Effects of Affective Polarization on Ideological Polarization due to Algorithmic Filtering

Jean Springsteen and William Yeoh

Washington University in St. Louis, Saint Louis, Missouri, United States
{jmspringsteen,wyeoh}@wustl.edu

Abstract. The introduction of social media websites touted the idea of global communication, exposing users to a worldwide audience and a diverse range of experiences, opinions, and debates. Unfortunately, studies have shown that social networks have instead contributed to growing levels of polarization in society across a wide variety of issues. Social media websites employ algorithmic filtering strategies to drive engagement, which can lead to the formation of filter bubbles and increased levels of polarization. In this paper, we introduce the concept of affective polarization – the tendency for an individual to distrust someone not because of their specific opinion, but because of their political party – into an opinion dynamics model. Our results show that incorporating affective polarization into the opinion dynamics model (1) affects the level of ideological polarization present among agents in the network; (2) changes the effectiveness of algorithmic filtering strategies; and (3) is exacerbated by the presence of extremists in the network. Hence, the inclusion of an affective polarization mechanism in opinion dynamics modeling is crucial in trying to understand the effects of algorithmic filtering strategies on ideological polarization on social networks.

Keywords: Social simulation · Opinion dynamics · Algorithmic filtering · Affective polarization

1 Introduction

Social media networks have become a common place for individuals to discuss ideas, opinions, and events. While social media networks have provided a platform for globalized communication, there have also been unintended consequences, such as increased polarization. Social media has the ability to expose individuals to diverse opinions and perspectives, but individuals seemingly gravitate towards opinions and posts they already agree with.

Popular opinion dynamics models explain how interaction between two agents on a network leads to their opinions becoming more similar. The foundational DeGroot model of opinion dynamics [4] has been widely studied extended to include the concept of bounded confidence [8], the stubbornness of agents to stay committed to their initial opinions [7], multi-dimensional extensions [15, 14], and the inclusion of a network administrator that can make changes to the social network [3].

While these models provide conditions for the agents in the network to reach a consensus, they suffers from the constraint that weights of interpersonal influence must be non-negative. We address this limitation by incorporating methods from social judgement theory. Social judgement theory states the attitude change of an individual is a judgemental process, where external stimuli and influence are judged relative to an individual's own opinion [17, 16]. In social judgement theory, there are three zones within which individuals judge external influence or attitudes. If an outside opinion is close enough to an individual's own views, this opinion is "acceptable"; whereas an opinion sufficiently far from an individual's own opinion would be "unacceptable." If the perceived opinion is neither close enough or sufficiently different, it falls in a zone of noncommitment. We follow this framework in [2] and allow agents to judge their neighbors' opinions relative to their own.

Recent work has studied the effects of filter bubbles in social networks; however, one limitation of previous work [11, 3] is that these models assume that an individual can still be exposed to the opinions of *all* of its neighbors in the social network. We address this limitation by investigating how several simple algorithmic filtering strategies that decide which opinions an individual is exposed to can impact the polarization of opinions in social networks. Further, most opinion dynamics models in the literature focus on ideological polarization, leaving out important group dynamics. We incorporate affective polarization into our model to study the effect of group identity. The phenomenon of affective polarization [10, 9] refers to the increasing emotional division and polarization between individuals from different groups or political parties. Importantly, individuals are likely to think of people belonging to the other group negatively. It is becoming increasingly likely that individuals rely on negative stereotypes and do not interact with the actual opinions of people in the out-group [6]. Therefore, when opinion dynamics models do not take type, or party, into account, they fail to recognize an important mechanism in opinion formation. We aim to address this by developing a model that allows people to interact with neighbors based on the ideological distance between their opinions and the potential difference in group identity. This enables us to study how the incorporation of these mechanisms drives polarization and is impacted by the presence of filtering strategies. Our experimental results show that incorporating affective polarization into the opinion dynamics model (1) affects the level of ideological polarization present among agents in the network; (2) changes the effectiveness of algorithmic filtering strategies; and (3) is exacerbated by the presence of extremists in the network. Therefore, to understand the impact of algorithmic filtering strategies and opinion dynamics, it is crucial to incorporate an affective polarization mechanism.

2 The Model

This section introduces the model we use to study the impact of affective polarization on opinion dynamics. To incorporate both social identity theory and social judgment theory into the model, we extend the model in Tsang & Lar-

son [18] to include a trust function that adapts to the magnitude of opinion difference between two agents, using the same conventions as Chau et al[2]. Additionally, the trust function depends on whether two agents are of the same type, as individuals respond to content based on similarity of social identity.

2.1 Opinion Dynamics

We model a network where n agents are embedded in a weighted, undirected graph $G = \langle V, E \rangle$. The vertices, $V = \{v_1, ..., v_n\}$ correspond to the agents, while an edge, $e_{i,j} = (v_i, v_j) \in E$ indicates that agents v_i and v_j are neighbors on the social network. If two agents are neighbors, they are able to interact with content from the other.

Our focus is on the propagation of an opinion on an issue during a set of discrete time steps, $t \in \{1, ..., T\}$. Each agent's opinion is confined to the $[0, 1]$ interval, where 0 and 1 are referred to as "extreme" opinions, and 0.5 represents a moderate opinion. Each agent, v_i, is also assigned a type, $p_i \in \{0, 1\}$, that corresponds loosely to their opinion. Specifically, $p_i \sim \text{Bernoulli}(x_i)$. This type does not change even while opinions shift. At each time step, agent v_i has an opinion, denoted x_i^t, and it shares that opinion with its neighbors, $N_i = \{v_j \in V | (v_i, v_j) \in E\}$. An agent's opinion at time t is updated based on the weighted opinion of their neighbors in the previous time step:

$$x_i^t = \frac{w_{i,i}^{t-1} x_i^{t-1} + \sum_{v_j \in N_i} w_{i,j}^{t-1} x_j^{t-1}}{w_{i,i}^{t-1} + \sum_{v_j \in N_i} w_{i,j}^{t-1}} \tag{1}$$

where $w_{i,j}^{t-1}$ indicates the weight agent v_i places on the opinion of agent v_j at time $t-1$. This value also evolves over time, according to Equation 2:

$$w_{i,j}^t = \alpha w_{i,j}^{t-1} + (1 - \alpha) T(x_i^t, x_j^t) \tag{2}$$

where $\alpha \in [0, 1]$ is a parameter that describes how set an agent is in their own opinion, and $T(x_i^t, x_j^t)$ is function that defines the trust between two agents, based on their absolute difference in opinion. To incorporate social judgement theory [17, 16, 2], the trust function has three components, where agents behave differently according to their absolute difference in opinion, $|x_i^t - x_j^t|$. This trust function is given by:

$$T(x_i^t, x_j^t) = \begin{cases} e^{\frac{(|x_i^t - x_j^t| - d_1)^2}{-(d_1 / \ln(2))^2}} - 1 & \text{if } |x_i^t - x_j^t| < d_1 \\ 0 & \text{if } d_1 \leq |x_i^t - x_j^t| \leq d_2 \\ 1 - e^{\frac{(|x_i^t - x_j^t| - d_2)^2}{-((1-d_2) / \ln(2))^2}} & \text{if } |x_i^t - x_j^t| > d_2 \end{cases} \tag{3}$$

where d_1 and d_2 are threshold parameters ($0 \leq d_1 \leq d_2 \leq 1$). Trust values are confined to the interval $[-1, 1]$, where the highest trust value is assigned to neighbors with the exact same opinion. Therefore, agents are more likely to assign high trust values to their neighbors of the same type and similar opinion

and the lowest trust values to agents of the opposite type with a large absolute difference of opinion.

Group identity also influences how two agents interact. As Druckman et al. [6] find, individuals not only mis-estimate the ideological extremity of those in the out-type, they rely on these misconceptions when making their own decisions. The trust function in Equation 3 allows agents to judge the distance between their opinion and the opinion of another agent, but it does not allow for an agent to take into account the type of the other individual. To address this, when interacting with agents of the other type, an agent does not judge their true opinion. Instead, they use an estimated opinion. Since individuals rely on more extreme stereotypes when estimating the opinion of people in the out-type, we model this opinion as the average of the most extreme 10% of the agents of the type. This quantity can evolve over time; an agent re-estimates the ideological extremity of an agent of the opposite type at each iteration. While Equation 3 still calculates the trust value for two agents of the same type, Equation 4 models how trust changes after interacting with agents of the opposite type.

$$T_{\text{opp}}(x_i^t, \hat{x}_j^t) = \begin{cases} e^{\frac{(|x_i^t - \hat{x}_j^t| - d_1)^2}{-(d_1/\ln(2))^2}} - 1 & \text{if } |x_i^t - \hat{x}_j^t| < d_1 \\ 0 & \text{if } d_1 \leq |x_i^t - \hat{x}_j^t| \leq d_2 \\ 1 - e^{\frac{(|x_i^t - \hat{x}_j^t| - d_2)^2}{-((1-d_2)/\ln(2))^2}} & \text{if } |x_i^t - \hat{x}_j^t| > d_2 \end{cases} \tag{4}$$

2.2 Algorithmic Filtering Strategies

To test the impact algorithmic filtering strategies can have on the distribution of opinions on a network, we implement several simple filtering strategies. At each time step, the filtering strategy selects k neighbors for each agent, v_i to interact with, where k is a user-defined parameter that is the same for each agent and at each time step. We use $S_i \subseteq N_i$ to denote the subset of neighbors agent v_i interacts with, formally defined as:

$$S_i \equiv \{\hat{v_1}, \hat{v_2}, ...\hat{v_k} | \hat{v_j} \in N_i \wedge \mathcal{P}\} \tag{5}$$

where $\mathcal{P}$ corresponds to the constraints of the algorithmic filtering strategies.

Random Neighbors: As a baseline strategy, k neighbors are chosen randomly for an agent, v_i. For this baseline strategy, there are no additional constraints beyond being in the neighbor set.

$$\mathcal{P} \equiv true \tag{6}$$

Least Polar Neighbors: In trying to reduce polarization of opinions on the network, an obvious filtering strategy is to only allow individuals to interact with their least polar neighbors, where polarity is measured here as the absolute distance from 0.5. This is formalized as the constraint:

$$\mathcal{P} \equiv \forall x \in X_i \setminus \{\hat{x}_1, \hat{x}_2, ..., \hat{x}_k\} : |x - 0.5| \geq |\hat{x}_j - 0.5| \tag{7}$$

where neighboring agent v_j has opinion $\hat{x}_j$, and X_i is the set of opinions corresponding to the neighboring agents, N_i.

Most Similar Neighbors: Another intuitive approach to reducing opinion polarization on a network is to allow individuals to interact with similar agents, reducing the number of interactions with less similar agents. This leads to more interactions in the assimilation zone and fewer in the boomerang zone. More formally,

$$\mathcal{P} \equiv \forall x \in X_i \setminus \{\hat{x}_1, \hat{x_2}, ..., \hat{x}_k\} : |x - x_i| \geq |\hat{x}_j - x_i| \tag{8}$$

where $\hat{x}_j$ is the opinion of agent v_j and X_i is the set of opinions corresponding to neighboring agents $v \in N_i$.

Most Popular Neighbors: Social networks have provided a way for individuals to follow and interact with influential people or organizations, often with a higher number of followers than a typical individual. For this filtering strategy, an agent is shown their k neighbors with the highest degree. Letting $\deg(v)$ denote the degree of agent in v in the graph, we model this constraint as:

$$\mathcal{P} \equiv \forall v \in N_i \setminus \{\hat{v}_1, \hat{v}_2, ..., \hat{v}_k\} : \deg(v) \leq \deg(\hat{v}_j) \tag{9}$$

Agents of the Same Type: On social media networks, cross-party ties exist. However, as Facebook reported in 2015, their algorithms lead individuals to experience slightly less cross-cutting content [1]. The same report echoes that people are more likely to interact with and consume information with which they already agree. To drive engagement and prevent individuals from believing they are interacting with extremists of the other type, one potential strategy is to only encourage interactions between agents of the same type. If an agent did not have neighbors of the same type, they interacted randomly with their neighbors.

$$\mathcal{P} \equiv \forall v \in N_i \setminus \{\hat{v}_1, \hat{v}_2, ..., \hat{v}_k\} : p_i = p_j \tag{10}$$

Clearly, these are relatively simple filtering strategies, especially in comparison to the filtering and ranking algorithms used by actual social media companies. However, as simple as they are, their impact on ideological polarization is pronounced. Therefore, if their algorithms do not explicitly address the impact they have on polarization, it is likely they are further exacerbating it.

3 Experimental Evaluation

To determine the effects that algorithmic filtering strategies can have on ideological polarization, we run a set of simulation experiments, varying whether agents take the type of another agent into account and varying the type of agents in the network. For each set of parameters, we run 25 trials, averaging quantities

of interest over the trials. In each run, the experiment terminates when all opinions have changed by no more than a small value, $\delta = 0.001$, or the experiment reaches the maximum number of iterations, $i_{\max} = 500$, though $i_{\max}$ was rarely reached. For each experiment, we first generate a graph, G, with $n = 200$ agents (nodes), where each agent, v_i, has opinion x_i and type p_i. In each simulation, G is a Barabási-Albert random graph with homophily. This allows us to model the tendency of individuals to self-select similar neighbors on the network. Instead of changing the structure of the graph, individual agents are allowed to interact with only a subset of its neighbors.

We also vary the number of "extremists" [18] present in the network. Extreme agents have fixed polarized opinions at one extreme (0 or 1) and do not update their opinions as a result of interacting with other agents. In half of the experiments, there are no extremists present in the network, and in the other half of experiments, 10% of agents in the graph are randomly assigned to be 0-extremists and 10% are assigned to be 1-extremists. The model also depends on an individual's tolerance for opinions different than their own. In the trust functions in Equations 3 and 4, there are two threshold parameters, d_1 and d_2, corresponding to the three zones from social judgement theory [17, 16]. We vary these parameters to analyze how different assimilation and boomerang zones affect opinion dynamics. For each experiment, we use three values for d_1 and d_2: $d_1 \in \{0.3, 0.5, 0.7\}$ and $d_2 \in \{0.5, 0.7, 0.9\}$, creating 8 different combinations (since $d_1 \leq d_2$ by definition). To understand the importance of incorporating type dynamics and the presence of affective polarization into the model, we run half of the experiments with this mechanism (Equation 4 for agents of different types) and half where all agents interact the same way (Equation 3).

For each of the six filtering types, we compare four results - whether or not agents reacted to a neighbor's type or used their true opinion and whether or not the network contains extremists. The polarity results and opinion distributions are presented without the opinions of agents who were designated as extremists. We use average distance from 0.5 as a measure of polarization, following the convention in Tsang & Larson [18], but we also investigate the variance of opinions in the final opinion distribution, incorporating another measure of polarity used in the literature [3, 12, 13].

3.1 Random Neighbors

The first strategy we implement allows agents to interact with their neighbors randomly. By construction of the network, it is still more likely that an agent will interact with someone similar, both in terms of opinion and type. Without the presence of a filtering strategy, we cannot rule out consistent interaction with extremists or agents of the opposite type. Figure 1 shows the polarization on the network after the experiment has terminated, for each of the d_1 and d_2 combinations. The first panel shows the polarization of moderates on a network without extremists where agents estimate the opinion of neighbors of the opposite type. In general, the agents on these networks reached a consensus, as there was very little variance in the 200 opinions. While the average opinion over the 25 trials

Fig. 1. Average polarization of moderates after random neighbor interactions. No extremists are present in the first two models. In the first and third model, agents estimate the opinion of opposite type.

was roughly 0.5 for each combination of threshold parameters, closer inspection of the results indicates that opinions rarely settled around 0.5. Opinions more frequently converged to values 0.35 and 0.65, meaning opinion formation was largely dependent on initial graph structure.

The second panel of Figure 1 shows the average polarity from networks free of extremists where agents use the true opinion of their neighbors to form their opinions. In this case, agents always reached a consensus, and there was little variance (< 0.006) in the opinions for all trials in all combinations of threshold parameters. The average polarization reflects the fact that consensus rarely occurred at the moderate opinion, 0.5, and was more likely to occur at a slightly polarized opinion. The random interactions that did occur greatly influenced the final consensus - meaning the filtering strategies should greatly influence the final opinion distribution. The third and fourth panel of Figure 1 show the polarization of the agents in the same experiments when there are extremists present in the network. Unsurprisingly, the average polarization on the network increases as a result. There is little difference between agents interacting with the true opinion of their neighbors as opposed to an estimated opinion, likely due to the fact that agents interact with extremists over the course of the experiment.

3.2 Least Polar Neighbors

Unsurprisingly, prioritizing neighbors with the most moderate opinions results in the lowest levels of polarity of agents in the network, as shown in Figure 2. The highest levels of polarity come from the experiments with the smallest threshold parameter corresponding to acceptance zones. With smaller values of d_1, agents may still boomerang away from moderate opinions. However, agents still overwhelmingly assimilate when interacting with their moderate neighbors.

Even in the case where agents estimate the opinion of moderates to be extreme because they are of the opposite type, average polarization remains relatively low. In the third panel of Figure 2, we see the highest levels of polarization for this strategy, corresponding to a network with extremists where agents estimate the opinion of the opposite type. However, both the average distance to 0.5 and the variance in the agents' opinions (~ 0.08) are relatively small when compared to other filtering strategies. Additionally, in these experiments, originally moderate agents (the 80% who were not extremists) typically reach a consensus,

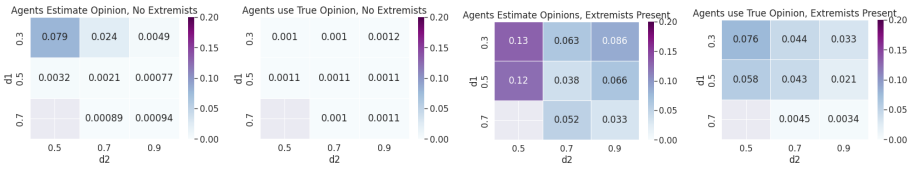

Fig. 2. Average polarization of moderates after least polar neighbor filtering strategy. No extremists are present in the first two models. In the first and third model, agents estimate the opinion of opposite type.

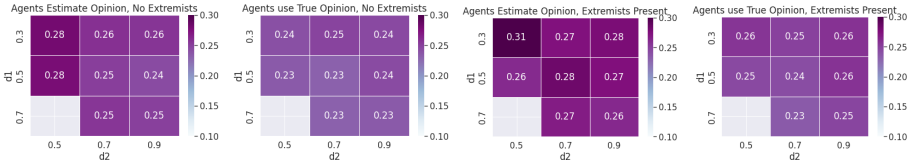

Fig. 3. Average polarization of moderates after most similar neighbor filtering strategy. No extremists are present in the first two models. In the first and third model, agents estimate the opinion of opposite type.

almost exactly at 0.5. Therefore, in our model, it is possible for agents to reach a consensus, even when over-estimating the extremism of agents of the other type, but it requires prioritizing the least polar opinions.

3.3 Most Similar Neighbors

In this filtering strategy, we expect there to be little variation in the results across threshold parameters since the filtering strategy prioritizes neighbors whose opinion is already close to that of the agent. Since agents are more likely to be neighbors with agents who have similar opinions and are thus more likely to be of the same type, we do not expect a very polarized opinion distribution. Figure 3 shows the average polarization of agents in the network after opinions converge under this filtering strategy. As expected, there is no significant difference based on threshold parameters, the presence of extremists, or whether agents are estimating the opinion of neighbors of the opposite type.

In comparing the values in Figure 3 to the levels of polarization from previous strategies, the values seem high, but not entirely extreme. Upon further investigation, the variance in these opinions was nearly always between 0.25 and 0.3, indicating that agents did not reach a consensus. Figure 4 shows an example final distribution of opinions for each of the four experiments. In some of the experiments, there are distinct peaks, indicating ideological groupings. For example, when agents use the true opinion of their neighbors in a network without extremists, we see two distinct peaks, one around 0.3 (corresponding to agents of type 0) and another, larger peak near 0.7. According to the definition of polarization in DiMaggio et al. [5], the bimodal distribution with relatively high levels

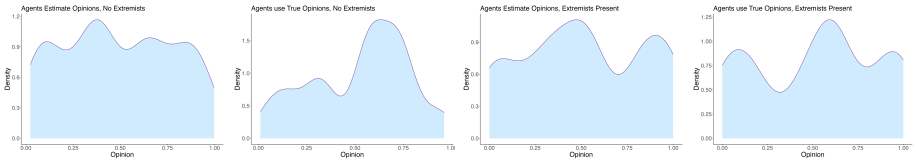

Fig. 4. Example of the distributions of final opinions after each type of experiment using the most similar neighbors filtering strategy.

Fig. 5. Average polarization of moderates after most popular neighbor filtering strategy. No extremists are present in the network in the first two models. In the first and third model, agents estimate the opinion of opposite type.

of variance indicate a polarized opinion distribution. The opinion distribution in a network with extremists where agents estimate the opinions of neighbors of the other type also results in a bimodal distribution with relatively high variance. In this case, the agents of type 1 are centered very close to the 1-extreme, while agents of type 0 have more variance in their opinions.

3.4 Most Popular Neighbors

This filtering strategy is unique because an agent will interact with the same neighbors at every iteration. Since the structure of the network does not change over the course of the experiment, an agent's most popular neighbors will not change. Therefore, we expect the presence of extremists and the threshold parameters to influence the polarity of the agents in the network. Additionally, the nodes with the highest degree in the network may be the most ideologically extreme, so the exact network structure plays a significant role in this filtering strategy. The results of this experiment are presented in Figure 5.

As expected, polarity varies based on the threshold parameters and the presence of extremists. When agents use the true opinions of their neighbors and there are no extremists in the network (panel 2 in Figure 5), there is very little polarization. In addition, the variance of opinions in these experiments is quite small (< 0.008). In this instance, agents tend to reach a consensus at the opinion of the agent with the highest degree. This can be at any value on the ideological spectrum, but over the course of 25 trials, the average highest-degree node has opinion 0.5. In general, this instance of strategy leads to consensus at the opinion of the node with the highest degree.

Fig. 6. Average polarization of moderates after agents of the same type filtering strategy. No extremists are present in the network in the first two models. In the first and third model, agents estimate the opinion of opposite type.

In the first panel of Figure 5, we see the threshold parameters influence the effectiveness of this strategy. When d_1 and d_2 are small, there is only a small zone of acceptable opinions for each agent. If the agents with the highest degree have opinions that fall outside of this zone, an agent will not assimilate towards that opinion, and since d_2 is also small, an agent is more likely to boomerang away from the opinion of the most popular agents. This results in higher levels of polarization than in experiments with larger d_1 thresholds. Additionally, we see higher levels of polarization with this strategy when agents overestimate the ideological extremity of their neighbors. This strategy is most affected by overestimation of ideology because of the repeated interaction with the same neighbors. When the most popular neighbors are of the other type, agents repeatedly interact with them, causing an overestimation of their ideological extremity.

3.5 Agents of the Same Type

Individuals seek out information they already agree with, and we investigated this as a filtering strategy above. We assumed most similar based on opinion, and did not consider type. What if we were most concerned with similarity between two agents based on type? In this case, we do not expect there to be much difference between experiments where agents estimate out-type ideology. However, we anticipate the presence of extremists to radically shift the levels of polarization. With extremists in the network, an agent is likely to interact with both in-type extremists and in-type moderates, increasing polarity on the network. This should intensify as the threshold parameter for acceptable opinions, d_1 increases, and agents become more susceptible to in-type extremist influence.

Figure 6 shows the results from this filtering strategy. As expected, the main difference in polarity is a result of the presence of extremists. However, we also see a difference in the first two panels of Figure 6, where the difference in the experiment was whether agents overestimated the ideological extremity of out-type neighbors. The difference comes from the agents who did not have neighbors of the same type. If an agent did not have in-type neighbors, they interacted randomly with neighbors of the out-type, which caused the higher levels of polarity, especially when the d_1 threshold was relatively low.

This filtering strategy was not effective in reducing polarization in the presence of extremists, as evidenced in panels 3 and 4 in Figure 6. The average

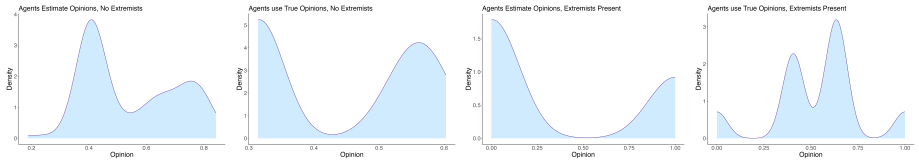

Fig. 7. Example of the distributions of final opinions after each type of experiment using the same type strategy.

distance to 0.5 was greater than 0.35 for all combinations of threshold parameters, indicating that under this strategy, most agents who began with a moderate opinion ended up at the extreme of their corresponding type. We explore this further by looking at the opinion distributions in Figure 7. In each instance of this filtering strategy, the final opinion distribution is characterized by two distinct peaks of opinions. The final opinions are extremely polarized in the case when extremists are present in the network and agents overestimate the ideological extremity of out-type neighbors. In this instance, there are no remaining moderates - every agent in the network has moved toward an ideological extreme.

These results show that including a mechanism for agents to mis-estimate the opinion of their neighbors of the other type affects the final distribution of opinions on a network. In general, when agents hold negative stereotypes about their neighbors of the opposite types, there are increased levels of polarity on the network, and this is only exacerbated by the presence of extremists on the network. While there are certainly other factors that influence the formation of an individual's opinion, we emphasize the importance of including a mechanism where an individual's *perceptions* of their neighbors influence how they interact.

4 Conclusions and Future Work

Understanding the formation of opinions on a network is a complicated question. Previous opinion dynamics models have focused on how individuals change their opinion after interacting with others, but each model relies on an individual interacting with their neighbor's actual opinion. We address this by allowing individuals to estimate the ideological extremity of neighbors in the out-type. We find that this leads to generally higher levels of polarity in the network, especially in networks with ideological extremists. Further, the effectiveness of filtering strategies is impacted by the presence of an affective polarization mechanism.

It is important to emphasize the role an affective polarization mechanism plays in opinion formation. If individuals do not interact with someone's true opinion, and they assume they are interacting with an ideological extremist, filtering strategies can contribute to increasing levels of polarity and the presence of extremists in the network only contributes to agents' negative stereotypes about the other type. Further, we cannot accurately study opinion dynamics if we do not incorporate the ways in which people actually interact. Given the evidence that individuals use negative stereotypes when interacting with people

of another type, opinion dynamics models should take this into account, and we have shown that the presence of such a mechanism may lead to higher levels of ideological polarization.

References

1. Bakshy, E., Messing, S., Adamic, L.A.: Exposure to ideologically diverse news and opinion on facebook. Science **348**(6239), 1130–1132 (2015)
2. Chau, H., Wong, C., Chow, F., Fung, C.H.F.: Social judgment theory based model on opinion formation, polarization and evolution. Physica A: Statistical Mechanics and its Applications **415**, 133–140 (2014)
3. Chitra, U., Musco, C.: Analyzing the Impact of Filter Bubbles on Social Network Polarization, p. 115–123. Association for Computing Machinery (2020)
4. DeGroot, M.H.: Reaching a consensus. Journal of the American Statistical Association **69**(345), 118–121 (1974)
5. DiMaggio, P., Evans, J., Bryson, B.: Have american's social attitudes become more polarized? American Journal of Sociology **102**(3), 690–755 (1996)
6. Druckman, J.N., Klar, S., Krupnikov, Y., Levendusky, M., Ryan, J.B.: (mis) estimating affective polarization. The Journal of Politics **84**(2), 1106–1117 (2022)
7. Friedkin, N.E., Johnsen, E.C.: Influence networks and opinion change. Advances in Group Processes **16**(1), 1–29 (1999)
8. Hegselmann, R., Krause, U.: Opinion dynamics and bounded confidence: models, analysis and simulation (2002)
9. Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., Westwood, S.J.: The origins and consequences of affective polarization in the united states. Annual review of political science **22**, 129–146 (2019)
10. Iyengar, S., Sood, G., Lelkes, Y.: Affect, not ideology: A social identity perspective on polarization. Public opinion quarterly **76**(3), 405–431 (2012)
11. Matakos, A., Aslay, C., Galbrun, E., Gionis, A.: Maximizing the diversity of exposure in a social network. IEEE Transactions on Knowledge and Data Engineering (2020)
12. Matakos, A., Terzi, E., Tsaparas, P.: Measuring and moderating opinion polarization in social networks. Data Mining and Knowledge Discovery **31**, 1480–1505 (2017)
13. Musco, C., Musco, C., Tsourakakis, C.E.: Minimizing polarization and disagreement in social networks. In: Proceedings of the International World Wide Web Conference. pp. 369–378 (2018)
14. Nedić, A., Touri, B.: Multi-dimensional hegselmann-krause dynamics. In: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC). pp. 68–73. IEEE (2012)
15. Parsegov, S.E., Proskurnikov, A.V., Tempo, R., Friedkin, N.E.: Novel multidimensional models of opinion dynamics in social networks. IEEE Transactions on Automatic Control **62**(5), 2270–2285 (2016)
16. Sherif, C.W., Sherif, M., Nebergall, R.E.: Attitude and attitude change: The social judgment-involvement approach. Saunders Philadelphia (1965)
17. Sherif, M., Hovland, C.I.: Social judgment: Assimilation and contrast effects in communication and attitude change. (1961)
18. Tsang, A., Larson, K.: Opinion dynamics of skeptical agents. In: Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems. p. 277–284 (2014)