

# Mitigating Polarization via Strategic News Exposure

Julia Smith, Katherine Henry, Pratyay Bishnupuri

## 1 Project Design

The question we plan to address throughout this project is how a recommendation algorithm can choose articles/political media such that after interacting with these new sources, a user’s opinion is as depolarized as possible. We plan to simulate these interactions by generating imaginary people and articles that will interact according to the model explained below. We are using a trust and opinion update rule from one of the relevant papers we have read, *linked*, but we have adjusted these equations to better fit our model. Each person generated in our simulation will have a randomly generated initial opinion value  $x_0 \in \mathbb{R} : x_0 \in [0, 1]$ , an opinion value of 0 indicates a very far left political stance and an opinion of 1 indicates a very far right stance. Additionally, a person has an empathy  $h$  value that will remain constant throughout the process, as we take this value to be inherent. Empathy represents a person’s willingness to be persuaded by someone/thing with a different opinion. At this point in our simulation process, we are randomly generating this value for empathy, however we plan to do further research on this topic to make empathy assignments a distribution dependent on the person’s initial opinion value. Although we have not decided how to model this feature yet, we have seen some notions in previous work indicating that people with polarized opinions have smaller empathy and people with more left wing views tend to have slightly higher empathy levels. Furthermore, people in this simulation will also have an “inertial” value  $i$  that indicates their political involvement, which we have

found based on previous work to be indicative of a person’s willingness to change opinions. We currently plan to randomly generate this number unless we find work indicating that the US population follows some sort of known distribution. These are all the attributes that people will have in our model, articles will simply have a static value indicating their political bias  $a \in \mathbb{R} : a \in [0, 1]$ .

## 2 Update Rule

Our current update rule is based on empathy and trust models current present in opinion research, *linked*. However, because our experimental setup is a bit different, the typical mathematical update rules for opinion have been modified. Our update functions assume that each person starts the experiment with a certain level of polarization,  $X_t$ , measured from 0 to 1. At each time step  $t$ , the agent is capable of reading one article that can affect their level of polarization. Our goal is to choose an article  $a_c$  that minimizes the agent’s polarization at time  $t$ . The core of our experiment is the trust function, a function that models the amount of trust a person has in any given article. We assume that the agent has an innate level of empathy  $h_x$ , which is the measure of how much they are willing to consider opinions different than their own. Using the difference in opinion between the agent and the article and the empathy value, we calculate trust using a Gaussian kernel. The Gaussian kernel ensures that when opinions are farther apart, they are significantly more alien to the agent. Thus,

at a certain time timestep  $t$ , the trust between an agent and an article can be modeled as:

$$T(a, x_t) = \exp\left(-\frac{a - x_t)^2}{h_x}\right)$$

This trust function informs our weight update. The weight between a chosen article and the agent’s polarization at time  $t$  is indicative of how much importance the agent gives to a certain article’s opinion. The weight between a chosen article and an agent at time  $t + 1$  is given as:

$$w_{a_c, x_{t+1}} \leftarrow w_{a_c, x_t} + T(a_c, x_t)$$

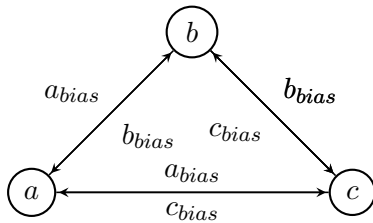
The actual update rule for an agent’s opinion at a time  $t + 1$  is dependent not only on the weight, but on the inertial factor  $i$ . Recent research by the *PewResearchCenter* suggests that voters who are more politically involved are more secure in their party definition, and thus will not change their opinion as easily as the politically inactive. Thus, as the level of political involvement increases, the inertial factor—the ease with which someone will change their opinion—also rises.

$$x_{t+1} \leftarrow \frac{(x * x_t) + (x_t * w_{a_c, x_t})}{(i + w_{a_c, x_t})}$$

### 3 Possible Factors

Though we are fairly confident that our chosen metrics are important in a person’s level of polarization, additional literature review suggests that other realistic factors could be added to our experiment. We will be exploring the significance of these factors in the coming weeks: Comment Effect, Level of Engagement, Level of Media Literacy

### 4 Implementation




---

#### Algorithm 1 simulatePath(paths, person)

---

```

startingStance  $\leftarrow$  person.getStance()
finalStances  $\leftarrow$  []
for path in paths do
    currentStance  $\leftarrow$  startingStance
    for article in path do
        currentStance  $\leftarrow$ 
            updateStance(article.getBias(),
                currentStance, person)
    end for
    finalStances.append({currentStance, path})
end for
return minStance(finalStances)

```

---

To simulate the process of an individual reading a series of articles, we will be using a complete graph. The use of a complete graph implies the idea that a person can read any other article after reading an article. So in the example above, when a person reads  $a$ , they can then either read article  $b$  or  $c$ . The complete graph is bidirectional, so when taking the path from article  $a$  to article  $b$ , the weight on the edge is article  $a$ ’s bias and when going from article  $b$  to article  $a$ , the weight on the edge is article  $b$ ’s bias. The weight on the edge is representative of the influence the article will have on the person where the level of influence is dependent on the update function defined above. The goal using this graph is to find a path such that the person’s polarization will be mitigated.

Given a set of possible paths a person can take, we can find the final political stance of the person using this algorithm below. For each path in paths, there will be a set of articles  $A = \{ \text{articles in path} \}$  that the person will read. We will have the person read each article  $a_i \in A$ , where  $i \dots A.length$ , sequentially in the path and update the person’s political stance using the update function. Then after calculating the final stance for every path, we will find the most moderate stance and its respective path of articles  $A$  which is what we will return.

You can find our initial implementation on our *GitHub*.