



Prevalence and Propagation of Fake News

Banafsheh Behzad, Bhavana Bheem, Daniela Elizondo & Susan Martonosi

To cite this article: Banafsheh Behzad, Bhavana Bheem, Daniela Elizondo & Susan Martonosi (2023) Prevalence and Propagation of Fake News, *Statistics and Public Policy*, 10:1, 2190368, DOI: [10.1080/2330443X.2023.2190368](https://doi.org/10.1080/2330443X.2023.2190368)

To link to this article: <https://doi.org/10.1080/2330443X.2023.2190368>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC



[View supplementary material](#)



Published online: 19 Apr 2023.



[Submit your article to this journal](#)



Article views: 2221



[View related articles](#)



[View Crossmark data](#)



Citing articles: 3 [View citing articles](#)

Prevalence and Propagation of Fake News

Banafsheh Behzad^a, Bhavana Bheem^b, Daniela Elizondo^c, and Susan Martonosi^b

^aInformation Systems Department, California State University, Long Beach, Long Beach, CA; ^bMathematics Department, Harvey Mudd College, Claremont, CA; ^cMathematics Department, Rutgers University, Piscataway, NJ

ABSTRACT

In recent years, scholars have raised concerns on the effects that unreliable news, or “fake news,” has on our political sphere, and our democracy as a whole. For example, the propagation of fake news on social media is widely believed to have influenced the outcome of national elections, including the 2016 U.S. Presidential Election, and the 2020 COVID-19 pandemic. What drives the propagation of fake news on an individual level, and which interventions could effectively reduce the propagation rate? Our model disentangles bias from truthfulness of an article and examines the relationship between these two parameters and a reader's own beliefs. Using the model, we create policy recommendations for both social media platforms and individual social media users to reduce the spread of untruthful or highly biased news. We recommend that platforms sponsor unbiased truthful news, focus fact-checking efforts on mild to moderately biased news, recommend friend suggestions across the political spectrum, and provide users with reports about the political alignment of their feed. We recommend that individual social media users fact check news that strongly aligns with their political belief and read articles of opposing political bias. Supplementary materials for this article are available online.

ARTICLE HISTORY

Received May 2022
Accepted March 2023

KEYWORDS

Bias; Policy recommendation;
Social media analysis;
Truthfulness

1. Introduction

In recent years, social media networks such as Facebook and Twitter have become major news sources; an estimated 62% of Americans claim to get news from social media (Gottfried and Shearer 2017). Concurrently, social media has seen a surge of *fake news* which, for the purpose of this article, is news containing a high frequency of falsehoods. In March and April 2021, fake news claims circulating through social media included:

“Per the CDC There Are Nearly Twice As Many Vaccine Related Deaths SO FAR in 2021 (1,755) Than All the Vaccine Deaths this Past Decade (994)”-March 2021, (Hoft 2021a), and “Stanford Study Results: Face masks are Ineffective to Block Transmission of COVID-19 and Actually Can Cause Health Deterioration and Premature Death.”

-April 2021 (Hoft 2021b)



Both of these claims were proven to be false, and their only purpose is to misinform people. Considering the high ramifications of just one of these claims propagating, mitigating the spread of untruthful news is of paramount importance (Vicario et al. 2016; Lazer, Baum, and Benkler 2018; Grinberg et al. 2019).


Social media has the unique characteristic of its users playing a decisive role in which content gets propagated. Alarming, research shows that users are less likely to fact check the news they receive on social media than the news they receive through other sources (Allcott and Gentzkow 2017; Jun, Meng, and Johar

2017). Furthermore, fake news takes advantage of the political echo chambers within social media to further its propagation (Vicario et al. 2016). With active users continuously making decisions on whether to propagate news (e.g., share, retweet), unverified and unreliable sources can potentially reach as large an audience on social media as national news sources (Allcott and Gentzkow 2017).

Techniques for mitigating the spread of fake news include increasing the spread of truthful news and flagging untruthful news articles or accounts (Farajtabar et al. 2017; Lazer, Baum, and Benkler 2018). The former technique increases the amount of truthful news in a social media user's feed, and the latter diminishes the credibility of fake news sources. One major source of fake news propagation are social bots, computer generated and controlled social media accounts. Since social bots target particular social media users whose usage characteristics make them more likely to further propagate fake news (Shu, Bernard, and Liu 2018; Sharma et al. 2019), one tactic for fake news mitigation is having more methods of detecting the difference between bot accounts and human accounts in order to deactivate bot accounts (Chu et al. 2012; des Mesnards and Zaman 2018; Inuwa-Dutse, Liptrott, and Korkontzelos 2018). Although current techniques are beneficial, social media users were still being exposed to fake news during crucial times like the 2016 U.S. presidential election and the 2020 COVID-19 pandemic (Allcott and Gentzkow 2017; Mikkelsen 2020).

The initial step in finding techniques to mitigate fake news is understanding how fake news spreads on both the macroscopic

CONTACT Banafsheh Behzad  banafsheh.behzad@csulb.edu  Information Systems Department, California State University, Long Beach, Long Beach, CA.

 Supplementary materials for this article are available. Please go to www.tandfonline.com/uspp.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

and microscopic levels. The macroscopic spread refers to the network level propagation of content. One study uses a network graph model to analyze how people believe and spread news; they conclude that social media cultivates echo chambers in which low quality news is likely shared (Brooks and Porter 2019). Vicario et al. (2016) find that cascade dynamics of fake news depend on polarized echo chambers. Since early detection is vital to diminishing the effects of fake news, Louni and Subbalakshmi focus on detecting the source of news within a cluster (Louni and Subbalakshmi 2014). Other studies effectively detect fake news on a network level by using propagation paths to classify different types of news (Liu and Wu 2018; Pierri, Piccardi, and Ceri 2019). Vosoughi, Roy, and Aral (2018) graph data from Twitter and conclude that fake news spreads faster and more broadly than true news when analyzing a social media network. Jang et al. (2018) use evolutionary tree analysis to similarly conclude that fake news spreads deeper through social media networks and undergoes frequent modifications. Research on the macroscopic spread of content provides an understanding of which clusters facilitate fake news spread and also of news propagation patterns. Findings from both Vosoughi, Roy, and Aral (2018) and Jang et al. (2018) inform techniques for early detection of fake news, which is crucial for mitigation.

Research on content spread at the microscopic level focuses on an individual user's decision to further propagate information from their social media feeds. Findings at the microscopic level on fake news propagation do not necessarily align with findings at a macroscopic level since an individual level decision to share news may not take into consideration the network effects. Some microscopic level studies use user surveys to understand user behavior, and other studies create models to pinpoint social media attributes that correlate to a higher likelihood of propagating fake news. Guess, Nagler, and Tucker (2019) use Poisson regressions to link a representative online survey to the respondents' sharing history on Facebook. They find that fake news is more likely to propagate through conservatives and older users. Allcott and Gentzkow (2017) similarly conduct surveys on a portion of the voting population and find that people with higher education, who spend more time on social media, and have no partisan attachment are the most accurate in distinguishing fake from real news (Allcott and Gentzkow 2017). Furthermore, Papanastasiou (2017) uses an agent-based sequential model to conclude that social media users tend to correlate popularity with credibility. Lee et al. (2014) successfully create a feature-based prediction model and a time estimation model to characterize the likelihood an account retweets a particular message based on certain traits of the Twitter account, such as its sociability and follower base. Characterizing the probability that a social media user will further spread untruthful or polarizing news based on individual factors, like political ideology, is vital in understanding how to stop the spread. It is noteworthy to mention that occasionally fake news is accompanied by flawed analysis and not necessarily intentional untruthfulness. As an example, a motion filed by Texas to the U.S. Supreme Court in 2020, consists of a claim supported by statistical analysis that it is unlikely that Biden has more votes than Trump in the four battleground states Pennsylvania, Georgia, Michigan, and Wisconsin. Miao, Pan, and

Gastwirth (2022) discuss why Texas's claim is logically flawed and against the principles of statistics.

This article further expands the research within the microscopic level. We develop a model for the probability an individual social media user with a certain political belief chooses to share a news article having a certain political bias and truthfulness. Different from previous work, our model considers the truthfulness of an article separately from the article bias, defining both characteristics on separate continuous numerical spectra. We allow for an article to be unbiased yet still untruthful, or biased while still largely truthful. Our model estimates the probability that a particular population will share a certain article. Numerical probabilities provide a more accurate understanding of how different social media populations share content relative to each other which helps inform recommendations for platforms and individual users to reduce the spread of untruthful or highly biased news.

From the analysis of the data, we find that if social media users engage only with users of similar political beliefs, then the rate of sharing news, fake or true, increases. If users engage with users of opposing political beliefs, then the rate of sharing decreases, and the decrease is more pronounced for fake news. This article recommends social media platforms: sponsor unbiased truthful news; focus existing fact-checking and flagging algorithms on moderately liberal- and conservative-biased news; suggest content and connections from across the political spectrum; and supply social media users with a score reflecting the political alignment of their feed. For individual users, this article recommends fact-checking news that aligns strongly with their belief, as people are more susceptible to propagating this type of news, and to engage with users with opposing political beliefs.

The methodology of this article follows a three step approach: developing a probabilistic model for the propagation of political news; analyzing the model to determine which attributes of content and user population contribute to the propagation of untruthful or biased content; finally, using the findings to recommend interventions against fake news. In the next section, we explain our model and outline the assumptions that are built into it. Section 3 analyzes the model from an optimization perspective to determine characteristics of content that a malicious agent seeking to propagate untruthful or highly biased news might target. Section 4 details the dataset used to validate our modeling assumptions. Section 5 presents an empirical analysis of the propagation rate of political news within populations of readers that have varying political beliefs. The analyses of Sections 3 and 5 inform the recommended techniques for mitigating the spread of fake news discussed in Section 6. Lastly, we provide our ideas for future work and conclusions in Sections 7 and 8, respectively.

2. Probability Model of Propagation

This section describes how we model the probability p that a reader propagates an encountered article (e.g., “share,” “retweet,” or “like,” depending on the social media platform of interest) as a function of the article's bias and truthfulness, and the reader's belief.

An article i is characterized by two attributes: bias b_i and truthfulness t_i . Bias is a continuous value on $[-1, 1]$ reflecting the political ideology represented by the topics, facts, and claims included in the article. A bias equal to -1 corresponds to a very liberal (“left-biased”) article, a value of $+1$ corresponds to a very conservative (“right-biased”) article, and 0 corresponds to a politically unbiased article. The truthfulness t_i represents both the frequency and severity of falsehoods in the article, ranging from 0 (completely false news) to 1 (perfectly true). The article’s reader j is characterized by their political belief B_j which also ranges from -1 to $+1$ in the same sense as b_i .

The shape of the function is based on two assumptions, informed by previous work:

1. A person will be more likely to propagate an article when its bias is close to their belief, even when the article exhibits low truthfulness (Iyengar and Hahn 2009; Bakshy, Messing, and Adamic 2015; Grinberg et al. 2019; Manickam et al. 2019). That is, $p(b_i, t_i, B_j)$ should increase as $|b_i - B_j|$ decreases. Users often share content as a means of political persuasion, and we assume that a user would not wish to disseminate content that could persuade others to adopt a dramatically different political belief.
2. The probability of propagation increases with article truthfulness, t_i . Articles that are too untruthful are unlikely to be propagated because they are too outlandish (Grinberg et al. 2019). In other words, we assume that given two articles exhibiting the same bias, a reader will be more likely to share the more truthful article.¹ However, we also assume diminishing marginal increases in sharing rate as t_i increases, due to sharing probabilities being bounded.

Additionally, we require a valid probability function, bounded between 0 and 1 for $(b_i, t_i, B_j) \in [-1, 1] \times [0, 1] \times [-1, 1]$. A probability function that satisfies these assumptions and is reasonably tractable is the logistic curve:

$$p(b_i, t_i, B_j) = \begin{cases} \frac{f_l}{1 + e^{-k_l(t_i - (b_i - B_j)^2)}} & \text{if } B_j < 0 \\ \frac{f_r}{1 + e^{-k_r(t_i - (b_i - B_j)^2)}} & \text{if } B_j \geq 0 \end{cases} \quad (1)$$

where f_l, f_r, k_l , and k_r are scaling parameters. f_l and f_r scale the overall rate of sharing for left-political and right-political readers, respectively. k_l and k_r scale the rate of decay in sharing probability associated with the squared deviation between article bias and reader belief and the truthfulness of the article. Several plots of level curves of the probability function are shown in [supplementary material \(SM\) Section 1](#).

Due to its tractability and suitability in satisfying our assumptions, we will use the model given by (1) throughout the remainder of the article.

3. Characterizing Optimal Behavior of a Malicious Agent

We first examine the model from a mathematical perspective to characterize the optimal behavior of a malicious agent wishing

to spread “fake” news. We assume that a malicious agent seeks to propagate untruthful (low t_i) or highly biased (high $|b_i|$) news at the highest possible rate, assuming a single reader having belief B_j (in [Section 3.1](#)), or a population of readers having a belief distribution (in [Section 3.2](#)). Specifically, we answer three questions:

1. What are the general characteristics of articles and populations that achieve the highest propagation rate?
2. For untruthful articles that have low t_i , what degree of bias achieves the highest propagation rate?
3. For highly biased articles that have high $|b_i|$, what degree of truthfulness achieves the highest propagation rate?

By answering these questions, we can understand how a malicious agent might select or create political content to maximize propagation by readers of a given belief. Content having these characteristics could be prioritized for fact-checking or other interventions.

Although not an assumption integrated into the model, for the purposes of optimization, this section assumes that more biased news is inherently less truthful (Xiang and Sarvary 2007). Thus, we constrain the feasible region of valid (b_i, t_i) pairs to the region $|b_i| + t_i \leq 1$. This assumption is not only reasonable but also necessary to make the optimization nontrivial; otherwise, the optimal strategy for maximizing propagation would always be to produce perfectly truthful articles, which contradicts what we observe on social media.

3.1. Single Reader

We first address each of the three questions above in the context of a malicious agent seeking to propagate content to a single reader having belief B_j .

What are the general characteristics of articles that achieve the highest propagation rate?

The probability function (1) has no local maximum, indicating that the global maximum will lie along the boundaries of the feasible region defined by $|b_i| + t_i \leq 1$, $-1 \leq b_i \leq 1$, and $0 \leq t_i \leq 1$. For the constraint $t_i = 1 - |b_i|$, the function’s maximum is a piecewise function split at $|B_j| = 0.5$. For extreme conservative beliefs ($B_j \geq 0.5$), the function’s maximum is at $b_i = -\frac{1}{2} + B_j$ and $t_i = \frac{3}{2} - B_j$. As the function is symmetric, for extreme liberal beliefs ($B_j \leq -0.5$), the function’s maximum is at $b_i = \frac{1}{2} + B_j$ and $t_i = \frac{3}{2} + B_j$. When reader belief is more moderate, that is, $|B_j| < 0.5$, the maximum value of the function occurs at maximum truthfulness: $b_i = 0$ and $t_i = 1$. Although the probability of sharing increases as $|b_i - B_j|$ decreases, the constraint $|b_i| + t_i \leq 1$ causes the optimal article bias to be substantially lower in magnitude than the reader’s belief. On the constraint boundary $b_i = 0$, the maximum propagation rate is attained when $t_i = 1$. Unbiased news achieves maximum propagation when it is perfectly truthful. However, along the constraint $t_i = 0$, the maximum is at $b_i = B_j$. Untruthful news achieves maximum propagation when its bias matches the reader’s belief.

Using our analysis, we provide policy recommendations, which come in two forms: for social media platforms and for

¹Readers may not always be able to distinguish truth from untruth with high precision. However, we are modeling the *likelihood* of sharing an article. External factors such as the reputability of the source can serve as signals to readers as to the truthfulness of content.

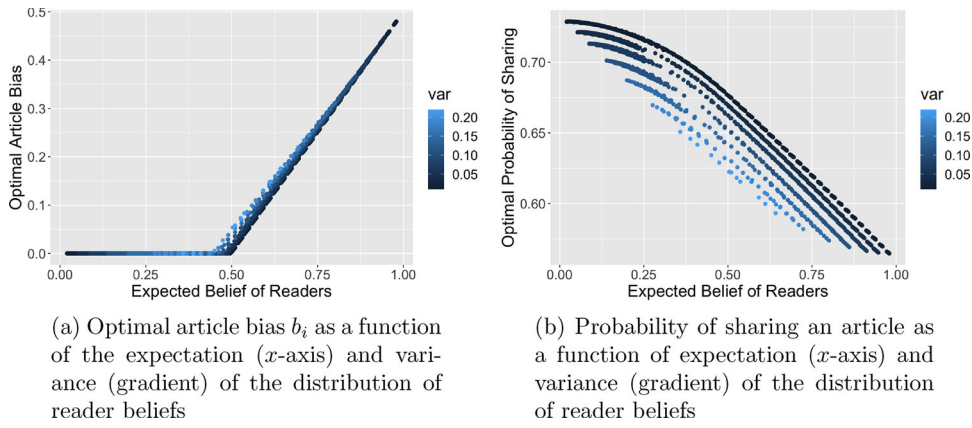


Figure 1. Optimal article bias and probability of sharing over all (b_i, t_i) pairs, for numerous discrete probability distributions of reader belief having given expected value and variance.

individual users.² From the above analysis, we provide the first of our recommendations for social media platforms (P) and users (U).

Recommendation P1: Populate users' feeds with unbiased truthful news as sponsored articles. An unbiased and highly truthful article attains a high propagation rate among many populations of readers, with the highest propagation rate among users with moderately skewed political beliefs. By sponsoring unbiased, truthful content, users will have more opportunities to engage with, and potentially share, legitimate content.

For untruthful articles that have low t_i , what degree of bias achieves the highest propagation rate?

We saw above that when $t_i = 0$, maximum propagation is achieved when $b_i = B_j$. This holds more generally for fixed t_i , provided that $t_i + |B_j| \leq 1$. When $t_i + |B_j| > 1$, there is no local maximum along the cross sectional probability curve. Thus, maximum propagation is attained at the endpoint $|b_i| = 1 - t_i$ (with the sign of b_i chosen to match the sign of B_j). To propagate a low-truth article to a single user, the malicious agent will try to match the article's bias to the user's belief, unless the user's belief is so extreme that doing so would violate the truth-bias constraint. In this latter case, the optimal bias of the article will be chosen along the constraint boundary (i.e., as extreme as possible). From this, we infer our next recommendation.

Recommendation U1: Fact check news that aligns strongly with a user's beliefs by using a third party fact-checking website for individual facts within the article. Users are more susceptible to untruthful news that aligns with their beliefs. A malicious agent acting optimally would align false content with a user's political beliefs, so users are susceptible to sharing untruthful news aligning with their beliefs. Thus, users should be proactive with the news that they read, and fact-check articles of this nature. If the platform does not suggest related content, users can also seek multiple news sources to corroborate the story.

For highly biased articles that have high $|b_i|$, what degree of truthfulness achieves the highest propagation rate?

For any fixed reader belief and article bias, the probability of sharing an article increases with article truthfulness, by assumption. Thus in our constrained optimization framework, the malicious agent wishing to propagate a highly biased article would select article truthfulness along the constraint boundary: $t_i = 1 - |b_i|$.

3.2. Distribution of Readers' Beliefs

Expanding on the case of a single reader, we examine the maximal propagation rate over a distribution of readers' beliefs. To start, given a discrete distribution of reader belief $f(B_j)$, the probability of a reader sharing an article having bias b_i and truthfulness t_i is given by:

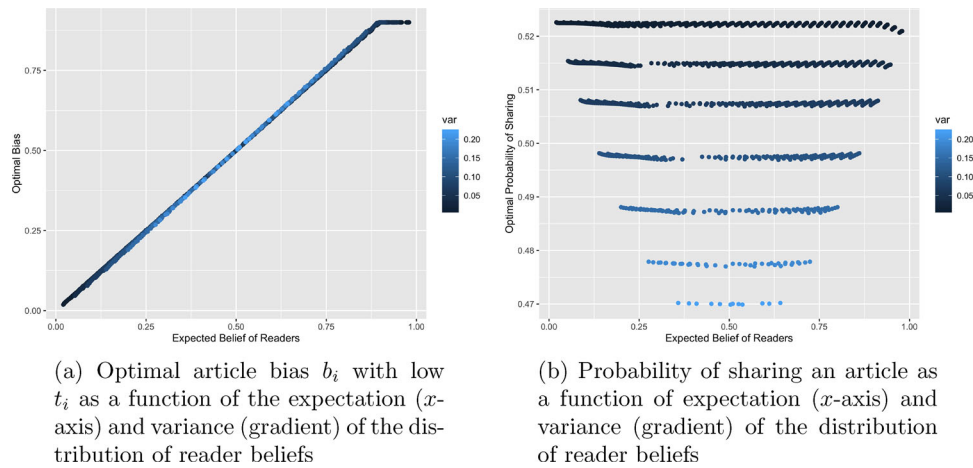
$$p(b_i, t_i) = \sum_{B_j} p(b_i, t_i, B_j) * f(B_j). \quad (2)$$

With the probability function described in (2), we solve numerically for the values of b_i and t_i that attain maximum propagation over systematically varied distributions of user belief having a variety of expectations and variances. For each distribution, we find the combination of article bias and truthfulness that maximizes the probability of propagation to answer each of the three questions posed earlier. We use the BARON solver (Tawarmalani and Sahinidis 2005; Sahinidis 2017) on the NEOS server (Gropp and Moré 1997; Czyzyk, Mesnier, and Moré 1998; Dolan 2001) to obtain the results.

What are the general characteristics of articles that achieve the highest propagation rate?

Plots of numerical solutions, shown in Figure 1, display the effects of the expected value and variance of the distribution of readers' beliefs on both optimal bias and the optimal probability of sharing. The propagation-maximizing article parameter values for the population of readers resemble those of the case of a single reader. However, increasing the variance of reader beliefs decreases the overall sharing probability and modestly increases the bias required to attain maximum propagation.

²A summary of all recommendations made in this article can be found in Table 5 in Section 6



(a) Optimal article bias b_i with low t_i as a function of the expectation (x -axis) and variance (gradient) of the distribution of reader beliefs

(b) Probability of sharing an article as a function of expectation (x -axis) and variance (gradient) of the distribution of reader beliefs

Figure 2. Optimal article bias and probability of sharing over all b_i values when $t_i = 0.1$, for numerous discrete probability distributions of reader belief having given expected value and variance.

This allows us to make our second recommendation for social media platforms:

*Recommendation P2: Recommend friends and connections across the political spectrum. Social media platforms generally recommend connections, content and advertising based on similarities. The opposite tendency, however, may help reduce the sharing rate of biased, untruthful news by increasing the variation in political belief of a user population. From the analysis, a population that has readers with opposing beliefs or moderate beliefs will exhibit a lower probability of sharing untruthful news. When suggesting new contacts to a user (e.g., a friend of a friend), platforms should suggest contacts that are across the political spectrum, as this will give users a chance to expose themselves to different political ideologies. This recommendation is counter to the current practice of many social media platforms to rank recommended connections based on *similarity* with the user.*

For untruthful articles that have low t_i , what degree of bias achieves the highest propagation rate?

Figure 2(a) displays the optimal article bias, when $t_i = 0.1$ as a function of the expected reader belief, shaded by the distribution variance; Figure 2(b) does the same for optimal probability of sharing. We see that when truth is fixed, the optimal bias equals $E[B_j]$ until the constraint boundary is reached, at which point the optimal bias equals $1 - t_i$, resembling the results of the single user case. The optimal bias is unaffected by variance. As in Figure 1(b), we see that the optimal sharing rate decreases as the variance of readers' beliefs increases. Once again, we conclude that recommending contacts across the political spectrum could reduce the rate at which untruthful content is shared (Recommendation P2).

For highly biased articles that have high $|b_i|$, what degree of truthfulness achieves the highest propagation rate?

As in the single reader case, sharing probability, for any value of $|b_i|$, increases with t_i , independent of reader beliefs. Thus, for highly biased content, the degree of truthfulness achieving the highest propagation rate lies along the constraint boundary, $t_i = 1 - |b_i|$.

Having explored the mathematical properties of our model in the context of a malicious agent seeking to propagate untruthful or polarizing content, we now introduce the dataset we use to validate the assumptions underlying the model.

4. Model Validation

We validate the assumptions of our model using social media data from Grinberg et al. (2019) and Media Bias Monitor (MBM) (Ribeiro et al. 2018) to estimate sharing probabilities and the three model parameters: bias, truthfulness and reader belief. We then compare the proposed probability function to the propagation rates of articles found on social media and verify that the model assumptions are satisfied.

4.1. Data

To obtain a set of readers along with their political beliefs, we use the data from Grinberg et al. which contains files with information about *panel members*, a set of Twitter users whose Twitter profiles can be uniquely linked to public voting records. There are 572 panel members, who are categorized by political affiliation: extreme left, left, lean left, center, lean right, right, or extreme right. In the original dataset, the categories lean left and left were combined to form the category left, and the same was done for the right side; the groups are not merged together in this article. We assign a belief value B_j to each panelist equal to the average political affiliation of reader j 's assigned belief group, which can be found in the last column of Table 4.

In the raw dataset, the panel members are heterogeneous in their social media behaviors, with some panelists categorized as *supersharers* (sharing more than 922 political articles), *superconsumers* (having more than 45,128 exposures to political articles), *bots* (nonhuman automated users responsible for mass distribution of content), and *apolitical* (having fewer than 100 exposures to political articles). As in Grinberg et al., to mitigate against outliers and simplify our analysis, we filter out these users, leaving a more homogeneous population of panel members for our analysis. A comparison of model parameters between typical and atypical users is given in SM Section 2.

The data from Grinberg et al. also holds a collection of tweets containing URLs and the political news sources of the URLs. The data has 3,141,106 tweets that the panel members were potentially exposed to, meaning the tweet was shared by an account the user follows. These tweets came from 106,838 distinct URLs: the mean number of exposures per URL was 29.4; the median number of exposures per URL was 3; and the maximum number of exposures per URL was 13,416. Thus, while there is some dependency in the exposure data, we assume it is modest given that the majority of articles had very few exposures among panel members, and these panel members were not necessarily connected to each other within the social network.

The URLs from these tweets are given unique identifiers making the actual link anonymous. Thus, it is impossible to attribute b_i and t_i values to individual URLs shared by the panel member. Instead, we use the recorded domain names of the news website from which the URL originates (Grinberg et al. 2019). We make the assumption that every article on a particular news website has the same bias and truthfulness value as the entire domain. As with user political beliefs, we categorize the article bias as extreme left (i.e., very liberal, having values of b_i close to -1), left, lean left, center, lean right, right, or extreme right (i.e., very conservative, having values of b_i close to $+1$).

To our knowledge there is no established research that calculates a numeric value for the content-based semantic bias for a wide range of domains. Therefore, we use other ways of estimating bias of an article. We estimate article bias using MBM, a database that has over 20,000 news websites (Ribeiro et al. 2018). Ribeiro et al. develop MBM as a novel scalable way of calculating domain bias using audience demographics (Ribeiro et al. 2018). They find a correlation coefficient of 87%, with a 95% confidence interval of [65%, 96%], between their bias estimates and those of Budak et al. who use machine learning and crowdsourcing techniques to determine the bias of articles from fifteen domains (Budak, Goel, and Rao 2016). The strong correlation makes MBM's bias estimates a sufficient proxy for article bias, b_i ³.

In order to estimate a numerical truthfulness score for each URL, we expand on color codes provided by Grinberg et al. that assess the overall truthfulness of the domain (Grinberg et al. 2019). Grinberg et al. provide justifications for each domain's assigned coding which we use to develop a numerical conversion for each color. Domains that are coded as black are given a truthfulness value of zero. All other domains are assigned a value of t_i from a range of values. These conversions are shown in Table 1. Note that a truthfulness rating above green refers to domains with rigorous editorial processes, such as academic journals. Since the dataset did not have such domains, we assign no domain a value of t_i in the interval [0.8, 1]. By assigning truthfulness ratings to each justification from Grinberg et al., we are able to average the domain's justification ratings to assign a t_i value to each URL. Lastly, for the purpose of the panel plot given in Figure 3, we redistribute the t_i values into the truthfulness categories of very low, low, mixed, high, and very high.

Table 1. Conversion of colors provided in Grinberg et al. (2019) into a numerical range compatible with the truthfulness scale of the model.

Truthfulness color	Grinberg et al. Description (Grinberg et al. 2019)	Numeric range	Figure 3 category
Black	"Sites that published almost exclusively fabricated stories"	0.0–0.1	Very Low
Red	Sites that "spread falsehoods that clearly reflected a flawed editorial process"	0.1–0.2	Low
Orange	Sites "where annotators were less certain that the falsehoods stemmed from a systematically flawed process"	0.2–0.3	Low
Yellow	"Mild or rare inaccuracies"	0.3–0.6	Mixed (0.3–0.5) High (0.5–0.6)
Green	"Factual and sourced reporting"	0.6–0.8	High (0.6–0.7) Very High (0.7–0.8)

NOTE: The numerical scale was further categorized into truthfulness bins for the purposes of Figure 3.

Lastly, we estimate the probability of propagation by dividing the number of people in each political affiliation group who shared URLs from a given website by the number of people in that group who were exposed to URLs from that website. The data from Grinberg et al. categorizes tweets as being a retweet, quote, or original posting. On Twitter, retweets and quotes function as shares. So, in calculating the probability, we consider only retweets and quotes. The panel members shared 1566 URL links from 37 distinct websites.

4.2. Validating Model Assumptions

We use the dataset described above to validate the assumptions outlined in Section 2. Our purpose in this section is not to make claims about the predictive power of the model but to demonstrate with data that the assumptions guiding our choice of model in (1) are reasonable.

The first assumption is that the probability of sharing increases as the bias of an article becomes more aligned with reader belief. Figure 3 plots propagation rate as a function of reader belief, across combinations of article truthfulness and bias. Each data point presented in Figure 3 is a website and a political belief bucket. Each point includes all articles that were shared from that particular website. Panels in the top two rows (corresponding to liberal article bias) show a decreasing trend in sharing rate as reader belief increases (becomes more conservative), and panels in the bottom three rows (corresponding to conservative article bias) show an increasing trend as reader belief becomes more conservative. A linear least squares regression of sharing rate as a function of $|b_i - B_j|$ reveals that the slope coefficient, β , is statistically negative for both liberal ($\beta = -0.0050, p = 0.00002, n = 148$) and conservative readers ($\beta = -0.0071, p = 0.00001, n = 146$), further supporting this assumption. Moreover, readers are less likely to share an article having a bias that opposes their political beliefs. Using two-sample t -tests, we find that the mean sharing rate of liberals sharing liberal news (LL) is statistically higher than the mean sharing rate of liberals sharing conservative news (LC): $\bar{x}_{LL} = 0.0074, \bar{x}_{LC} = 0.0016, p = 0.00054$, pooled degrees

³The Grinberg et al. data includes *alignment* scores, the weighted average of panelists exposed to a website that are registered with either the Democratic or Republican party, for only 245 distinct websites.

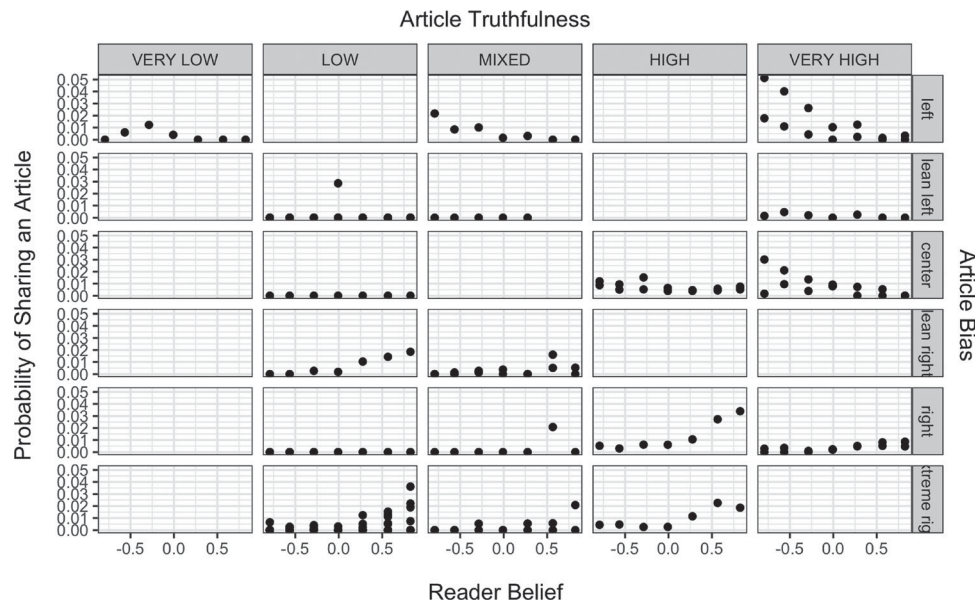


Figure 3. Probability of sharing an article depending on the political belief of a reader (x -axis), truthfulness of an article (columns), and bias of an article (rows) (Grinberg et al. 2019; Ribeiro et al. 2018).

of freedom ≈ 50 . Likewise, we find that the mean sharing rate of conservatives sharing conservative news (CC) appears to be higher than the mean sharing rate of conservatives sharing liberal news (CL): $\bar{x}_{CC} = 0.0049$, $\bar{x}_{CL} = 0.0024$, $p = 0.00856$, pooled degrees of freedom ≈ 129 .

The model's second assumption has two components, the first being that higher truthfulness increases the probability of sharing, and the second being that the rate of increase with increasing truthfulness itself decreases (diminishing marginal increases in sharing rate). Figure 3 confirms that articles in the “High” and “Very High” truthfulness columns have higher probabilities of sharing than in the “Low” and “Very Low” truthfulness columns. This is also supported by a statistical test of the slope coefficient β using a simple linear regression of sharing rate as a function of t_i . This coefficient is statistically positive among liberal readers ($\beta = 0.015$, $p = 0.00000$, $n = 148$) and weakly so for conservative readers ($\beta = 0.0051$, $p = 0.02535$, $n = 146$). We cannot validate the assumption of diminishing marginal increases in sharing rate as truthfulness increases.

We acknowledge that imprecision in the estimation of article bias (b_i), article truth (t_i) and user belief (B_j) from the dataset introduces additional uncertainty, weakening the conclusions one can draw from the reported p -values. Nonetheless, we have shown that the data generally support the assumptions underlying the functional form of the model given in (1). Having validated most assumptions of the model on real social media data, we can now analyze it empirically.

5. Empirical Analysis

In this section, we examine qualitative trends in model-estimated sharing probabilities over simulated populations of users as a function of the bias and truthfulness of the article being propagated. We do not use the model for statistical prediction of propagation rates. Instead, we use it to understand

general characteristics of the population and the content being disseminated that result in larger or smaller propagation rates. We investigate the types of populations over which the propagation of untruthful news will be highest, and for a given population, the types of news that the model suggests will be widely shared. Understanding which content is more likely to be shared, and the characteristics of populations that are more susceptible to the sharing of untruthful or highly biased news, provides insights into how users and platforms could prioritize their interventions. Using these qualitative results, we are able to make several recommendations for social media users and platforms.

5.1. Estimating Model Parameters

To examine trends in model-estimated propagation rates of political news over a population of users, we first estimate reasonable values for the scaling parameters f_l , k_l , f_r , and k_r . We fit a nonlinear least squares regression model to the observed values of (b_i, t_i, B_j) in our dataset. Table 2 includes the nonlinear least squares regression estimates, their p -values, standard errors, and the overall residual standard errors. The estimated coefficients are used in the remainder of this section and comprise the “Base Scenario” given in Table 3. To account for uncertainty in these parameter estimates, the column labeled “Sensitivity Analysis” provides low and high values of each parameter used to test the sensitivity of our recommendations to the parameters. (The recommendations resulting from our model analysis are robust to changes in these parameter values, as seen in the sensitivity analysis results given in SM Section 3.)

5.2. Empirical and Synthetic Population Distributions

Using these parameter estimates, we compute the model-estimated probability of sharing an article for all possible combinations of b_i and t_i , and over a distribution of readers'

Table 2. The fitted parameters found using a nonlinear least squares regression.

	f_l	k_l	f_r	k_r
Estimate	0.010	4.465	0.007	5.581
p-value	1.12e-12***	0.036	1.25e-09***	0.11
Standard Error	0.001	2.108	0.001	3.475
Residual Standard Error	0.007		0.007	
Degrees of Freedom	146		144	

NOTE: Significance value: 0.0001 '***', 0.001 '**', 0.005 '*'.

Table 3. Estimated parameter values used for model analysis.

Parameter	Base scenario	Sensitivity analysis
f_l	0.010	[0.005, 0.014]
k_l	4.465	[2.232, 6.697]
f_r	0.007	[0.004, 0.011]
k_r	5.581	[2.791, 8.372]

NOTE: The Base Case values are used in the analysis of Section 5.3. The sensitivity analysis is found in SM Section 3.

beliefs, using (2). The probability distribution function $f(B_j)$ is the fraction of the population in the belief group having average political belief B_j . The purpose of this analysis is not to make exact predictions about sharing probabilities, but rather to understand general relative trends of content sharing within different populations of readers.

We use the distribution found in the data from Grinberg et al., henceforth referred to as the *empirical distribution*, and five more constructed distributions: two bimodal distributions (representing partisan and hyperpartisan populations) and three unimodal distributions (representing left-unimodal, centrist, and right-unimodal populations). Table 4 gives the proportions of each distribution type having each political belief.

The five constructed distributions of readers represent different types of social media populations. The two bimodal distributions reflect populations with partisan and hyperpartisan beliefs. The three unimodal distributions represent populations concentrated on the left side of the political spectrum, in the center of the spectrum, or on the right side of the spectrum. Comparing the trends in the model-estimated propagation rate over bimodal and unimodal distributions of readers' political beliefs will help us understand the effects of partisanship on the overall propagation rate of different types of political content.

Table 4. Distribution of readers' political belief for the empirical and synthetic distributions.

	Empirical distribution	Partisan distribution (bimodal on the left and right)	Hyperpartisan distribution (bimodal on the extreme left and extreme right)	Left-unimodal distribution	Centrist-unimodal distribution	Right-unimodal distribution	Belief center point
Extreme left	0.092	0.080	0.400	0.200	0.020	0.020	-0.857
Left	0.230	0.400	0.080	0.400	0.080	0.040	-0.571
Lean left	0.225	0.020	0.020	0.200	0.200	0.060	-0.286
Center	0.184	0	0	0.080	0.400	0.080	0
Lean right	0.131	0.020	0.020	0.060	0.200	0.200	0.286
Right	0.091	0.400	0.080	0.040	0.080	0.400	0.571
Extreme right	0.046	0.080	0.400	0.020	0.020	0.200	0.857
Expectation	-0.146	0	0	-0.400	0	0.400	
Variance	0.214	0.382	0.643	0.167	0.114	0.167	

5.3. Results

In the sections that follow, we present population-wide model-estimated sharing rates of political content as a function of the content's bias and truthfulness, for each of the six population distributions described above. Figure 4 presents the probability of article propagation as a function of bias for right-biased articles over the six distributions. We focus on the propagation of right-biased articles, as the analysis for left-biased articles is symmetric. Moreover, the results and recommendations that follow are robust to variability in the estimated parameter values, as shown in the sensitivity analysis of SM Section 3.

We examine the same three questions as were posed previously, now focused on both content and population characteristics: (a) What are the general characteristics of articles and populations that achieve the highest propagation rate? (b) For untruthful articles with low t_i what characteristics yield the highest propagation rate? (c) For highly biased articles with high $|b_i|$, what characteristics yield the highest propagation rate?

5.3.1. What are the General Characteristics of Articles and Populations that Achieve the Highest Propagation Rate?

To determine the conditions under which the highest overall propagation rate is achieved, we find the location of the maximum of each graph. As seen most clearly in Figure 4, in all distributions except the right-unimodal distribution (Figure 4(d)), overall propagation rate is maximized for unbiased, high-truth content. The centrist-unimodal population (Figure 4(e)) exhibits the highest sharing rate of this type of content. In the right-unimodal population, the sharing rate is highest for *mildly* biased, moderately truthful content. Thus, we conclude that unbiased, truthful content can achieve high propagation rates in most population types. This is consistent with Recommendation P1 (*Populate users' feeds with unbiased truthful news as sponsored articles.*). However, political skew in the population of readers promotes the propagation of biased and less truthful news.

We can compare the centrist population to the partisan and hyperpartisan populations to understand how the variability of beliefs in a population impacts propagation rate even as the average belief is held constant at 0. Contrasting Figures 4(b), (c), and (e), we see that partisan and hyperpartisan populations share biased news at the highest rate of the distributions studied, while a centrist population shares highly truthful news at a higher

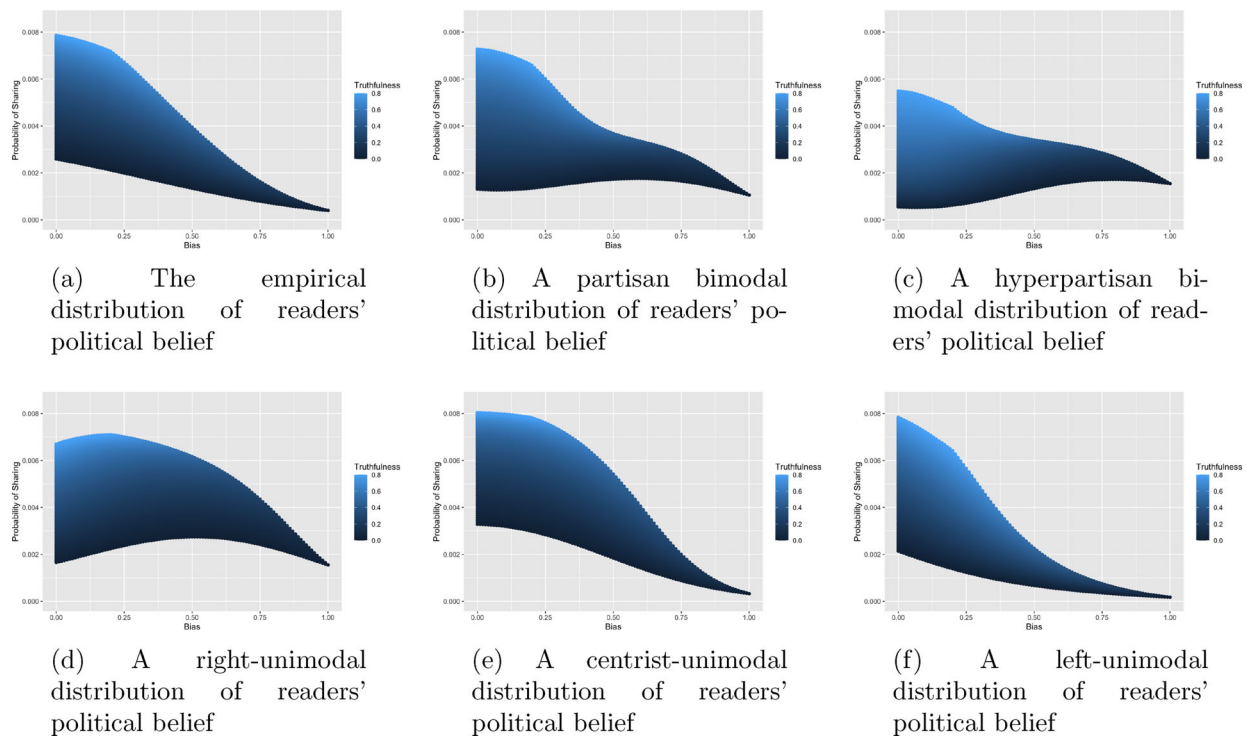


Figure 4. The probability of sharing an article as a function of right political bias (x -axis) and truthfulness (gradient).

rate than partisan and hyperpartisan populations. This leads to the conclusion that partisanism incentivizes the propagation of biased, untruthful news, even when the mean belief of the population is centered at zero.

Based on this analysis we can conclude that if users engage only with users of similar political beliefs, then the rate of sharing news, fake or true, increases. If users engage with users of opposing political beliefs, then the rate of sharing decreases, and the decrease is more pronounced for fake news.⁴ This suggests a tradeoff between mitigating the spread of untruthful news and increasing the flow of information. On one hand, dissemination and flow of information is considered to be positive (Rosling 2018), and based on the aforementioned result, it increases when users engage only with users of similar political beliefs. On the other hand, spread of fake news is harmful (Giachanou and Rosso 2020) and this as well increases when users engage only with users of similar political beliefs. Thus in the extreme, in an echo chamber the level of information flow might be high but it might be predominated by misinformation. This result lends additional support for Recommendation P2 (*Recommend friends and connections across the political spectrum.*)

5.3.2. For Untruthful Articles, What Characteristics Yield the Highest Propagation Rate?

To determine the conditions under which the propagation of *untruthful* (low t_i) content is highest, we can examine the dark regions of Figure 4 forming the bottom curve of each plot, and identify where the maximum occurs along this curve.

We see that for the empirical (Figure 4(a)), centrist-unimodal (Figure 4(e)), and left-unimodal (Figure 4(f)) distributions, the maximum of the bottom curve occurs at the left side of the

graph: propagation of untruthful right-leaning content is highest in these populations when the content is unbiased. Thus, a unimodal population with centrist beliefs will prioritize truthful unbiased news, but such a population is also susceptible to propagating untruthful unbiased news. This suggests that fact-checking efforts should not focus only on extremely biased content; the greatest propagation of untruthful content in these populations will occur with unbiased content.

However, for the partisan bimodal (Figure 4(b)) and right-unimodal (Figure 4(d)) distributions, propagation of untruthful content is highest for moderately biased content ($|b_i| \approx 0.5$). And for the hyperpartisan bimodal population (Figure 4(c)), propagation of untruthful content is highest for strongly biased content ($|b_i| \approx 0.75$). The probability of sharing any partisan news, true or fake, increases when the expected belief is similarly partisan. Thus, a malicious agent trying to maximize the probability of sharing an untruthful article in a partisan population would choose to propagate biased content, with the optimal bias increasing with the partisanism of the population.

Recommendation P3: Prioritize the fact-checking of moderately biased news, based on the user population. In unimodal populations, untruthful content is most likely to be shared if it exhibits little to no bias. However, in partisan bimodal populations, untruthful content is most likely to be shared if it exhibits moderate bias. Strongly biased untruthful content is highly likely to be shared only in hyperpartisan populations. Algorithms should fact-check articles of mild or moderate bias, as these are the articles, when untruthful, readers are more susceptible to. Only in hyperpartisan bimodal populations do untruthful articles exhibiting strong bias achieve maximum propagation.

⁴This result can be supported mathematically, as shown in SM Section 4.

It is worth noting that this suggestion aligns with research on persuasion. According to Johansen and Joslyn (2008), in environments with one-sided flow of information, significant persuasive effects would be expected. However, it is expected that the level of persuasion decreases in environments with users across the political spectrum and more specifically among users with opposing beliefs. In such environments, an extremely biased article is unlikely to influence the user's views, because its contents are improbable under the user's prior, whereas mildly biased news is significantly more likely to persuade the user. In other words, mildly biased news might update user's belief, frictionlessly (Bettinger et al. 2020). With this recommendation, our model contributes to the studies which analyze the evolution of beliefs/opinions in social media (see Nordio et al. (2017) among all).

5.3.3. For Highly Biased Articles with High $|b_i|$, What Characteristics Yield the Highest Propagation Rate?

The probability of sharing *strongly-right-biased* content in populations having the empirical, centrist-unimodal and left-unimodal distributions is low overall. Moreover, there is a general decreasing trend in propagation rate of strongly-right-biased content as we look across the tails of the graphs from right-unimodal (Figure 4(d)), to centrist-unimodal (Figure 4(e)), to left-unimodal populations (Figure 4(f)). The hyperpartisan population (Figure 4(c)) has a higher propagation rate of extremely biased, low truthfulness news, and a lower overall propagation rate of unbiased news of all truthfulness levels. This further supports Recommendation P2 above (*Recommend friends and connections across the political spectrum*) to reduce extreme bimodality in the population. This also supports three additional recommendations:

Recommendation P4: Suggest related articles that present an issue from a diversity of viewpoints. When a user accesses an article on their social media feed, the platform can recommend related, fact-checked, articles from multiple news sources having a variety of political alignments. In addition to permitting a user to assess the verity of the content when presented multiple ways, untruthful content will be less likely to be shared. It is noteworthy this suggestion runs counter to the common practice of many social media platforms of showing users content that is relevant to them (Agarwal 2016). This represents a tradeoff a platform might have to make between profitability and social good.

Recommendation P5: Give users a report about the alignment of their feed. To reduce the polarity of a population, platforms can provide each user with a score measuring the alignment of their feed. This will help users to be more cognizant of whether they expose themselves to different types of news and allow them to broaden their networks and content beyond those that only work to confirm their beliefs.

Recommendation U2: Read fact-checked articles with opposing biases to mitigate the effects of confirmation bias. Once presented with content from across the political spectrum and given a metric on the political alignment of their own media use, users should consciously engage with content that runs counter to their beliefs. When users read articles that oppose their beliefs, the probability of sharing an untruthful news article is relatively low, and they prioritize sharing truthful news.

From this analysis it is clear that the choice of content propagated by a malicious agent will be influenced by the population it is targeting. Even when the average population belief is centered at zero, bimodal partisanism incentivizes the propagation of biased, untruthful news, whereas a centrist population is more likely to share truthful news. This suggests that interventions to reduce bimodality in the social media user population could reduce the propagation of biased and untruthful content. Moreover, fact-checking mildly or moderately biased news could have more impact than fact-checking strongly biased news, as users are more susceptible to sharing untruthful news when it exhibits mild-to-moderate bias.

6. Discussion

The spread of fake news is intricate and occurs at both the individual and network level. Our work is novel in that it characterizes the effect of a news article's bias and truthfulness on its likelihood of being propagated by individual users. While we do not consider how political content travels through a network of social media users, we have characterized how a population of individuals interacts with news once received.

Table 5 summarizes the platform and user recommendations. We can consider the platform recommendations in two groups. The first group seeks to change/decrease the supply of biased news; P1, P2, and P4 are in this group. With these interventions in place, the platform minimizes the access of the users to highly biased news. The second group, consisting of P3 and P5, is comprised of informational interventions that provide

Table 5. Platform and user recommendations derived from the analysis of this article, with references to sections of this article's analysis that generate each.

Platform Recommendation	Description	Paper Section(s)
P1	Populate users' feeds with unbiased truthful news as sponsored articles.	3.1, 3.2, 5.3.1
P2	Recommend friends and connections across the political spectrum.	3.2, 5.3.1, 5.3.3
P3	Prioritize the fact-checking of moderately biased news, based on the user population.	5.3.2
P4	Suggest related articles that present an issue from a diversity of viewpoints.	5.3.3
P5	Give users a report about the alignment of their feed	5.3.3
User Recommendation	Description	Paper Section(s)
U1	Fact check news that aligns strongly with a user's beliefs.	3.1, 3.2
U2	Read fact-checked articles with opposing biases to mitigate the effects of confirmation bias.	5.3.3

users with information intended to modify their behavior, as suggested in the user recommendations U1 and U2. In many domains, informational interventions have been shown to result in substantial changes in behavior of individuals (Bettinger et al. 2020).

Our work supports the conclusion that both unimodal skew or bimodal partisanship in the reader population contributes to the spread of biased, untruthful news. Some platforms already have ways to mitigate the spread of fake news. Facebook has fact-checkers that review content and rate whether it is false or true (Facebook 2020). Twitter has recently implemented a feature that flags tweets that could be misconstruing information or be misleading (Roth and Pickles 2020). These are important steps to mitigate spread of fake news, and our model provides a mechanism for prioritizing this fact-checking effort. Additionally, while fact-checking content that has already been shared is a retroactive endeavor, several of our recommendations above serve to more proactively prevent the sharing of fake news.

7. Limitations and Future Directions

The model developed in this article is a novel way of characterizing the spread of political content on social media. Unlike previous models, it disentangles the effects of article truthfulness and bias on the sharing rate. Whereas previous work treats bias and truthfulness as categorical or binary values, our model allows for more granularity as the parameters are continuous and vary individually. This aspect of our model enables us to examine each parameter's respective relationship to the spread of fake news in more detail.

The model and similar work are limited by the lack of available data sources for measuring bias and truthfulness of news articles. Currently, work in the literature on calculating content-based semantic bias is not scalable, limiting the size of the dataset used in this article. Although alignment has a strong correlation with bias, future papers could continue the process of finding credible and scalable ways to calculate content-based semantic bias.

Currently, the model describes the probability an individual user will propagate a certain article based on three parameters: bias of the article, truthfulness of the article and political belief of the reader. The model is based on accurate assumptions, as validated with data by Grinberg et al. and MBM (Ribeiro et al. 2018); however, political news propagation is based on more than just these three variables. For example, the popularity of an article has a direct correlation to the probability that the article will be shared (Papanastasiou 2017). Additionally, the model makes an assumption that a reader's belief stays static even as they are consuming news. The model also does not differentiate between potentially different motivations behind sharing content, and assumes all sharing reflects support of the content.

Lastly, the current model does not address how fake news spreads in a network, nor how sharing behavior might change in response to changes in content. Network effects govern the types of content an individual is likely to be exposed to, while our analysis examines the conditional sharing rate assuming that an exposure to certain content has occurred. Moreover, the recommendations we make consider short-term impacts

of interventions. If the bias distribution of articles to which a reader is exposed changes as a result of interventions, sharing behavior might change over time in response. Future research could use our model to examine these temporal dynamics in a social network.

8. Conclusion

This article provides a probabilistic model that describes the likelihood of an individual social media user sharing an online news article based on the article's bias and truthfulness, as well as their own political belief. Assumptions underlying the model are validated using data from Grinberg et al. (2019) and Ribeiro et al. (2018). We are able to examine the immediate potential impact of interventions to reduce the likelihood of untruthful news being shared, by modeling the decision of an individual user to share (or not) political content.

We characterize a malicious agent as someone who wants to spread untruthful and/or highly biased news, and we use the model to determine the agent's choice of bias and truthfulness that maximizes the probability of propagation. Understanding the content and population characteristics that a malicious agent would target permits us to prioritize among intervention strategies, such as which content to fact-check. Additionally, we examine trends in model-estimated propagation rates in six different populations to characterize the conditions under which "fake" news is most likely to propagate.

Our analysis suggests that social media platforms should promote unbiased, truthful news. We find that users are more susceptible to sharing untruthful content when it is mildly or moderately biased, so platforms should prioritize allocating fact-checking resources to this type of content. Social media platforms should suggest contacts and fact-checked content that reflect differing political opinions, so that users can avoid politically homogeneous clusters. Additionally, platforms should provide users with information about the political alignment of their feed to encourage users to seek out articles whose bias opposes their own belief. Individuals should also attempt to fact check news that aligns strongly with their beliefs.

Two of these recommendations run counter to the business model of most social media platforms: P2 (*Recommend friends and connections across the political spectrum*) and P4 (*Suggest related articles that present an issue from a diversity of viewpoints*). Platforms such as Facebook, Twitter, and Instagram make content and contact recommendations based on *similarity* with the user; in doing so, they provide content that is perceived to increase the user's enjoyment and use of the platform. However, our recommendation that platforms should suggest content and contacts that oppose a user's belief presents a tradeoff that a platform must make in order to mitigate fake news spread.

In summary, our model characterizes an individual user's decision of whether or not to share political news on social media. Using this model, we present policy recommendations for social media platforms and users. Mitigating the spread of fake news has become increasingly urgent worldwide, as we've seen in recent elections and the global Covid-19 pandemic. Our

model provides a useful tool for identifying realistic actions that platforms and users can take to mitigate the spread of fake news.

Supplementary Materials

The supplementary materials include (1) level curves of the probability model; (2) comparison of model parameters between typical and atypical users; (3) sensitivity of population results to estimated model parameters; and (4) proof that population diversity reduces sharing rates.

Acknowledgments

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank Michael Gao and Steven Witkin for their contributions to earlier versions of this work, and Deyana Marsh for her substantial contributions to the computational testing. Additionally, the authors thank David Lazer and Nir Grinberg for providing access to the dataset from (Grinberg et al. 2019). Finally, the authors thank the anonymous reviewers for detailed feedback that greatly improved this manuscript.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

This material is based upon work supported by the National Science Foundation under grant no. DMS-1757952. The authors also acknowledge financial support from Harvey Mudd College and the California State University at Long Beach.

References

- Agarwal, A. (2016), "What do Social Media Algorithms Mean for You?" available at <https://www.forbes.com/sites/insights-capitalone/2021/03/10/growth-now-4-opportunities-for-cxos-to-build-their-companies-in-the-coming-year/?sh=18d5f8d05558>, accessed on 04 May 2021. [10]
- Allcott, H., and Gentzkow, M. (2017), "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, 31, 211–236. DOI:10.3386/w23089. [1,2]
- Bakshy, E., Messing, S., Adamic, L. A. (2015), "Exposure to Ideologically Diverse News and Opinion on Facebook," *Science*, 348, 1130–1132. DOI:10.1126/science.aaa1160. [3]
- Bettinger, E., Cunha, N., Lichand, G., and Madeira, R. (2020), "Are the Effects of Informational Interventions Driven by Salience?" ECON - Working Papers 350, Department of Economics - University of Zurich, available at <https://ideas.repec.org/p/zur/econwp/350.html>. [10,11]
- Brooks, H. Z., and Porter, M. A. (2019), "A Model for the Influence of Media on the Ideology of Content in Online Social Networks," *ArXiv abs/1904.09238*. [2]
- Budak, C., Goel, S., and Rao, J. M. (2016), "Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis," *Public Opinion Quarterly*, 80, 250–271. [6]
- Chu, Z., Gianvecchio, S., Wang, H., and Jajodia, S. (2012), "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" *IEEE Transactions on Dependable and Secure Computing*, 9, 811–824. DOI:10.1109/TDSC.2012.75. [1]
- Czyzyk, J., Mesnier, M. P., and Moré, J. J. (1998), "The NEOS Server," *IEEE Journal on Computational Science and Engineering*, 5, 68–75. [4]
- des Mesnards, N. G., and Zaman, T. (2018), "Detecting Influence Campaigns in Social Networks using the Ising Model," *CoRR abs/1805.10244*, available at <http://arxiv.org/abs/1805.10244>. [1]
- Dolan, E. D. (2001), "The NEOS Server 4.0 Administrative Guide," Technical Memorandum ANL/MCS-TM-250, Mathematics and Computer Science Division, Argonne National Laboratory. [4]
- Facebook. (2020), "What Publishers Should Know About Third Party Fact Checking on Facebook," available at <https://www.facebook.com/business/help/182222309230722>, accessed on 29 July 2020. [11]
- Farajtabar, M., Yang, J., Ye, X., and Xu, H., Trivedi, R., Khalil, E., Li, S., Song, L., and Zha, H. (2017), "Fake News Mitigation via Point Process based Intervention," in *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of Proceedings of Machine Learning Research, eds. D. Precup and Y. W. Teh, pp. 1097–1106, Sydney: International Convention Centre, PMLR. Available at <http://proceedings.mlr.press/v70/farajtabar17a.html>. [1]
- Giachanou, A., and Rosso, P. (2020), *The Battle Against Online Harmful Information: The Cases of Fake News and Hate Speech*, New York: Association for Computing Machinery. DOI:10.1145/3340531.3412169. [9]
- Gottfried, J., and Shearer, E. (2017), "News Use Across Social Media Platforms 2016," *Pew Research Center's Journalism Project*. Available at <https://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/>. [1]
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. (2019), "Fake News on Twitter during the 2016 U.S. Presidential Election," *Science*, 363, 374. DOI:10.1126/science.aau2706. [1,3,5,6,7,11,12]
- Gropp, W., Moré, J. J. (1997), "Optimization Environments and the NEOS Server," in *Approximation Theory and Optimization*, eds. M. D. Buhman and A. Iserles, pp. 167–182, Cambridge: Cambridge University Press. [4]
- Guess, A., Nagler, J., and Tucker, J. (2019), "Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook," *Science Advances*, 5, eaau4586. DOI:10.1126/sciadv.aau4586. [2]
- Hoft, J. (2021a), "Per the CDC There Are Nearly Twice as Many Vaccine Related Deaths So Far in 2021 (1,755) Than All the Vaccine Deaths This Past Decade (994)," available at <https://archive.is/mqXNN#selection-595.11-595.150>, accessed on 04 May 2021. [1]
- (2021b), "Stanford Study Results: Facemasks are Ineffective to Block Transmission of Covid-19 and Actually Can Cause Health Deterioration and Premature Death," available at <https://archive.is/LmMT#selection-603.0-603.147>, accessed on 04 May 2021. [1]
- Inuwa-Dutse, I., Liptrott, M., and Korkontzelos, I. (2018), "Detection of Spam-Posting Accounts on Twitter," *Neurocomputing*, 315, 496–511. DOI:10.1016/j.neucom.2018.07.044 [1]
- Iyengar, S., and Hahn, K. S. (2009), "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use," *Journal of Communication*, 59, 19–39. DOI:10.1111/j.1460-2466.2008.01402.x [3]
- Jang, S. M., Geng, T., Li, J. Y. Q., Xia, R., Huang, C. T., Kim, H., and Tang, J. (2018), "A Computational Approach for Examining the Roots and Spreading Patterns of Fake News: Evolution Tree Analysis," *Computers in Human Behavior*, 84, 103–113. DOI:10.1016/j.chb.2018.02.032 [2]
- Johansen, M. S., and Joslyn, M. R. (2008), "Political Persuasion During Times of Crisis: The Effects of Education and News Media on Citizens' Factual Information about Iraq," *Journalism & Mass Communication Quarterly*, 85, 591–608. DOI:10.1177/107769900808500307 [10]
- Jun, Y., Meng, R., and Johar, G. V. (2017), "Perceived Social Presence Reduces Fact-Checking," *Proceedings of the National Academy of Sciences* 114, 5976–5981. DOI:10.1073/pnas.1700175114 [1]
- Lazer, D. M., Baum, M. A., and Benkler, Y. (2018), "The Science of Fake News," *Science*, 359, 1094–1096. DOI:10.1126/science.aao2998 [1]
- Lee, K., Mahmud, J., Chen, J., Zhou, M., and Nichols, J. (2014), "Who Will Retweet This?: Automatically Identifying and Engaging Strangers on Twitter to Spread Information," *Proceedings of the 19th International Conference on Intelligent User Interfaces, IUI '14*, pp. 247–256, New York: ACM. DOI:10.1145/2557500.2557502 [2]
- Liu, Y., and Wu, Y. F. (2018), "Early Detection of Fake News on Social Media through Propagation Path Classification with Recurrent and Convolutional Networks," *AAAI Conference on Artificial Intelligence*. Available at <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16826/15707>. [2]
- Louni, A., and Subbalakshmi, K. P. (2014), "A Two-Stage Algorithm to Estimate the Source of Information Diffusion in Social Media Networks," *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. DOI:10.1109/infcomw.2014.6849253 [2]

- Manickam, I., Lan, A. S., Dasarathy, G., and Baraniuk, R. G. (2019), “Ideo-trace: A Framework for Ideology Tracing with a Case Study on the 2016 U.S. Presidential Election,” *arXiv:1905.08831v2*. [3]
- Miao, W., Pan, Q., and Gastwirth, J. L. (2022), “A Misuse of Statistical Reasoning: The Statistical Arguments Offered by Texas to the Supreme Court in an Attempt to Overturn the Results of the 2020 Election,” *Statistics and Public Policy*, 9, 67–73. DOI:10.1080/2330443X.2022.2050327 [2]
- Mikkelsen, D. (2020), “Are Non-respirator Masks Ineffective at Protecting against COVID-19?” Available at <https://www.snopes.com/fact-check/cloth-masks-ineffective-covid19/>, accessed on 04 May 2021. [1]
- Nordio, A., Tarable, A., Chiasserini, C. F., and Leonardi, E. (2017), “Belief Dynamics in Social Networks: A Fluid-based Analysis,” *IEEE Transactions on Network Science and Engineering*, 5, 276–287. DOI:10.1109/TNSE.2017.2760016 [10]
- Papanastasiou, Y. (2017), “Fake News Propagation and Detection: A Sequential Model,” *SSRN Electronic Journal*. DOI:10.2139/ssrn.3028354 [2,11]
- Pierri, F., Piccardi, C., and Ceri, S. (2019), “Topology Comparison of Twitter Diffusion Networks Reliably Reveals Disinformation News,” arXiv preprint arXiv:1905.03043. [2]
- Ribeiro, F., Henrique, L., Benevenuto, F., Chakraborty, A., Kulshrestha, J., Babaei, M., and Gummadi, K. (2018), “Media Bias Monitor: Quantifying Biases of Social Media News Outlets at Large-Scale,” *International AAAI Conference on Web and Social Media*. Available at <https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17878>. [5,6,7,11]
- Rosling, H. (2018), *Factfulness: Ten Reasons We’re Wrong About the World - And Why Things Are Better Than You Think*, New York: Flatiron Books. [9]
- Roth, Y., Pickles, N. (2020), “Updating Our Approach to Misleading Information,” *Twitter*. Available at https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html, accessed on 04 May 2021. [11]
- Sahinidis, N. V. (2017), *BARON 17.8.9: Global Optimization of Mixed-Integer Nonlinear Programs*, User’s Manual. Available at <http://www.minlp.com/downloads/docs/baron%20manual.pdf> [4]
- Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., Liu, Y. (2019), “Combating Fake News: A Survey on Identification and Mitigation Techniques,” *ACM Transactions on Intelligent Systems and Technology*, 10, 1–42. DOI:10.1145/3305260 [1]
- Shu, K., Bernard, H. R., and Liu, H. (2018), “Studying Fake News via Network Analysis: Detection and Mitigation,” in *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, Lecture Notes in Social Networks, eds. N. Agarwal, N. Dokoohaki, and S. Tokdemir, pp. 43–65, Cham: Springer. DOI:10.1007/978-3-319-94105-9_3 [1]
- Tawarmalani, M., and Sahinidis, N. V. (2005), “A Polyhedral Branch-and-Cut Approach to Global Optimization,” *Mathematical Programming*, 103, 225–249. [4]
- Vicario, M. D., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., Quattrociocchi, W. (2016), “The Spreading of Misinformation Online,” *Proceedings of the National Academy of Sciences*, 113, 554–559. DOI:10.1073/pnas.1517441113 [1,2]
- Vosoughi, S., Roy, D., and Aral, S. (2018), “The Spread of True and False News Online,” *Science*, 359, 1146–1151. DOI:10.1126/science.aap9559. [2]
- Xiang, Y., and Sarvary, M. (2007), “News Consumption and Media Bias,” *Marketing Science*, 26, 611–628. DOI:10.1287/mksc.1070.0279. [3]