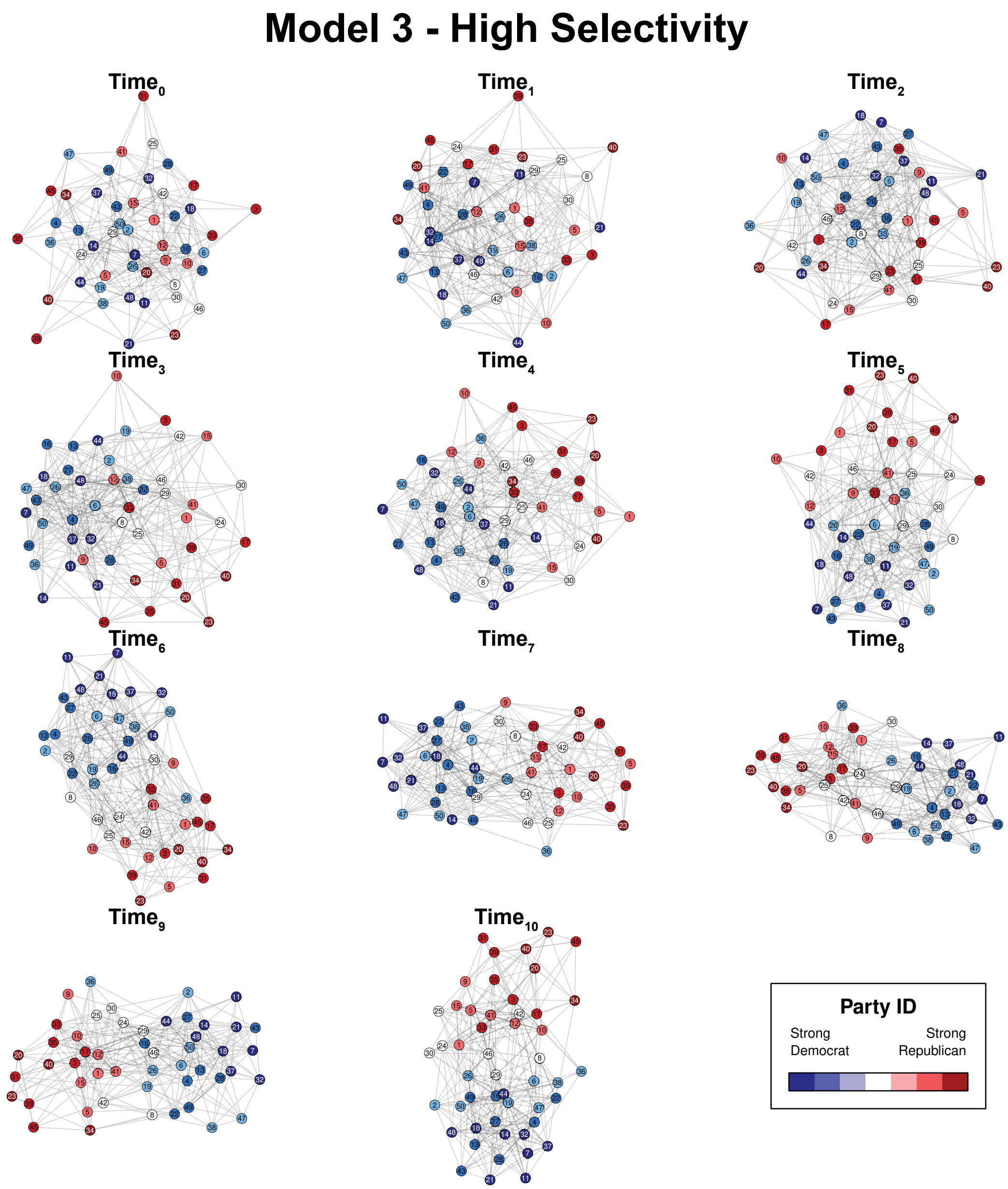
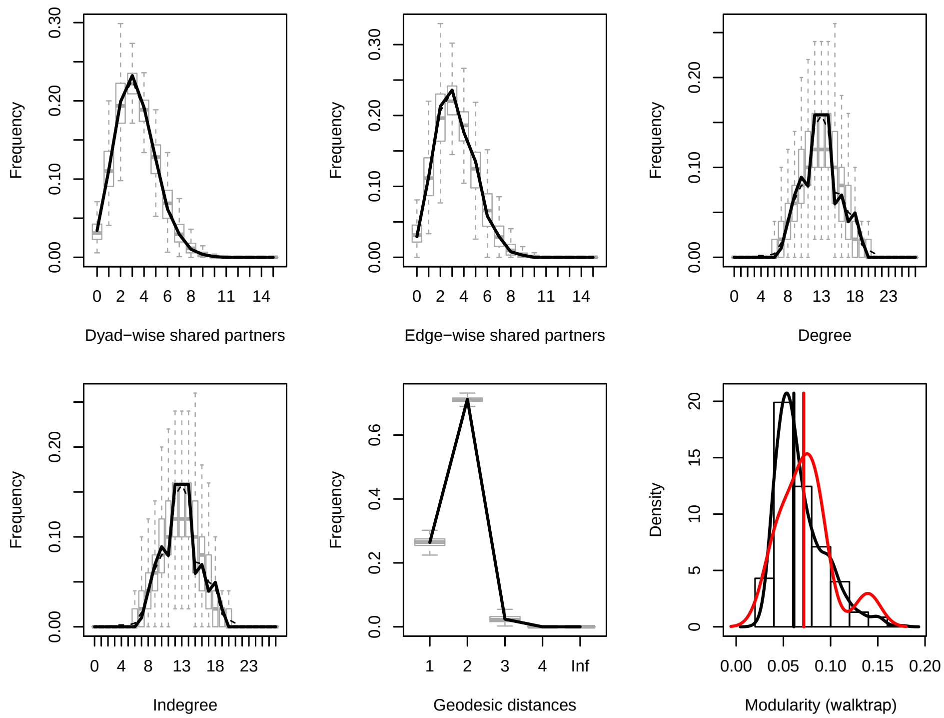


Figure 4. Network graphs of the ABM with high selectivity from $Time_0$ through $Time_{10}$. Nodes represent individual agents with a party id score indicated by the node color. Homophily would be indicated by stark segregation of these nodes by colors.

A.



B.

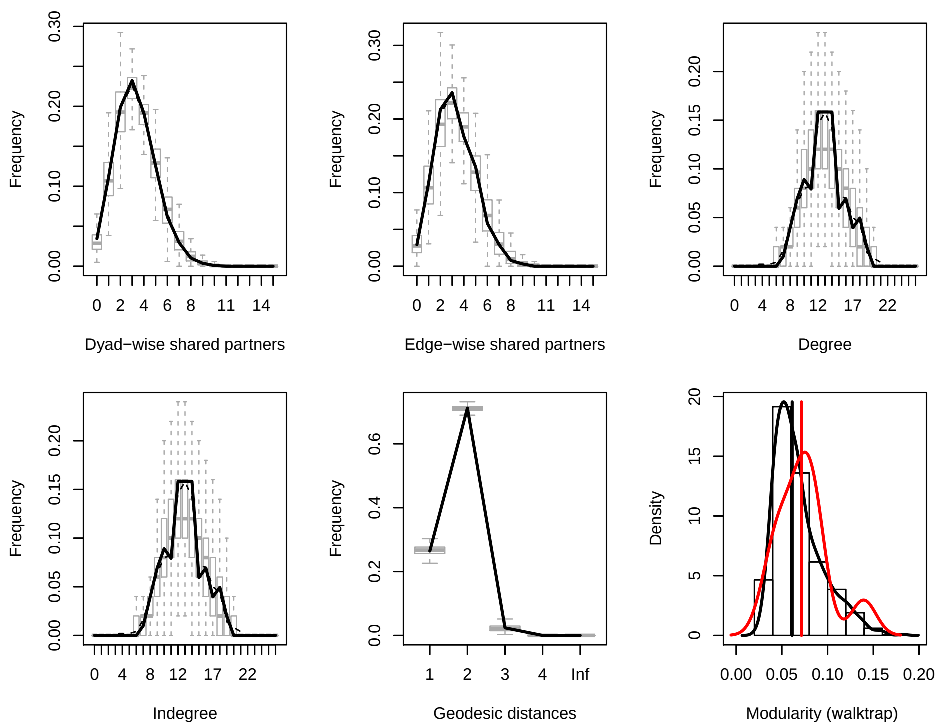
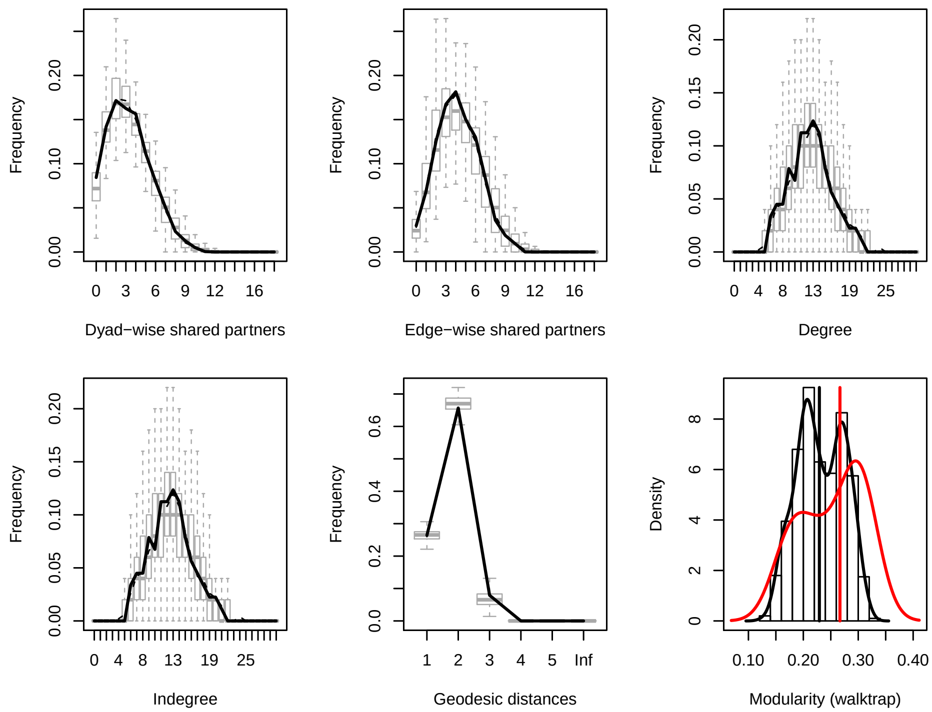


Figure 5. BTERGM model fit statistics of the ABM of moderate selectivity with (panel A) and without (panel B) homophily included in the model specification.

A.



B.

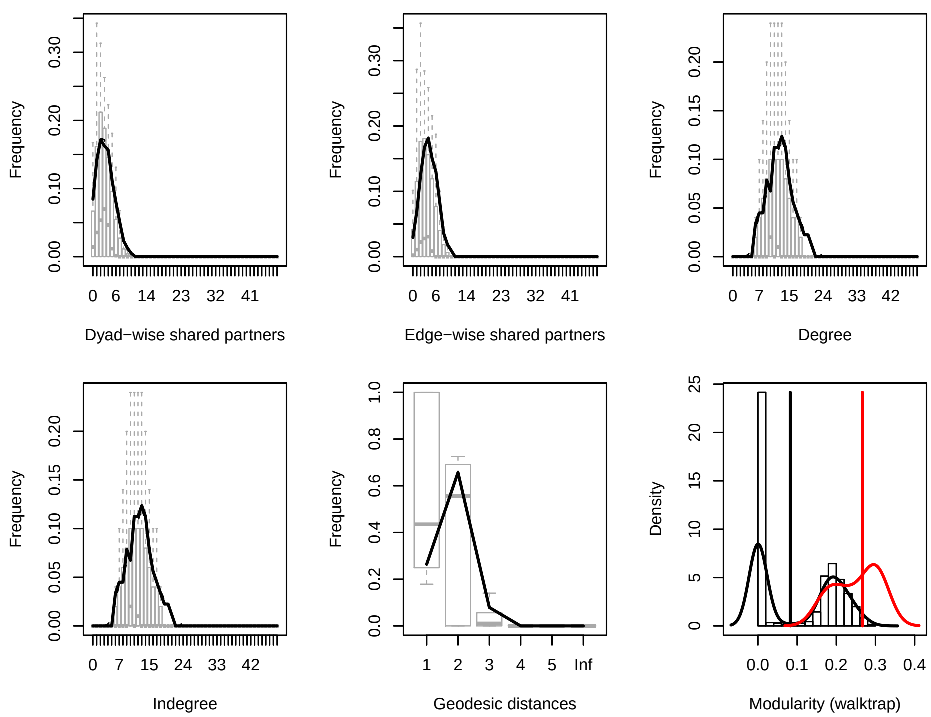


Figure 6. BTERGM model fit statistics of the ABM of high selectivity with (panel A) and without (panel B) homophily included in the model specification.