Understanding Dynamics of Polarization via Multiagent Social Simulation

Amanul Haque
Dept. of Computer Science
North Carolina State University
Raleigh, NC, USA
ahaque2@ncsu.edu

Nirav Ajmeri*
Dept. of Computer Science
University of Bristol
Bristol, UK
nirav.ajmeri@bristol.ac.uk

Munindar P. Singh
Dept. of Computer Science
North Carolina State University
Raleigh, NC, USA
mpsingh@ncsu.edu

Abstract

It is widely recognized that the Web contributes to user polarization and such polarization affects not just politics but also peoples' stances about public health, such as vaccination. Understanding polarization in social networks is challenging because it depends not only on user attitudes but also their interactions and exposures to information. We adopt Social Judgment Theory to operationalize attitude shift and model user behavior based on empirical evidence from past studies. We design a social simulation to analyze how content sharing affects user satisfaction and polarization in a social network. We investigate the influence of varying tolerance in users and selectively exposing users to congenial views. We find that (1) higher user tolerance slows down polarization and leads to lower user satisfaction; (2) higher selective exposure leads to higher polarization and lower user reach; and (3) both higher tolerance and higher selective exposure lead to a more homophilic social network.

Keywords: Echo Chambers, Selective Exposure, User Tolerance, Social Networks

1 Introduction

As the COVID-19 pandemic crosses the two-year mark, we can see that it has established a new normal, not only in the objective challenges it poses to society and business but also

*Corresponding author.

in terms of widespread attitudes and behaviors that are antivax, antimask, and antiscience. Polarization on such topics is a societal problem since it makes rational decision-making and resource allocation difficult. The Web enables fast information diffusion across traditional boundaries, which unfortunately has contributed to polarization. Specifically, social media influences users in subtle ways, especially regarding politics (Nahon, 2015); moreover, online and offline political participation is correlated (Johnson et al., 2020; Bode et al., 2014).

We simulate two factors identified by prior research that influence polarization. *First*, selective exposure to congenial (attitude-conforming) information exacerbates confirmation bias, polarizing opinions further (Stroud, 2010; Garrett et al., 2014; Kim, 2015; Westerwick et al., 2017). Selective exposure arises in and strengthens echo chambers, wherein a person encounters only beliefs or opinions that coincide with their own so that their existing views are reinforced, and alternative ideas are suppressed. Conversely, cross-cutting exposure (to uncongenial i.e., attitude-disconfirming information) has a depolarizing effect (Kim, 2015), though with caveats (Garrett et al., 2014; Kim, 2019).

Second, user tolerance for ideas that contradict their own mitigates polarization (Coscia and Rossi, 2022).

We analyze the effects of selective exposure and tolerant users on polarization among users at large. Specifically, we investigate the following research questions.

$\mathbf{RQ_{tolerance}}$. *Does higher tolerance among users in a social network help mitigate polarization?*

$\mathbf{RQ_{exposure}}$. *Does selective exposure to congenial information contribute to polarization?*

We develop a multiagent social simulation to investigate these research questions.

To address $\mathrm{RQ}_{\mathrm{tolerance}}$, we model tolerant users by having a higher tolerance level toward both opposing and congenial views. We operationalize tolerance in users using Social Judgment Theory (Sherif and Hovland, 1961), which defines tolerant people as those having a wider latitude of noncommitment. For $\mathrm{RQ}_{\mathrm{exposure}}$, we emulate selective exposure by filtering posts based on the receiving user's stance toward a given issue.

For $\mathrm{RQ}_{\mathrm{tolerance}}$, we find that tolerant users do mitigate polarization but achieve lower user satisfaction than users with lower tolerance. Surprisingly, higher tolerance also leads to a more homophilic social network. For $\mathrm{RQ}_{\mathrm{exposure}}$, we find that higher selective exposure leads to more polarization, and a more homophilic social network. Higher selective exposure leads to higher aggregate user satisfaction in the social network but with fewer satisfied users.

Analyzing the dynamics of polarization based on information sharing on social media can help us identify potential interventions. Since most content filtering (algorithmic selective exposure) in use today is based on artificial intelligence (AI), this work can help us better understand the social and political aspects of using AI. Our findings suggest avenues for further theoretical development in tandem with consideration of interventions to reduce polarization in online social networks.

Organization. The rest of the paper is organized as follows: Section 2 describes the background and discusses related work. Section 3 explains our methodology, including definitions and the simulation design, assumptions and limitations. Section 4 details the experimental

setup, results of our experimentation, and statistical analysis of the results. Section 5 includes a discussion on results, and threats to the validity of this work and concludes with future directions.

2 Background and Related Work

The Theory of Cognitive Dissonance (Festinger, 1957) asserts that when a person is confronted with contrasting ideas, it causes psychological discomfort making that person more selective in their information consumption, potentially causing confirmation bias. *Confirmation bias* is the tendency of people to accept "confirming" evidence at face value while subjecting "dis-confirming" evidence to critical evaluation (Lord et al., 1979), resulting in people gravitating toward information that aligns with (confirms) their existing views. Bias exists in the selection and sharing of information, especially news (Hart et al., 2009; Knobloch-Westerwick, 2014).

Selective exposure is a tendency of people to choose and spend more time on information that is consistent with their existing beliefs (Klapper, 1960; Redlawsk, 2002; Taber and Lodge, 2006), though some prior works suggest that partisan selective exposure may be a myth (Kinder and Sears, 1981; Zaller, 1992). Freedman and Sears (1965) argue against voluntary selective exposure in favor of *de facto selectivity*. They claim that most examples of selectivity in mass communication can be attributed to complex factors such as demography, education, social connections, and occupation, which are incidental to their supportiveness to the receiver's existing beliefs. People prefer supportive information in some situations while dissonant information in other situations (Hargittai et al., 2008). Individuals with strong preferences are more likely to spend more time reading negative (uncongenial) information about their choice (Meffert et al., 2006), perhaps to critique it (Hargittai et al., 2008).

2.1 Social Media and Politics

The number of users on social media platforms has increased rapidly over the years. Only 8% of the Internet users in the US used some social networking platform in 2005 (Lenhart, 2009), whereas in 2021, 69% use Facebook, and 40% use Instagram (Auxier and Anderson, 2021). The use of social networking sites for political discussions has also increased over the years. Social media is now among the most common ways in which people, particularly young adults, obtain their political news (Infield, 2020). A meta-analysis from 36 past studies assessing the relationship between social media use and participation in civic and political life found a positive correlation between the two, with more than 80% of the coefficients as positive (Boulianne, 2015). Polarization measured based on online social interactions shows a good correlation with offline polarization (Morales et al., 2015). Adults who use social networking platforms as a political tool are more likely to participate in politics (Bode et al., 2014). This is true across various cultural and geographical boundaries, including empirical evidence from the US (Infield, 2020), Pakistan (Ahmad et al., 2019), and Taiwan (Zhong et al., 2022).

Selective exposure to political information is correlated with polarizing people's opinions to align with the values of the political party they support (Stroud, 2010; Garrett et al.,

2014; Kim, 2015; Westerwick et al., 2017). Though the causal direction, i.e., whether selective exposure leads to polarization or the other way around, is less obvious (Stroud, 2010). Stroud (2010) investigate the causal relationship between partisan selective exposure and polarization and find strong evidence suggesting selective exposure leads to polarization while finding limited evidence suggesting the reverse causal direction. Schkade et al. (2007) find that intragroup deliberation on social issues among like-minded people leads to more extreme and less diverse ideological beliefs, while Bail et al. (2018) observe that exposure to opposing views on social media can increase political polarization. Habitual online news users are less likely to exercise selectivity to get attitude-consistent exposure, which reduces their likelihood of participating in the political system (Knobloch-Westerwick and Johnson, 2014). The longer individuals spend on attitude-consistent content associated with biased sources, the more immediate attitude reinforcement occurs, and its influence can be detected even after a couple of days of exposure (Westerwick et al., 2017).

Cross-cutting exposure refers to being exposed to oppositional viewpoints. Cross-cutting exposure in social networks fosters political tolerance and makes individuals aware of legitimate rationales for oppositional viewpoints (Mutz, 2002b). Exposure to disagreeing viewpoints contributes to people's ability to generate reasons, particularly why others might disagree with their view (Price et al., 2002). Kim and Chen (2016) find that exposure to cross-cutting perspectives results in a higher level of political engagement, though this increase may depend on the social media platform used.

Cross-cutting exposure, widely assumed to encourage an open and tolerant society, is not necessarily the environment that produces enthusiastically participatory individuals. People belonging to social networks involving greater political disagreement are less likely to participate in politics (Mutz, 2002b,a). Constant exposure to disagreement may necessitate trade-offs in other social network characteristics such as relationship intimacy and frequency of communication (Mutz, 2002b). Conflict-avoiding individuals, in particular, are more likely to respond negatively to cross-cutting exposure by limiting their political participation to avoid confrontation and putting their social relationships at risk (Mutz, 2002a).

Garrett et al. (2014) examine survey data following elections in the US and Israel and find consistent results despite cultural differences. Their findings suggest that pro and counter-attitudinal information exposure has a distinct influence on perceptions of and attitudes toward members of opposing political parties.

Mutz (2002a) analyzes the consequences of cross-cutting exposure on political participation. They find that people whose social networks involve greater political disagreement are less likely to participate in politics and are more likely to hold politically ambivalent views.

Though many studies have investigated polarization using empirical data from social media, a common limitation has been that past studies either look at one-time exposure or study these effects in isolation. For instance, Stroud (2007) studies the effects of selective exposure using empirical evidence but relies on data from one-time exposure and studies the immediate effects without differentiating the long-term effects. However, the evidence from past studies suggests that political participation and its effect is a long-term process that unfolds over time based on multiple exposures (Gerber et al., 2003; Valentino and Sears, 1998). Further, existing research has focused chiefly on effect at an individual level, i.e., relying on self-reported data of how an individual's stance is influenced by exposure to potentially polarizing content. However, self-reporting is susceptible to user bias and

overlooks how changes in one part of the social network can influence other parts.

2.2 Multiagent Social Simulation

Many earlier models on opinion and influence propagation are based on a centralized diffusion process, overlooking the decentralized nature of information diffusion in social networks.

Kempe et al. (2003) design two fundamental diffusion models for influence maximization, namely, the Independent Cascade Model (ICM) and the Linear Threshold Model (LTM). Influence in these models is transferred through the correlation graph starting from a set of seed nodes (activated nodes). Influence decreases when hopping further away from the activated node.

Jiang et al. (2017) design a preference-aware and trust-based influence maximization model called the Preference-based Trust Independent Cascade Model (PTICM) that takes into account user preferences and trust between users in computing influence propagation.

Li et al. (2019) design a novel agent-based seeding algorithm for influence maximization named Enhanced Evolution-Based Backward selection that models individual user preferences and social context based on social influence and homophily. Their results suggest that individuals are influenced by their social context much more than retaining their own opinions. Though the Prior Commitment Level (PCL) of a user is an essential factor for influence propagation, users tend to revise their PCL over time.

Chen et al. (2020) propose a group polarization model based on the SIRS epidemic model and factor in the relationship strength based on the J-A (Jager and Amblard) model. They use a BA network model due to its closeness to the real-world social network structure and a Monte Carlo method to conduct simulation experiments.

Kozitsin and Chkhartishvili (2020) develop an agent-based model to explore how agents' activity patterns affect the formation of echo chambers. They use a personalizing system algorithm to control mutual interactions among agents and decide what information the agents are exposed to. They find that the critical parameter that guides agents' opinion dynamics is the probability of publishing a post, i.e., agents who often publish posts tend to enter echo chambers.

Hązła et al. (2019) use a geometric model of polarization and demonstrate that societal opinion polarization often arises as an unintended byproduct of influencers attempting to promote a product or an idea. Gaitonde et al. (2021) extend this work to show that the exact form of polarization in such models is quite nuanced. Even when strong polarization does not hold, weaker notions of polarization can attain nonetheless.

Baumann et al. (2020) propose a radicalization model that uses a reinforcement mechanism to drive opinions to extremes starting from moderate initial conditions. They show that the transition from a global consensus to a radicalized state is mostly governed by social influence and the controversialness of topics discussed.

Wang et al. (2019) model a rumor-propagation framework based on information entropy to understand information distortion and its polarization effects in social networks. They find that mass polarization toward a positive or negative consensus occurs when a synergistic mechanism between preferential trust and polarization tendencies is sustained. The segregation of the population into groups of different polarities happens under certain conditions.

We design a multiagent social simulation to emulate information diffusion on social networks. We model user behavior based on existing social science theories and empirical evidence from prior studies.

3 Methodology

We now describe our social simulation model and agents' interaction.

3.1 Social Simulation Definitions

Definition 1 (Social Network) *A* social network *is an undirected graph with nodes representing users and the links connecting the nodes representing a relationship between two users.*

A social network is represented as $G = (nodes, edges)$, where $nodes = \{a_1, \ldots, a_n\}$ are users and $edges = \{(a_1, a_2), (a_4, a_9), \ldots, (a_x, a_y)\}$ represent a direct connection between pair of users in the social network. An agent can only interact with its neighbors in the social network.

Definition 2 (Agent) *An* agent *represents a user in the social network.*

Each agent is independent and has attributes defining its preferences such as *user activity*, and *sharing preference*. User activity captures how active an agent is, and sharing preference captures agents' willingness to share a post on the social network. Both range over [0, 1] (0 represents most inactive/unwilling and 1 most active/willing). An agent is capable of taking two actions, sharing a post, and providing sanctions to posts.

Definition 3 (Post) *A* post *is a message shared by an agent with its neighbors in the social network.*

Agents in a social network interact by sharing posts that can be represented as $Post = (a, t, s)$, where a is the author, t is the topic mentioned in (or discussed in) the post, and s is the stance of the post towards the topic (continuous value in $[-1, 1]$, where -1 represents extreme opposition and 1 extreme support for the issue).

A post serves as a timestep and is used to track changes in the social network over time. Updates to the social network and agent's attributes are made after a post has completed diffusion in the social network (i.e., it has reached as many agents as possible).

Definition 4 (Sanction) *A* sanction *is a reaction an agent has for a post it receives.*

Sanctioning provides a foundation for how participants in a sociotechnical system (STS) may seek to influence each other's decision making and steer the STS towards their preferred direction (Nardin et al., 2016). Agents provide positive sanctions to congenial posts and negative to uncongenial posts based on their stance on a given topic being discussed in the post. Sanctioning is analogous to providing likes and comments to a post and captures whether a user approves (likes) or disapproves (dislikes) the topic in a received post.

Definition 5 (Issue) *An* issue *refers to the topic being discussed in a post.*

Issues are predefined, and all agents hold a stance on each issue. An agent's stance toward an issue is represented as a continuous value between [−1,1], with −1 indicating extreme opposition, and 1 extreme support for the issue. Each agent has an overall POV (point-of-view) that depends on its stance on various issues. The POV of an agent is computed as the mean of its stance on all issues. POV ranges between [−1, 1], with −1 representing extreme support for POV-1 (<0), 0 means neutral POV, and 1 extreme support for POV-2 (>0).

With respect to a post, an agent can be in one of the four states: (1) *Nonreceiver:* Agents who have not yet received the post (all agents other than the author are in this state at the start of the simulation); (2) *Receiver:* Agents who have received the post (but not yet shared it); (3) *Spreader:* Agents who have shared the post with their friends; and (4) *Disinterested:* Agents who received the post but chose not to share it further and lost interest in the post.

3.2 Social Simulation Model

The simulation starts with an agent (a_x) sharing a post (p_k) with its neighbors in the social network. The receiver then decides whether to share the received post further with a probability of sharing that depends on the content of the post and the receiver's preferences. An agents' preference involves its sharing preference, how active the agent is on the social network, and the agent's stance toward the issue (supporting vs. opposing). The content of a post includes the issue mentioned in the post and the post's stance toward the issue. Equation 1 describes the computation for sharing probability $sP(a_{\mathrm{x}}, p_{\mathrm{k}})$ for the agent a_{x} to share the post p_{k} it received.

$$sP(a_x, p_k) = c_1 \times uA(a_x, p_{k-1}) \times |uS(a_x, i, p_{k-1}) \times pS(p_k, i)| \times sPref(a_x, p_{k-1}) \tag{1}$$

c_1 is a constant, a_x is the receiver, p_k is the k^{th} post being shared in the social network, and i is the issue being discussed in the shared post. $uA(a_x, p_{k-1})$ is the user activity of user a_x before the post p_k is shared, $uS(a_x, i, p_{k-1})$ is the user a_x's stance towards issue i before the post p_k is shared, $pS(p_k, i)$ is the stance of the post towards issue i, and $sPref(a_x, p_{k-1})$ is the sharing preference of user a_x before the post p_k is shared. An agent with low $sPref(a_x, p_{k-1})$ is more likely not to share a post further and may enter the state *Disinterested*. Disinterested agents are not candidates for sharing the post (p_k) further.

The agents who receive the post provide a sanction. Sanctions can be positive or negative. Sanctions by the receiver depend on how active the receiver is, its stance toward the issue at hand, and the post's stance toward the issue. Sanction by an agent a_y for a post p_k it received from agent a_x is computed using Equation 2.

$$Sanc(a_y, p_k, a_x) = c_1 \times uA(a_y, p_{k-1}) \times uS(a_y, i, p_{k-1}) \times pS(p_y, i) \tag{2}$$

$Sanc(a_y, p_k, a_x)$ is a sanction provided by agent a_y for the post p_k it received from agent a_x. Sanction scores affect user activity and the stance of each agent towards an issue. Agents

prefer positive sanctions (social acceptance), which increases their activity on the platform, while negative sanctions discourage agents from sharing their views in the future, hence reducing their participation (user activity). The update in user activity depends on the sanctions received by an agent for the posts it shared. An agent's user activity ($uA(a_x, p_k)$) after sharing a post p_k is computed using Equation 3.

$$uA(a_x, p_k) = uA(a_x, p_{k-1}) + c_2 \times \sum_{a_i \in neighbor(G, a_x, p_k)} Sanc(a_i, p_k, a_x) \tag{3}$$

c_2 is a constant, $uA(a_x, p_{k-1})$ represents the user activity of agent a_x before the post p_k is shared, and $uA(a_x, p_k)$ represents the user activity of agent a_x after the post p_k is shared. $neighbor(G, a_x, p_k)$ refers to all neighbors of agent a_x in the social network G that receive the post p_k directly from agent a_x.

An agent's stance toward an issue is influenced by the sanctions it receives from other agents. We model this shift in the stance of an agent using Social Judgment Theory (SJT) (Sherif and Hovland, 1961), which describes how individuals change their position when confronted with a competing position on a given issue. According to SJT, an individual shifts their stance in the direction of the competing stance if the competing stance falls within their *latitude of acceptance* (assimilation). In contrast, they will shift away from the competing stance if the competing stance falls beyond their *latitude of rejectance* (contrast). For instance, for an agent a_x, that has a stance of $uS(a_x, i, p_k)$ towards issue i, a threshold determining the latitude of acceptance u_{xi} and a threshold determining the latitude of rejection t_{xi} with $t_{xi} > u_{xi}$. When this agent a_x interacts with another agent a_y, the following rules are applied to compute the shift in the stance of agent a_x towards an issue i.

$$\text{diff\_Stance}(a_x, a_y, i, p_k) = |uS(a_x, i, p_k) - uS(a_y, i, p_k)| \tag{4}$$

diff_Stance(a_x, a_y, i, p_k) is the absolute difference in the stances of agent a_x and agent a_y on issue i as the post p_k is being shared.

$$\begin{aligned} \text{If diff\_Stance}(a_x, a_y, i, p_k) < u_{xi}, &\quad \delta uS(a_x, a_y, i, p_k) = \mu \times (uS(a_y, i, p_k) - uS(a_x, i, p_k)) \\ \text{If diff\_Stance}(a_x, a_y, i, p_k) > t_{xi}, &\quad \delta uS(a_x, a_y, i, p_k) = \mu \times (uS(a_x, i, p_k) - uS(a_y, i, p_k)) \\ &\text{else } \delta uS(a_x, a_y, i, p_k) = 0 \end{aligned} \tag{5}$$

μ represents the strength of the influence between two agents. We assume the same strength of influence between all pairs of connected agents in the social network; hence the value of μ is 1. The shift in the stance of an agent a_x for sharing posts p_k on issue i is computed using the received sanction scores and the difference in stance (toward the issue at hand) between the author or spreader (i.e., a_x) of the post, and the receiver (i.e., a_y) (Equation 6).

$$\Delta S(a_x, a_y, i, p_k) = c_2 \times \frac{Sanc(a_y, p_k, a_x)}{\delta uS(a_x, a_y, i, p_k) + 1} \tag{6}$$

$\Delta S(a_x, a_y, i, p_k)$ is the shift in stance (of agent a_x) due to a sanction (by agent a_y) for a post p_k it shared on issue i.

User stance after sharing post p_k can be computed using Equation 7.

$$uS(a_x, i, p_k) = uS(a_x, i, p_{k-1}) + \sum_{a_j \in neighbor(G, a_x, p_k)} \Delta S(a_x, a_j, i, p_k) \tag{7}$$

$uS(a_x, i, p_{k-1})$ is the stance of the agent a_x on issue i before it shares post p_k, and $uS(a_x, i, p_k)$ is the stance of an agent a_x on issue i after the posts p_k is shared and sanctions for it received from all other agents. The maximum allowed change in stance due to one post is 0.20 and we bound user stance within $[-1, 1]$ by restricting the values.

The codebase[1] of our social simulation is publicly available.

3.3 Agent Goals and Actions

The simulation progresses with agents sharing posts with other agents, causing each post to diffuse further in the social network. Each post receives a sanction from all agents that receive it, and these sanctions, in turn, influence its authors' (spreaders') activity score and stance toward various issues. An agent supports a POV with which its aggregate stance toward various issues is in agreement. Agents can take two actions, sharing a post and sanctioning a received post. Agents in the simulation try to maximize their influence and popularity in the social network by sharing relevant content and providing appropriate sanctions. Accordingly, we define two goals for each agent—*Promoting Views* and *User Satisfaction*.

Promoting Views. All agents try to promote their views (POVs) on different issues by sharing relevant posts with their friends (neighbors in the social network). Agents also achieve this by providing positive sanctions to what agrees with their views and negative sanctions to what does not.

User Satisfaction. All agents try to maximize their satisfaction. User satisfaction is computed based on the sanctions received from other agents. Agents change their stance toward issues to ensure more aggregate positive sanctions over time.

3.4 Simplifying Assumptions and Limitations

We make simplifying assumptions to operationalize user attributes and online sharing behavior. *First*, we assume views (on an issue) to be binary in this simulation, i.e., either supporting POV-1 or POV-2, meaning agents with no POV are nonparticipating. This is a design choice as we intend to analyze the scenario where only motivated agents (i.e. agents who have a POV) try to influence and promote their views. As an agent becomes neutral in its POV (i.e., an agent with POV as zero), it stops sharing posts and providing sanctions. We assume all agents have some POV at the start of the simulation, and no agent has a neutral POV.

Second, we assume the initial user attributes and stance of each post based on a probability distribution. We use a random normal distribution to populate initial user attributes including the agent's stance towards an issue, sharing preference, and post's stance. This

[1] https://github.com/ahaque2/MultiAgent-Social-Simulation.git

ensures a balance of stance toward each party across issues and provides a reasonable starting condition for the simulation.

Third, we assume all agents prefer getting positive sanctions over negative or none. They accordingly change their stance on issues over time to ensure social acceptance (i.e., to get aggregate positive sanctions from their neighbors). Sanctions also influence user activity; positive sanctions cause higher user activity while negative sanctions cause it to decline.

Our simulation models user preferences and emulates user behavior on social networks to analyze polarization dynamics. However, our model has a few limitations that stem from the simplifications (of user behavior and its influence).

First, for simplicity, sharing of posts and opinion shifts are sequential in this simulation, i.e., only one post is being shared in the network at any given time. Another post starts diffusing in the network only when the previous post has completely diffused (i.e., has reached all agents it could have). This limits the simulation to not factor in the effects of parallel exposure to different (maybe conflicting) information, i.e., being exposed to several posts relating to an issue before forming (shifting) an opinion about it.

Second, the social network in this simulation is static, i.e., neither a new link is formed nor an existing one severed at any time. However, selective exposure partially makes the network dynamic by filtering posts based on the difference in stance between two agents towards an issue. A dynamic social network demands far more computational resources and some knowledge of the offline world to link or delink agents over time appropriately.

4 Experiments and Results

We now describe the experimental setup and the metrics used to measure changes in the social network followed by results.

4.1 Initial Simulation Setup

We use the Facebook social network from Leskovec and Mcauley (2012) to seed the simulation. The social network consists of 4,039 nodes (agents) and 88,234 edges (neighbors) and an average clustering coefficient of 0.61.

The agents in the social network interact by sharing posts from a pool of artificially generated posts without replacement. The stance of the posts follows a bounded normal distribution (μ=0.00, σ=0.52, min=-1, max=1) such that there is equal support and opposition for each issue. We predefine six issues and generate an equal number of posts for each issue. We use a total of $\approx$5,000 posts that are shared between agents in each run of the experiment. Each simulation run ends when all posts in the pool of generated posts have been shared in the social network.

We create ten independent initial distributions to assign different initial user attributes for each simulation run. We set initial user satisfaction to zero for all agents. Each agent is initialized with a sharing preference based on a random normal distribution bounded between 0 and 1 (average over all distributions, μ=0.5, σ=0.14, min=0, and max=1). User activity is initialized based on a tailed distribution bounded between 0 and 1, skewed towards higher values (average over all distributions: μ=0.874, σ=0.17, min=0, and max=1). Higher initial

user activity ensures greater activity and faster results. We compute kurtosis (Zwillinger and Kokoska, 1999) for all user activity distributions. The average kurtosis (over all ten distributions of user activity) was 1.54 (for a normal distribution, kurtosis is zero).

We assume two POVS (Point-Of-Views), POV-1 and POV-2. Each agent has a POV in [−1, 1] that depends on its stance on various issues. Each agent's stance towards different issues is initialized based on a random normal distribution bounded in [−1, 1] centered around zero. The stance distribution is such that on aggregate there is equal support and opposition for each issue. The POV of each agent is computed as the average stance on issues favoring each POV resulting in a normal distribution in [−1, 1] approximately centered around zero (average over all distributions, μ =0.01, σ =0.11, min=-0.40, and max=0.44). This ensures there is approximately equal support for each POV at the start of the simulation.

We ensure consistency between the agent stance who authors and shares the post and the stance of the post by choosing the authors appropriately. If an agent supports issue A, it will only start a supportive post on issue A, whereas an agent who opposes it starts only a critical one on that issue. Agents are chosen to be authors of a post based on their activity score and sharing preference half of the time and at random for the other half. Agents who are more active or have a higher sharing preference are more likely to start sharing a post.

4.2 Metrics

We define primary and secondary metrics to measure various changes in the network over time. Primary metrics focus on measuring polarization and user satisfaction, while secondary metrics compare initial and final user distribution for different user attributes for each experiment.

4.2.1 Primary Metrics

Primary metrics include the following.

Polarization. Polarization measures the extent to which the resulting distribution of opinions is polarized. We adopt the polarization index measure proposed by Morales et al. (2015) to measure overall polarization in the social network. The polarization index is inspired by the electric dipole moment and measures polarization as the distance between two opposing ideologies. Polarization lies in [0, 1] with 0 indicating least polarization and 1 indicating most.

To compute polarization we define A^- as the relative population with POV-1 (i.e., negative POV, <0) and A^+ as the relative population with POV-2 (i.e., positive POV, >0). We compute the normalized difference in the populations using the relative populations A^- and A^+.

$$\Delta A = |A^+ - A^-| \tag{8}$$

We then compute the gravity center (mean) of each population, gc^- and gc^+, and define the pole distance, d, as the normalized distance between the two gravity centers. d can be expressed as.

$$d = \frac{|gc^+ - gc^-|}{|max(A^+) - min(A^-)|} \tag{9}$$

$\max(A^+)$ expresses the maximum possible value for positive opinions (POV>0), and $\min(A^-)$ expresses the minimum possible value for negative opinions (POV<0).

The network polarization ($Polarization(G, p_k)$) after the post p_k is shared on the social network is defined based on the function of the difference in size between the population of both POVs (ΔA) and the pole distance d.

$$Polarization(G, p_k) = (1 - \Delta A)d \tag{10}$$

Polarity. Polarity is indicative of the POV that has greater aggregate support in the social network. We measure polarity as the mean POV of all agents. Polarity ranges over $[-1, 1]$, with -1 indicating absolute support (by all agents) for one POV (POV-1) and $+1$ for the other (POV-2), and 0 for neutral POV.

$$Polarity(G, p_k) = \sum_{a_i \in G} \frac{POV(a_i, p_k)}{numAgents(G)} \tag{11}$$

Homophily. Homophily measures the homogeneity of a network structure with respect to some attribute (i.e., the agents' POV in this case). Homophily is shown to be useful in link prediction between users in a social network (Yuan et al., 2014). Higher homophily is indicative of greater segregation in the social network. We use the *assortativity* of a social network (Newman, 2003) to measure homophily. The value of homophily ranges over $[-1, 1]$, with 1 indicating a perfectly assortative network and values in $[-1, 0]$ indicating a perfectly disassortative network.

$$Homophily(G, p_k) = \frac{\sum_i e_{\mathrm{ij}} - \sum_i a_{\mathrm{i}} b_{\mathrm{j}}}{1 - \sum_i a_{\mathrm{i}} b_{\mathrm{j}}} \tag{12}$$

where e_{ij} is the fraction of edges in a network that connects a vertex of type i to one of type j, and a_i and b_j are the fractions of each type (based on the agents' POV) of the end of an edge attached to vertices of type i, and type j respectively. The type depends on the agent's POV, and we group agents into 20 equally spaced groups based on their POV. We use the networkx [2] implementation of assortativity to compute network homophily.

User Satisfaction. User satisfaction measures how satisfied the overall social network is based on the outcome of individual user actions. To operationalize the computation for user satisfaction (for each agent), we use the sanction scores that an agent gets for sharing posts with other agents in the social network to compute the update in user satisfaction (Equation 13). We take the mean of each user's satisfaction to compute overall network satisfaction (Equation 14).

[2] https://networkx.org/documentation/stable/reference/algorithms/assortativity.html

$$uSat(a_x, p_k) = uSat(a_x, p_{k-1}) + c_2 \sum_{a_i \in neighbor(G, a_x, p_k)} Sanc(a_i, p_i, a_x) \tag{13}$$

$$netSat(G, p_k) = \sum_{a_i \in G} \frac{uSat(a_i, p_k)}{numAgents(G)} \tag{14}$$

where $uSat(a_x, p_k)$ refers to the user satisfaction of agent a_x after the post p_k has been shared, $uSat(a_x, p_{k-1})$ refers to the user satisfaction of agent a_x before the post p_k has been shared, and $netSat(G, p_k)$ measures the overall network user satisfaction after post p_k has been shared.

4.2.2 Secondary Metrics

We define secondary metrics to compare user distribution (based on count) in the initial (at the start of the simulation run) and final (after completion of each simulation run) populations. We define three secondary metrics based on user attributes (such as user activity and user's POV), and the primary metric on user satisfaction. Secondary metrics are computed after all posts are shared. Secondary metrics include the following.

Satisfied users. User distribution (percentage) in initial and final populations with negative (<0), zero (=0), or positive (>0) user satisfaction.

Active users. User distribution (percentage) in initial and final populations with low (<0.75), medium (>0.75 and <0.90), or high (>0.90) user activity.

Polarized users User distribution (percentage) in initial and final populations with high (>0.10 or <−0.10) or low (>−0.10 and <0.10) intensity of POVs.

Table A.2 describes the secondary metrics and lists their thresholds.

4.3 Experiments

To address $\text{RQ}_{\text{tolerance}}$ (*Does higher tolerance among users in a social network help mitigate polarization?*), we vary agents' tolerance levels. To address $\text{RQ}_{\text{exposure}}$ (*Does selective exposure to congenial information contribute to polarization?*), we vary the levels of selective exposure in our simulation. We analyze the influence of changing these configurations on the primary and secondary metrics.

To mitigate the effects of stochasticity we run the simulation ten times with different initial distributions for the agent's attributes while keeping the social network and shared posts the same to ensure a fair comparison. For each experiment, we compute the primary and secondary metrics. The reported results are averages of ten simulation runs.

Figures 1 and 2 compare how polarization, polarity, homophily, and user satisfaction change with more posts being shared under different experimental settings. Tables 1 and 2 summarize our findings for the two experiments. Tables 4 and 5 include results from the statistical analysis. Tables A.1 and A.2 include a description of the notation used to explain

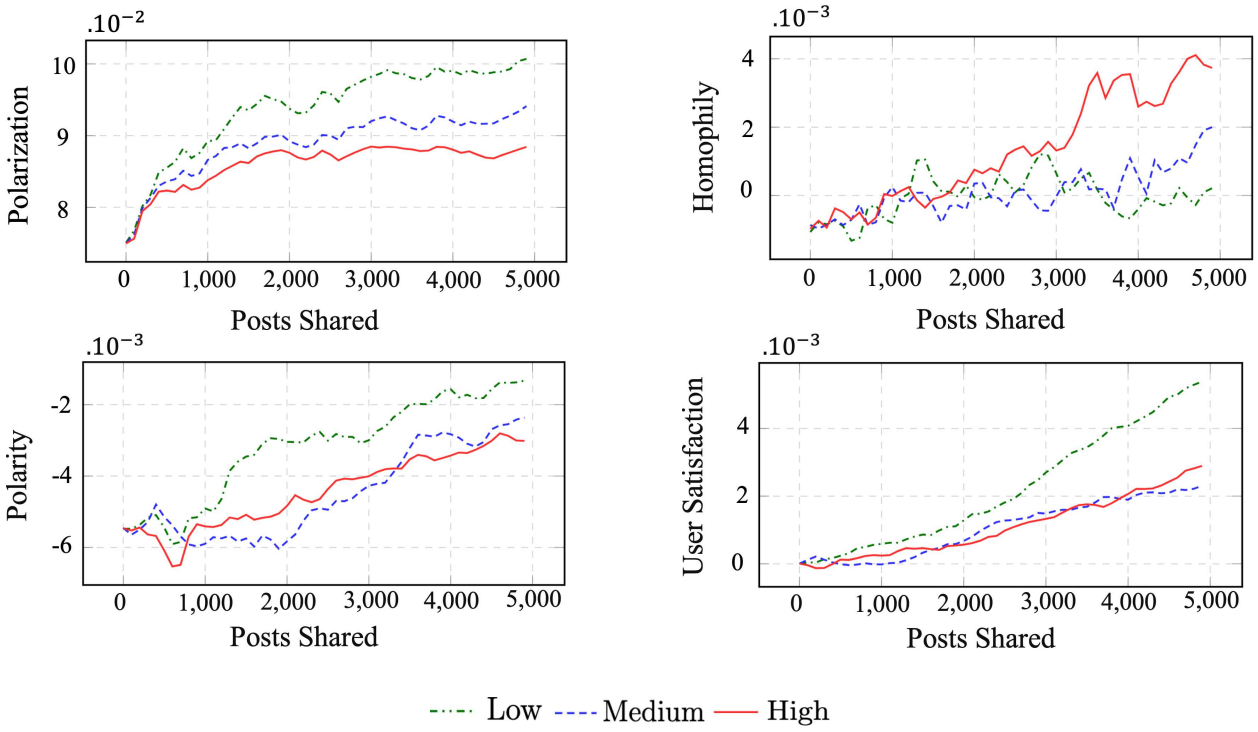

Figure 1: Experiment 1 (Tolerance): Comparing polarization, homophily, network polarity, and user satisfaction of agents in a social network with different tolerance levels.

the simulation design and metrics, respectively. Sections 4.3.1 and 4.3.2 describe the experimental setup and results of the two experiments in detail.

4.3.1 Experiment 1: Tolerant Users

The *tolerance* of an agent is defined based on its latitude of noncommitment (Sherif and Hovland, 1961), i.e., the difference between the latitude of acceptance (assimilation) and latitude of rejectance (contrast). The higher difference implies more tolerance. A more tolerant agent is less reactive to sanctions it receives from other agents for its shared posts, i.e., a more tolerant agent is less likely to change its stance on issues based on sanctions from agents who differ from its stance above a threshold (level of tolerance).

We run the simulation with three levels of tolerance, namely, HIGH, MEDIUM, and LOW. HIGH tolerant agents have a higher latitude of noncommitment (70%) and change their stance only based on sanctions from agents within a smaller (30%) difference in stance (between receiver and spreader) toward an issue. If a HIGH tolerant agent receives a sanction from an agent who differs in stance (on the issue in the shared post) greater than 30%, it discards that sanction and does not update its stance. MEDIUM tolerant agents have a latitude of noncommitment as 40%, and LOW tolerant agents have a latitude of noncommitment as 10%.

Figure 1 shows changes in the primary metrics as more and more posts are shared. When agents have a HIGH tolerance, polarization grows slower than when tolerance is MEDIUM or LOW. The polarization is constantly lower when tolerance in agents is HIGH compared to

MEDIUM or LOW. Homophily grows faster when the agent's tolerance is HIGH, compared to MEDIUM or LOW, and social networks whose agents have higher tolerance end up with higher homophily after all posts are shared. The overall user satisfaction at LOW tolerance is constantly higher than HIGH or MEDIUM.

Table 1 shows the proportion of receiver (spreader and disinterested) and nonreceiver agents after all posts are shared. The number of receivers (agents who receive a post) is highest when tolerance is MEDIUM and lowest when tolerance is HIGH. The number of disinterested agents is highest when tolerance is HIGH.

Table 2 lists values for secondary metrics after all posts are shared. Secondary metrics compare the proportion of satisfied, active, and polarized users in the initial (before any posts are shared) and final (after sharing 5000 posts) populations based on thresholds defined for secondary metrics (Table 2). The number of positively satisfied users is highest when tolerance in users is HIGH and lowest when tolerance is LOW. User activity shows minor variation across different levels of tolerance. LOW tolerance leads to highest increase in highly polarized agents, whereas it is lowest when tolerance in agents is HIGH.

Takeaway (tolerance). *Higher tolerance in users slows down polarization leading to a less polarized network, higher network homophily, lower user satisfaction, and a low number of highly polarized users than when tolerance in users is lower.*

4.3.2 Experiment 2: Selective Exposure

We emulate selective exposure in our simulation by exposing each agent only to posts from other agents who have a similar stance on the issue discussed in the post. To operationalize selective exposure, we use a threshold value of the difference in the stances of two agents beyond which they stop seeing each other's posts. An agent sees posts only from other agents whose stance differs from its stance on an issue in the post below a threshold. We experiment with four threshold values for selective exposure, NONE (allow all agents to see all content shared by neighbors without any filtering, i.e., no selective exposure), LOW (allow a difference of 80% in the stance between sharing and receiving agents toward the issue in the post), MEDIUM (allow 50% difference), and HIGH (allow 20% difference). We maintain the level of tolerance among users at MEDIUM for all scenarios in this experiment.

Figure 2 compares the influence of different levels of selective exposure on all primary metrics. HIGH selective exposure leads to the highest polarization, and NONE leads to the lowest. Polarization in a social network is constantly higher for higher levels of selective exposure. Homophily is highest when selective exposure is HIGH, and shows minor variations across lowers levels of selective exposure. User satisfaction is highest when selective exposure is HIGH and shows minor differences across lower levels of selective exposure.

Table 1 shows the proportion of receiver (spreader and disinterested) and nonreceiver agents after all posts are shared. HIGH selective exposure experiences the lowest proportion of receiver agents, while NONE selective exposure leads to most.

Table 2 compares the proportion of satisfied, active and polarized users in the initial (before any posts are shared) and final (after sharing 5000 posts) populations based on thresholds defined for secondary metrics Table 2. MEDIUM selective exposure experiences the highest number of positively satisfied users, whereas the highest number of negatively

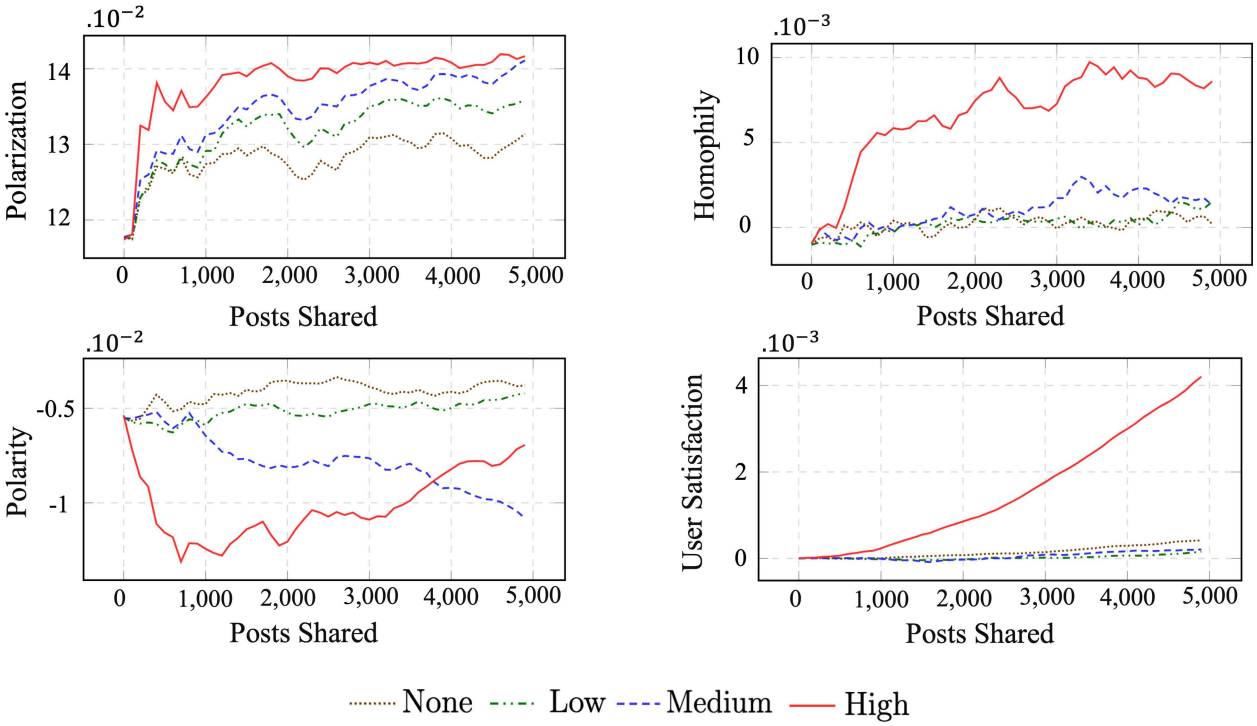

Figure 2: Experiment 2 (Selective Exposure): Comparing polarization, homophily, network polarity, and user satisfaction of agents in a social network with different levels of selective exposure.

satisfied users is with NONE selective exposure. HIGH selective exposure leads to the lowest number of negatively satisfied users. The number of highly active users experiences the most decline when selective exposure is MEDIUM, and the least when selective exposure is HIGH. HIGH selective exposure leads to the highest number of highly polarized users, whereas NONE and LOW selective exposure lead to the lowest.

Takeaway (selective exposure). *Higher selective exposure leads to higher polarization, higher network homophily, higher overall user satisfaction, and a higher number of polarized users than when selective exposure is lower.*

4.3.3 Statistical Analysis

We conduct statistical analysis to test if different levels of selective exposure and tolerance lead to statistically significant differences in users' POV (point-of-view) and primary metrics (network polarization, homophily, polarity, and user satisfaction). For users' POV we compare the final distributions (after all posts are shared) of users' POV at different levels of selective exposure and tolerance to establish if the differences are statistically significant. For primary metrics, we compare the distributions of each primary metric (computed after sharing of each post) at different levels of selective exposure and tolerance to identify differences in the overall social network metrics.

To choose the applicable statistical tests appropriately we first evaluate the distributions.

Exp	Config	Agent State		
			Receiver	
		Nonreceiver	Spreader	Disinterested
Tolerant Users	Low	60.12	14.49	25.39
	Medium	53.95	17.30	28.75
	High	62.99	13.36	23.65
Selective Exposure	None	54.76	16.88	28.36
	Low	55.44	16.48	28.08
	Medium	58.90	13.80	27.30
	High	82.63	4.97	12.40

Table 1: Distribution of agents across different states in the final population for each experimental setting. Results are from averages of ten simulation runs for each experiment. Values are in % of the total population.

Exp	Config	User Satisfaction			User Activity			User Polarity	
		Neg	Zero	Pos	Low	Medium	High	Low	High
Initial Distribution		0.00	100.00	0.00	1.56	64.82	33.62	99.33	0.67
Tolerant Users	Low	52.09	23.79	24.11	4.51	64.79	30.70	97.50	2.50
	Medium	51.92	22.26	25.82	4.51	64.52	30.97	98.19	1.81
	High	50.11	23.42	26.47	4.33	65.39	30.28	98.54	1.46
Selective Exposure	None	51.74	23.25	24.01	4.48	64.82	30.70	98.69	1.31
	Low	51.40	22.93	25.67	5.08	64.62	30.30	98.69	1.31
	Medium	45.26	26.64	28.10	7.30	64.42	28.28	97.03	2.97
	High	23.75	53.97	22.28	4.35	63.98	31.67	96.73	3.27

Table 2: Comparison between initial and final distributions of agents on secondary metrics for different experiments. Results are from averages of ten simulation runs. The values are in % of the total population.

We test the normality of distribution using the Shapiro-Wilks normality test (Shapiro and Wilk, 1965). We use parametric statistical tests, namely paired t-test and one-way ANOVA to compare distributions that are normal, and nonparametric tests, namely the Kruskal-Wallis test for distributions that are not normal.

In addition to the statistical significance test, we also compute the effect size for each test. For parametric statistical tests we use Cohen's d (Cohen, 1988) to compute the effect size as the distributions under comparison have similar standard deviations and the sample size is large (≈4,000). To interpret the effect size computed using Cohen's d, we adapt the interpretation from Cohen (1988) (see Table 3). For nonparametric statistical tests (Kruskal-Wallis test) we use epsilon square (ϵ^2) (Kelley, 1935) to compute the effect size based on recommendations from Tomczak and Tomczak (2014). To interpret the effect size computed

using epsilon square (ϵ^2) we adapt the interpretation from (Rea and Parker, 2014) for the correlation coefficient and square threshold values of each bin as ϵ^2 is a squared metric. The resulting interpretation for ϵ^2 effect size we use is as shown in Table 3. We chose ϵ^2 over other popular alternatives such as omega-squared (ω^2) (Albers and Lakens , 2018), as ϵ^2 is less biased (Okada , 2013).

For all statistical significance tests, we assume the null hypothesis to indicate similar distribution between compared entities while the alternate hypothesis to indicate that there exist statistically significant differences in the compared distributions.

We use the significance level, i.e., alpha, as 0.05 to accept or reject the null hypothesis.

We use the Kruskal-Wallis test to compare all primary metrics for different levels of selective exposure and user tolerance. For selective exposure, we compare how different levels (i.e., LOW, MEDIUM, and HIGH) compare against NONE selective exposure, whereas for user tolerance we compare each level of tolerance against each other in pairs.

Table 4 shows the results of the statistical significance test for all primary metrics at different levels of selective exposure and tolerance. The compared distributions correspond to the value of each metric after each post is shared on the social network. We are effectively comparing how the social network evolves (in terms of the metrics) as more and more posts are shared. The p-values for each pair of distribution comparing the metrics indicate that the difference in the distributions is statistically significant and the null hypothesis can be rejected, though the effect sizes vary. Based on the effect size, the difference between network homophily when selective exposure is MEDIUM and HIGH (compared to NONE selective exposure) is very strong. The difference in polarization at HIGH selective exposure (compared with NONE) and high tolerance (compared with LOW) is strong. Similarly, the difference in homophily between LOW and NONE selective exposure, and user satisfaction between HIGH and NONE selective exposure is also strong. For different levels of user tolerance, relatively strong differences exist in polarization between LOW and MEDIUM, MEDIUM and HIGH; in homophily between LOW and MEDIUM, HIGH and LOW; and in polarity between HIGH and LOW. For different levels of selective exposure, a relatively strong difference (in comparison to NONE selective exposure) exists in polarization at HIGH; in polarity at MEDIUM; and in user satisfaction at LOW. Other comparisons have an effect size of either moderate or weak.

Table 5 shows the results of the statistical significance test comparing users' POV at different levels of selective exposure and tolerance. The compared distributions correspond to the POV of each user after all posts are shared on the social network. We are effectively comparing how the POV of users differ as a consequence of different levels of selective exposure and tolerance at the start and end of each simulation run. The p-values for some of the differences show that the differences are statistically significant, though the effect sizes are either small or very small.

5 Discussion

Polarization is slowed down substantially when tolerance in users is HIGH. HIGH tolerant users experience the least network polarization and have less network polarity than when users' tolerance is LOW. The low polarization is plausibly because HIGH tolerant users are less likely to change their stance on issues based on sanctions they receive than LOW tolerant

	Effect Size	Interpretation
Epsilon-Square (ϵ^2) Interpretation based on (Rea and Parker, 2014)	[0.00, 0.01)	Negligible
	[0.01, 0.04)	Weak
	[0.04, 0.16)	Moderate
	[0.16, 0.36)	Relatively strong
	[0.36, 0.64)	Strong
	[0.64, 1.00]	Very strong
Cohen's d (Cohen, 1988)	0.20	Small
	0.50	Medium
	0.80	Large

Table 3: Effect sizes and their interpretations (according to the cited works).

users, hence, slowing down change to a users' POV. The number of highly polarized users is lowest when user tolerance is HIGH. Our results are consistent with the earlier work (Coscia and Rossi, 2022), which found lower levels of network polarization with high user tolerance in a social network.

Figure 1 shows user satisfaction when tolerance is LOW is constantly higher than when tolerance is HIGH, leading to a higher overall user satisfaction. However, the number of users with positive satisfaction is higher when tolerance is HIGH, compared to when tolerance is MEDIUM or LOW (Table 2). This indicates that the sharing of posts in a social network whose users have lower tolerance leads to higher overall user satisfaction but concentrated among fewer users.

Surprisingly, HIGH user tolerance leads to a more homophilic network (based on users' POV) than when user tolerance is LOW or MEDIUM. Also, User reach (number of users who receive a post) is lower when tolerance in users is HIGH compared to LOW and MEDIUM.

HIGH selective exposure leads to higher polarization than MEDIUM, LOW, and NONE selective exposure, in that order. This is plausibly because when selective exposure is HIGH users are more likely to see congenial posts (posts that agree with their existing stance) and are subject to fewer posts that may challenge their stance. Our finding that higher selective exposure leads to higher polarization agrees with earlier findings from prior works (Stroud, 2010; Garrett et al., 2014; Kim and Chen, 2016). However, it is important to elucidate the difference in the methodology between our work and prior works to understand the results better. While ours is a multiagent simulation that captures the evolution of polarization as caused by the social interactions between users, prior works (Stroud, 2010; Garrett et al., 2014; Kim and Chen, 2016) primarily rely on self-reported survey data for their conclusions. Further, prior works focus on how exposure to some information may polarize an individual's attitude in isolation rather than as a consequence of social interactions between multiple users.

As expected, user satisfaction is higher for higher levels of selective exposure (Figure 2). High user satisfaction may result because users receive more congenial posts with higher selective exposure, leading to more positive sanctions and higher user satisfaction for some users. The number of users with zero user satisfaction (i.e., users whose satisfaction didn't

Exp	Metric	Dist1	Dist2	H-statistic	p-value	Effect Size
Tolerant Users	Polarization	Low	Medium	2784.62	<0.01	0.27
		Medium	High	2852.45	<0.01	0.28
		High	Low	4178.42	<0.01	0.42
	Homophily	Low	Medium	1894.71	<0.01	0.19
		Medium	High	15.27	<0.01	0.00
		High	Low	2353.32	<0.01	0.24
	Polarity	Low	Medium	67.88	<0.01	0.01
		Medium	High	1516.77	<0.01	0.15
		High	Low	1981.18	<0.01	0.20
	User Satisfaction	Low	Medium	1111.50	<0.01	0.11
		Medium	High	10.60	<0.01	0.00
		High	Low	1075.30	<0.01	0.11
Selective Exposure	Polarization	None	Low	1336.62	<0.01	0.13
		None	Medium	2918.22	<0.01	0.29
		None	High	4317.15	<0.01	0.43
	Homophily	None	Low	5038.38	=0.04	0.50
		None	Medium	7316.85	<0.01	0.73
		None	High	7485.42	<0.01	0.75
	Polarity	None	Low	4.00	<0.01	0.00
		None	Medium	1813.12	<0.01	0.18
		None	High	6349.25	<0.01	0.63
	User Satisfaction	None	Low	2927.38	<0.01	0.29
		None	Medium	1232.89	<0.01	0.12
		None	High	4286.36	<0.01	0.42

Table 4: Statistical significance test results comparing primary metrics across different levels of selective exposure and user tolerance. *Dist1* and *Dist2* refer to the distributions of the corresponding primary metric for the overall social network (after all posts are shared) at the specified levels of tolerance and selective exposure as applicable based on the corresponding experiment (Exp). *H-statistic* represents the Kruskal-Wallis test statistic. *Effect size* is computed using epsilon-squared (ϵ^2).

change during the simulation run) is highest when selective exposure is HIGH and the number of negatively satisfied users is substantially lower ($\approx 2\times$) than lower levels of selective exposure. This indicates selective exposure ensures fewer users end up with aggregate negative satisfaction.

Higher selective exposure leads to the lowest user reach (i.e., highest number of nonreceivers, Table 1). This is most likely caused as a consequence of filtering out uncongenial posts for each user which leads to fewer users receiving any given post than when no selective exposure is applied. The number of disinterested is lowest in the case of HIGH selective expo-

Exp	Test	Dist1	Dist2	Test Statistic	p-value	Effect Size
Tolerant Users	Paired t-test	Low	Medium	1.35	=0.18	0.02
		Medium	High	0.72	=0.47	0.01
		High	Low	2.06	=0.04	0.03
	One-way ANOVA	Low	Medium	1.08	=0.30	0.02
		Medium	High	0.26	=0.61	0.01
		High	Low	2.41	=0.12	0.03
Selective Exposure	Paired t-test	None	Low	1.03	=0.30	0.02
		None	Medium	10.20	<0.01	0.17
		None	High	3.99	<0.01	0.07
	One-way ANOVA	None	Low	0.56	=0.45	0.02
		None	Medium	57.66	<0.01	0.17
		None	High	9.48	<0.01	0.07

Table 5: Statistical significance test results comparing a user's POV (point-of-view) in the final population (after all posts are shared) across different levels of selective exposure and user tolerance. *Dist1* and *Dist2* refer to the distributions of users' POV at the specified levels of tolerance and selective exposure as applicable based on the corresponding experiment (Exp). Effect size is computed using Cohen's d.

sure demonstrating that selective exposure makes it less likely for a post to reach potentially disinterested (i.e., users with a potentially uncongenial POV toward the post). This comes at the cost of a low number of spreaders when selective exposure is HIGH.

HIGH selective exposure witnesses the least drop in highly active users between the start and the end of the simulation. Our findings on higher selective exposure leading to more highly active users are consistent with some empirical findings from prior work. Prior work (Stroud, 2010) found selective exposure to congenial political information increases participation. At the same time, it undermines earlier work that found a positive role of cross-cutting exposure on political participation (Kim and Chen, 2016).

HIGH selective exposure leads to the highest number of highly polarized users at the end of the simulation. HIGH selective exposure also leads to a social network with the highest homophily. Homophily shows some of the highest effect sizes in the statistical significance test analysis with values indicating a very strong relation implying that the change in overall network homophily is statistically significant. The effect size is highest when selective exposure is HIGH, followed by MEDIUM, and then LOW indicating an increasing pattern of homophily with higher selective exposure.

Our findings have practical and valuable implications for social networking platforms that have become an integral part of our lives. These platforms try to maximize user satisfaction and often employ content filtering (algorithm selective exposure) to choose content based on user preference. Our simulation shows achieving user satisfaction via selective exposure can potentially increase polarization in the social network. High selectivity in exposure to congenial content may lead to better user satisfaction (due to increased likelihood of viewing

congenial posts), but it also leads to more polarized users. On the other hand, social networks whose users have a higher tolerance experience far less polarization among their users for the same number of shared posts. However, the user satisfaction when users' tolerance is higher is lower.

Interestingly, network homophily (the tendency of being connected to users with similar POV) increases in both experiments, i.e., higher selective exposure and higher tolerance in users both lead to networks with higher homophily. Social networks with higher homophily are more prone to forming echo chambers (wherein a person encounters only beliefs or opinions that coincide with their own), which is a growing challenge for social media platforms. While it is not incumbent on social networking platforms to mitigate its ill effects, such as polarization among users and the formation of echo chambers, there are some benefits to it. For instance, our simulation shows higher selective exposure leads to the lowest user reach (i.e., highest number of nonreceivers).

Our simulation model is a step toward understanding the social interactions between users in a social network and how it influences user behavior and polarization. A better understanding the potential consequences of the interactions on a social network can show us ways to mitigate the ill effects while still making the most of these social networking platforms.

5.1 Threats to Validity

Modeling user behavior is a challenging task that demands an intricate understanding of human psychology and an extensive operationalization of human traits. Though we model each user based on theories from social science and relevant observations from previous related works, the simplifications done to formalize the setup incur some threats to validity.

First, we assume equal strength of ties between each pair of connected users. In reality, people have varying strengths of ties, affecting how they react to posts from others and how it influences them.

Second, we only consider a user's own preferences and content of the post when deciding to share a post, and providing sanctions. In reality, there may be a myriad of factors that affect such decisions.

Third, the simulation runs on artificially generated data. User attributes and the posts being shared are artificially generated based on suitable probability distributions. Though we ensured appropriate distributions for initial user attributes, this does not guarantee a reasonable replication of a real-world social network. Any generalizations based on these findings need to be verified with empirical data.

Forth, the results are based on simulation runs each of which ends after sharing $\approx$ 5,000 posts. While most plots indicate the simulation stabilizing (near the end of the simulation) with the general direction of the plots being stable, there is no certainty that the same trends will continue forever.

The results should be taken with caution. Although our model is based on assumptions grounded in prior studies on polarization on social media, we use artificially generated data for this analysis. Further, reliably modeling user behavior is nontrivial and requires a fine-grained understanding of user behavior. We make simplifying assumptions in our model.

5.2 Future Directions

This work brings forth exciting directions for further research. First, it would help to develop richer simulation models that capture the dynamics of social networks, such as forming and severing ties between users and diffusing several posts simultaneously in the network. Second, it would help to seed the simulation with data collected from real users via a human-subject study. Third, it would be interesting to extend our model to incorporate methods of intervention that can help mitigate polarization in a social network.

Acknowledgments

This research was partially supported by the National Science Foundation under grant IIS-1908374 and a gift from Facebook.

References

Taufiq Ahmad, Aima Alvi, and Muhammad Ittefaq. The use of social media on political participation among university students: An analysis of survey results from rural Pakistan. *SAGE Open*, 9(3):1–9, 2019. doi: 10.1177/2158244019864484.

Casper Albers and Daniël Lakens. When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*. Elsevier. 74:187–195, 2018. doi: 10.1016/j.jesp.2017.09.004.

Brooke Auxier and Monica Anderson. Social media use in 2021. Technical report, Pew Research Center, 2021. URL `https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/`.

Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221, 2018. doi: 10.1073/pnas.1804840115.

Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, 124: 048301:1–048301:6, Jan 2020. doi: 10.1103/PhysRevLett.124.048301.

Leticia Bode, Emily K. Vraga, Porismita Borah, and Dhavan V. Shah. A new space for political behavior: Political social networking and its democratic consequences. *Journal of Computer Mediated Communication*, 19:414–429, 2014. doi: 10.1111/jcc4.12048.

Shelley Boulianne. Social media use and participation: A meta-analysis of current research. *Information, Communication & Society*, 18(5):524–538, 2015. doi: 10.1080/1369118X.2015.1008542.

Tinggui Chen, Jiawen Shi, Jianjun Yang, Guodong Cong, and Gongfa Li. Modeling public opinion polarization in group behavior by integrating SIRS-based information diffusion process. *Complexity*, 2020:4791527:1–4791527:20, 2020. doi: 10.1155/2020/4791527.

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 1988. ISBN 9780203771587. doi: 10.4324/9780203771587.

Michele Coscia and Luca Rossi. How minimizing conflicts could lead to polarization on social media: An agent-based model investigation. *PloS One*, 17(1):e0263184, 2022. doi: 10.1371/journal.pone.0263184.

Leon Festinger. *A Theory of Cognitive Dissonance*, volume 2. Stanford University Press, Stanford, CA, 1957.

Jonathan L. Freedman and David O. Sears. Selective exposure. *Advances in Experimental Social Psychology*, 2:57–97, 1965. ISSN 0065-2601. doi: 10.1016/S0065-2601(08)60103-3.

Jason Gaitonde, Jon Kleinberg, and Éva Tardos. *Polarization in Geometric Opinion Dynamics*, page 499–519. Association for Computing Machinery, New York, 2021. ISBN 9781450385541. doi: 10.1145/3465456.3467633.

R. Kelly Garrett, Shira Dvir Gvirsman, Benjamin K. Johnson, Yariv Tsfati, Rachel Neo, and Aysenur Dal. Implications of pro and counter-attitudinal information exposure for affective polarization. *Human Communication Research*, 40(3):309–332, 2014. doi: 10.1111/hcre.12028.

Alan S. Gerber, Donald P. Green, and Ron Shachar. Voting may be habit-forming: Evidence from a randomized field experiment. *American Journal of Political Science*, 47(3):540–550, 2003. doi: 10.1111/1540-5907.00038.

Eszter Hargittai, Jason Gallo, and Matthew Kane. Cross-ideological discussions among conservative and liberal bloggers. *Public Choice*, 134(1-2):67–86, 2008. doi: 10.1007/s11127-007-9201-x.

William Hart, Dolores Albarracín, Alice H. Eagly, Inge Brechan, Matthew J. Lindberg, and Lisa Merrill. Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4):555–558, 2009. doi: 10.1037/a0015701.

Jan Hązła, Yan Jin, Elchanan Mossel, and Govind Ramnarayan. A geometric model of opinion polarization. *arXiv*, 2019. doi: 10.48550/arXiv.1910.05274.

Tom Infield. Americans who get news mainly on social media are less knowledgeable and less engaged. Technical report, Pew Research Center, 2020. URL `https://www.pewtrusts.org/en/trust/archive/fall-2020/americans-who-get-news-mainly-on-social-media-are-less-knowledgeable-and-less-engaged`.

Chang Jiang, Weihua Li, Quan Bai, and Minjie Zhang. Preference aware influence maximization. In *Multi-Agent and Complex Systems*, pages 153–164, Singapore, 2017. Springer. ISBN 978-981-10-2564-8. doi: 10.1007/978-981-10-2564-8_11.

Benjamin K. Johnson, Rachel L. Neo, Marieke EM Heijnen, Lotte Smits, and Caitrina van Veen. Issues, involvement, and influence: Effects of selective exposure and sharing on polarization and participation. *Computers in Human Behavior*, 104:106155, 2020. doi: 10.1016/j.chb.2019.09.031.

Truman L. Kelley. An unbiased correlation ratio measure. *Proceedings of the National Academy of Sciences of the United States of America*, 21(9):554–559, 1935. ISSN 00278424. URL `http://www.jstor.org/stable/86523`.

David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, page 137–146, New York, 2003. Association for Computing Machinery. ISBN 1581137370. doi: 10.1145/956750.956769.

Yonghwan Kim. Does disagreement mitigate polarization? How selective exposure and disagreement affect political polarization. *Journalism & Mass Communication Quarterly*, 92(4):915–937, 2015. doi: 10.1177/1077699015596328.

Yonghwan Kim. How cross-cutting news exposure relates to candidate issue stance knowledge, political polarization, and participation: The moderating role of political sophistication. *International Journal of Public Opinion Research*, 31(4):626–648, 2019. doi: 10.1093/ijpor/edy032.

Yonghwan Kim and Hsuan-Ting Chen. Social media and online political participation: The mediating role of exposure to cross-cutting and like-minded perspectives. *Telematics and Informatics*, 33(2):320–330, 2016. ISSN 0736-5853. doi: 10.1016/j.tele.2015.08.008.

Donald R. Kinder and David O. Sears. Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology*, 40(3):414–431, 1981. doi: 10.1037/0022-3514.40.3.414.

Joseph T. Klapper. *The Effects of Mass Communication*. Free Press, Glencoe, 1960.

Silvia Knobloch-Westerwick. *Choice and Preference in Media Use: Advances in Selective Exposure Theory and Research*. Routledge, New York, 2014.

Silvia Knobloch-Westerwick and Benjamin K. Johnson. Selective exposure for better or worse: Its mediating role for online news' impact on political participation. *Journal of Computer Mediated Communication*, 19:184–196, 2014. doi: 10.1111/jcc4.12036.

Ivan V. Kozitsin and Alexander G. Chkhartishvili. Users' activity in online social networks and the formation of echo chambers. In *Proceedings of the 13th International Conference on Management of Large-Scale System Development (MLSD)*, pages 1–5. IEEE, 2020. doi: 10.1109/MLSD49919.2020.9247720.

Amanda Lenhart. The democratization of online social networks. Technical report, Pew Research Center, 2009. URL `https://www.pewresearch.org/internet/2009/10/08/the-democratization-of-online-social-networks/`.

Jure Leskovec and Julian Mcauley. Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems*, volume 25, Lake Tahoe, Nevada, 2012. Curran Associates, Inc. URL `https://proceedings.neurips.cc/paper/2012/file/7a614fd06c325499f1680b9896beedeb-Paper.pdf`.

Weihua Li, Quan Bai, and Minjie Zhang. A multi-agent system for modelling preference-based complex influence diffusion in social networks. *The Computer Journal*, 62(3):430–447, 2019. doi: 10.1093/comjnl/bxy078.

Charles Lord, Lee Ross, and Mark Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37:2098–2109, 11 1979. doi: 10.1037/0022-3514.37.11.2098.

Michael Meffert, Sungeun Chung, Amber Joiner, Leah Waks, and Jennifer Garst. The effects of negativity and motivated information processing during a political campaign. *Journal of Communication*, 56:27–51, 2006. doi: 10.1111/j.1460-2466.2006.00003.x.

Alfredo Jose Morales, Javier Borondo, Juan Carlos Losada, and Rosa M Benito. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):033114:1–9, 2015. doi: 10.1063/1.4913758.

Diana C. Mutz. The consequences of cross-cutting networks for political participation. *American Journal of Political Science*, 46(4):838–855, 2002a. ISSN 00925853, 15405907. doi: 10.2307/3088437.

Diana C. Mutz. Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review*, 96(1):111–126, 2002b. ISSN 00030554, 15375943. doi: 10.1017/S0003055402004264.

Karine Nahon. Where there is social media there is politics. In *The Routledge Companion to Social Media and Politics*, pages 39–55. Routledge, New York, 2015. ISBN 9781315716299.

Luis G. Nardin, Tina Balke-Visser, Nirav Ajmeri, Anup K. Kalia, Jaime S. Sichman, and Munindar P. Singh. Classifying sanctions and designing a conceptual sanctioning process model for socio-technical systems. *The Knowledge Engineering Review (KER)*, 31(2): 142–166, March 2016. doi: 10.1017/S0269888916000023.

Mark E. J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, 2003. doi: 10.1103/PhysRevE.67.026126.

Kensuke Okada. Is Omega Squared Less Biased? A Comparison of Three Major Effect Size Indices in One-Way ANOVA. *Behaviormetrika*. Springer. 40(2):129–147, 2013. doi: 10.2333/bhmk.40.129.

Vincent Price, Joseph N. Cappella, and Lilach Nir. Does disagreement contribute to more deliberative opinion? *Political Communication*, 19(1):95–112, 2002. doi: 10.1080/105846002317246506.

Louis M. Rea and Richard A. Parker. *Designing and Conducting Survey Research: A Comprehensive Guide*. John Wiley & Sons, 2014. URL `https://repository.vnu.edu.vn/handle/VNU_123/90042`.

David P. Redlawsk. Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *The Journal of Politics*, 64(4):1021–1044, 2002. doi: 10.1111/1468-2508.00161.

David Schkade, Cass R. Sunstein, and Reid Hastie. What happened on deliberation day? *California Law Review*, 95:915, 2007. doi: 10.2139/ssrn.911646.

S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. ISSN 00063444. doi: 10.2307/2333709.

Muzafer Sherif and Carl I. Hovland. *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*. Yale University Press, New Haven, CT, 1961.

Natalie Jomini Stroud. Media effects, selective exposure, and Fahrenheit 9/11. *Political Communication*, 24(4):415–432, 2007. doi: 10.1080/10584600701641565.

Natalie Jomini Stroud. Polarization and partisan selective exposure. *Journal of Communication*, 60(3):556–576, 08 2010. ISSN 0021-9916. doi: 10.1111/j.1460-2466.2010.01497.x.

Charles S. Taber and Milton Lodge. Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3):755–769, 2006. doi: 10.1111/j.1540-5907.2006.00214.x.

Maciej Tomczak and Ewa Tomczak. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 21 (1):19–25, 2014. URL `http://www.tss.awf.poznan.pl/files/3_Trends_Vol21_2014__no1_20.pdf`.

Nicholas A. Valentino and David O. Sears. Event-driven political communication and the preadult socialization of partisanship. *Political Behavior*, 20(2):127–154, 1998. ISSN 01909320, 15736687. URL `http://www.jstor.org/stable/586579`.

Chao Wang, Jin Ming Koh, Kang Hao Cheong, and Neng-Gang Xie. Progressive information polarization in a complex-network entropic social dynamics model. *IEEE Access*, 7:35394–35404, 2019. doi: 10.1109/ACCESS.2019.2902400.

Axel Westerwick, Benjamin K. Johnson, and Silvia Knobloch-Westerwick. Confirmation biases in selective exposure to political online information: Source bias vs. content bias. *Communication Monographs*, 84(3):343–364, 2017. doi: 10.1080/03637751.2016.1272761.

Guangchao Yuan, Pradeep K. Murukannaiah, Zhe Zhang, and Munindar P. Singh. Exploiting sentiment homophily for link prediction. In *Proceedings of the 7th ACM Conference on Recommender Systems (RecSys)*, pages 17–24, Foster City, California, October 2014. ACM. doi: 10.1145/2645710.2645734.

John R. Zaller. *The Nature and Origins of Mass Opinion*. Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press, Cambridge, 1992. doi: 10.1017/CBO9780511818691.

Fangqi Zhong, Pengpeng Li, and Jinchao Xi. A survey on online political participation, social capital, and well-being in social media users—based on the second phase of the third (2019) TCS Taiwan communication survey database. *Frontiers in Psychology*, 12: 730351–730351, 2022. ISSN 1664-1078. doi: 10.3389/fpsyg.2021.730351.

Daniel Zwillinger and Stephen Kokoska. *CRC Standard Probability and Statistics Tables and Formulae*. CRC Press, Boca Raton, 1999. ISBN 9780367802417. doi: 10.1201/9780367802417.

Appendix

Notation	Description
c_1	A constant (scale factor) to scale up smaller values. We use the value of 10.
c_2	A constant (scale factor) to scale down the larger values. We use the value of 0.1.
a_x	Agent x
p_k	k^{th} post shared on the social network
$uS(a_x, i, p_k)$	Stance of a_x toward issue i after p_k is shared
$pS(p_k, i)$	Stance of p_k toward issue i
$uA(a_x, p_k)$	Activity score for a_x after p_k is shared
$sPref(a_x, p_k)$	Sharing preference of a_x after p_k is shared
$sP(a_x, p_k)$	Probability of a_x to share p_k
$Sanc(a_y, p_k, a_x)$	Sanction a_y provides on receiving p_k from a_x
$\delta S(a_x, a_y, i, p_k)$	difference in stance between the spreader (a_x) and the receiver (a_y) on issue i as post p_k is beingshared.
$\Delta S(a_x, a_y, i, p_k)$	shift in stance (of a_x) due to a sanction (by a_y) for p_k it shared on the issue i.
$POV(a_x, p_k)$	POV of a_x after p_k has diffused in the social network
$neighbor(G, a_x, p_k)$	all neighbors of a_x in the social network G which receive p_k from a_x
$numAgents(G)$	Total number of agents in the social network G

A.1: Notations used to describe the simulation design.

Metric	Description
Negative Satisfied	Agents with user satisfaction less than zero
Zero Satisfied	Agents with user satisfaction equal to zero
Positive Satisfied	Agents with user satisfaction greater than zero
Low Activity	Agents with user activity lower than or equal to 0.75
Medium Activity	Agents with user activity between [0.75, 0.90]
High Activity	Agents with user activity greater than or equal to 0.90
Low Polarized	Agents with POV in $[-0.1, 0.1]$
High Polarized	Agents with POV greater than 0.1 or lower than -0.1

A.2: Secondary metrics to compare initial and final user distribution based on agent's state.