Using Agent-Based Modelling to Evaluate the Impact of Algorithmic Curation on Social Media

ANNA GAUSEN, Imperial College London, United Kingdom
WAYNE LUK, Imperial College London, United Kingdom
CE GUO, Imperial College London, United Kingdom

Social media networks have drastically changed how people communicate and seek information. Due to the scale of information on these platforms, newsfeed curation algorithms have been developed to sort through this information and curate what users see. However, these algorithms are opaque and it is difficult to understand their impact on human communication flows. Some papers have criticised newsfeed curation algorithms that, while promoting user engagement, heighten online polarization, misinformation, and the formation of echo chambers. Agent-based modelling offers the opportunity to simulate the complex interactions between these algorithms, what users see, and the propagation of information on social media. This paper uses agent-based modelling to compare the impact of four different newsfeed curation algorithms on the spread of misinformation and polarization. This research has the following contributions. (1) Implementing newsfeed curation algorithm logic on an agent-based model. (2) Comparing the impact of different curation algorithm objectives on misinformation and polarization. (3) Calibration and empirical validation using real Twitter data. This research provides useful insights into the impact of curation algorithms on how information propagates and on content diversity on social media. Moreover, we show how agent-based modelling can reveal specific properties of curation algorithms, which can be used in improving such algorithms.

CCS Concepts: • **Mathematics of computing** → *Stochastic processes*; • **Computing methodologies** → **Modeling and simulation**; **Distributed artificial intelligence**.

Additional Key Words and Phrases: misinformation, filter bubbles, polarization, agent-based modelling, simulation, social media networks

ACM Reference Format:
Anna Gausen, Wayne Luk, and Ce Guo. 2022. Using Agent-Based Modelling to Evaluate the Impact of Algorithmic Curation on Social Media. In *ACM Journal of Data and Information Quality on "Truth and Trust Online"*. ACM, New York, NY, USA, 24 pages. https://doi.org/xx.xxxx/xxxxxxx.xxxxxx

1 INTRODUCTION

Social media has reshaped our relationship with communication and information sharing. The scale of social media companies is unprecedented, Facebook was recorded to have 1.93 billion daily active users at the third quarter of 2021 [41]. Though misinformation and polarization are not new phenomena, the scale of social media has exacerbated these issues. This can undermine democratic processes [20], public health policies [31] and trust in traditional media [27]. There is extensive research into countering misinformation on social media, much of which is focused on fake news detection. Whilst this is an important approach, there are issues with accuracy, censorship and bias in what is detected. Social media platforms cannot control who joins and what users post until they have already broken the community

Manuscript submitted to ACM

guidelines. What social media platforms can control is the recommendation algorithms they have developed to curate users' newsfeeds and influence what users see. These algorithms could be optimised for higher information quality, minimising the spread of misinformation and reducing polarization. This paper introduces a novel approach that could be used for such optimisations, based on real data.

It is challenging to understand the complex consequences of human interactions with digitized technologies [9]. The newsfeed curation algorithms and social media users interact to form a complex socio-technical system in which human communication flows are altered by the algorithm. There have been some existing studies into the impact of curation algorithms on the formation of echo chambers and information quality. However, as many studies are conducted using "puppet accounts" [5] they can only be carried out on a small scale with a limited number of users.

Agent-based modelling offers the opportunity to simulate and understand the complex interactions between users, algorithms and flows of information on social media. These benefit from the ability to model a high number of agents, for this research $N = 1,000$, and to compare the impact across different curation algorithms. This paper will use an agent-based model of Twitter to compare the impact of four different newsfeed curation algorithms on the propagation of information, echo chamber formation and polarization.

Research Questions

Our study is intended to address the following questions.

(1) What impact do curation algorithms have on the spread of information on social media?
(2) What impact do curation algorithms have on the formation of echo chambers and polarization on social media?
(3) What is the relationship between the content diversity of users' newsfeeds and the propagation of misinformation?

The first research question will be answered through quantitatively measuring the difference in the propagation of a tweet, using number of (re)tweets per timestep, through the simulated network across four different newsfeed curation algorithms. The second research question will be answered by comparing the belief purity of the agents' newsfeeds on the simulated network between the four algorithms. Belief purity provides a measure of reduced content diversity, which can lead to polarization and echo chamber formation. The final question will be investigated by comparing both the measured tweet propagation of the false story and the belief purity of the agents' newsfeeds for each of the four curation algorithms. Further discussions concerning these questions can be found in Section 6.

Contributions

This research builds on work by Gausen et al. [17] that uses agent-based modelling to evaluate misinformation countermeasures on social media. This paper extends the agent-based model to include logic to capture a newsfeed curation algorithm and it measures additional social metrics to capture polarization and echo chamber formation. These extensions enable the model to be used to evaluate the social impact of different newsfeed curation algorithms and to address the research questions outlined above. This research has the following novel contributions:

(1) Developing an agent-based model of Twitter to compare newsfeed curation algorithms in terms of filter bubble formation and the spread of misinformation.
(2) Implementing four curation algorithms with different objectives in the logic of the agent-based model: chronological, shared-beliefs, random and popularity.

(3) Calibrating and empirically validating the model using Twitter cascades of both false and non-false stories which spread with a chronological newsfeed.

The results of this study have produced new insights into the impact of different newsfeed curation algorithm objectives. The results highlight that, without accounting for human cognitive biases or selective exposure, simply varying the objective of the newsfeed curation algorithm has a significant impact of the spread of information and echo chamber formation. It has shown that belief-based and popularity-based newsfeed curation algorithms lead to considerable increases in tweets shared compared to a chronological newsfeed. In the case of false news or misinformation this will lead to an amplification of the spread, which could be damaging. In addition, it shows that a belief-based curation algorithm can lead to a 30-60% increase in belief purity of an user's newsfeed, which can have implications for echo chamber formation and online polarization. The results of this paper do not exhibit a consistent relationship between the average belief purity of agents' newsfeeds and the spread of misinformation. However, in this analysis the misinformation case-study is not associated with a particular belief and is based on popular-culture. Future work could analyse the relationship between belief-based misinformation and echo chamber formation or polarization, such as partisan misinformation.

Paper Organisation

The remainder of this paper is structured as follows: Section 2 analyses existing work related to newsfeed curation on social media and the application of agent-based modelling to social networks. Section 3 provides a description of the agent-based model, the implementation of the curation algorithms and outlines the model assumptions. Section 4 describes the methodology for calibrating, validating and simulating the model, including the datasets used. Section 5 presents the results of the simulations. These are discussed in Section 6 in relation to the three research questions posed above. Finally, Section 7 summarises the findings and outlines avenues for future work.

2 BACKGROUND AND RELATED WORK

This section will review existing literature on the topics of social media newsfeed curation algorithms, existing approaches to auditing these algorithms and agent-based model of social media networks.

2.1 Curation Algorithms

Newsfeed curation algorithms have been developed to sort through the vast amounts of information on social media platforms. Users are physically unable to view everything so, according to Twitter, the algorithms aim to curate the information "to feature compelling, original, and diverse content" [44]. There has been some criticism that the algorithms lead to increased polarization by recommending tweets that are most aligned with user beliefs. "Despite optimistic talk about 'collective intelligence,' the Web has helped create an echo chamber where misinformation thrives" [42]. However, there is mixed evidence for this. Dylko et al. [14] found empirical evidence that customisable technologies increased political polarization through selective exposure. Whilst Cinelli et al. [10] found that individual choices still play a significant role. Howard et al. argue that solutions like "introducing random news stories and ensuring exposure to high quality information would be a simple — and healthy — algorithmic adjustment to social media platforms" [19].

Whilst curation algorithms often find criticism for their supposed role in polarization, echo chamber formation and misinformation; there are a number of papers exploring how these algorithms could be programmed to be societally beneficial. Helberger et al [18] propose using exposure diversity as a design principle for recommender systems. The

EU High Level Group on Media Freedom and Pluralism [34] says that "increasing filtering mechanisms makes it more likely for people to only get news on subjects they are interested in, and with the perspective they identify with." The paper warns that this could have a "negative impact on democracy" [34]. However, it suggests that "diversity-sensitive design" [34] and approaches that benefit "normative, social objectives" [34] over corporate preference could be beneficial. Similarly, Bozdag et al. [7] compare different software designs to try to prevent echo chambers. Another important consideration they highlight is how social norms are embedded in these tools. Currin et al. [12] explore the use of a random dynamical nudge to increase the diversity of content users see.

2.2 Auditing Curation Algorithms

Social media newsfeed curation algorithms are under increased scrutiny as they control how information propagates online. The details of how these algorithms are programmed is very opaque. There have been a number of different approaches to auditing these algorithms.

Two studies have tried to audit the algorithms using "puppet accounts" to provide improved understanding of the impact they have on the information users see [5] [6]. Bandy et al. [5] found that the curation algorithm, when compared to a chronological newsfeed, resulted in an increase in unique accounts and partisan echo chambers and a decrease in dominance of accounts with high number of followers and exposure to links. Barltey et al. [6] found that the curation algorithm showed more popular and less recent tweets compared to a chronological timeline. Koumchatzky et al. [23] provide additional details on what is involved in Twitter's deep learning curation algorithms. These studies provide interesting information about the impact the algorithm has on what appears on users' newsfeeds. However, they are only carried out with a small number of users and are unable to provide information on the impact these algorithms have on how information propagates across the network.

Other papers have proposed frameworks that map out stakeholder interests and ethics-oriented metrics [8]. However, these frameworks only offer static and qualitative evaluations and do not cover the social metrics proposed in this paper. Mozilla designed an audit of the YouTube recommendation algorithm that leveraged crowdsourced complaints. However, this relies on user engagement and there may be biases in who reports and what is reported [30].

From this review, it is evident that, despite efforts, it is very challenging to understand the objectives of the curation algorithms used by social media companies and equally challenging to understand the impact of these algorithms. This paper will be novel in its use of agent-based models to simulate and compare the impact of curation algorithms with different objectives on both the spread of information and content diversity. Unlike the auditing approaches described above, agent-based modelling enables dynamic, long-term analysis and scenario-testing with large populations of agents.

2.3 Agent-based Models of Social Networks

Agent-based models (ABMs), or artificial societies, are a way to model populations of individuals. The behaviour of each agent is controlled by a set of rules, which are often simple at the agent level but when combined can capture more complex phenomenon, known as emergent behaviour [25]. In the literature there are some existing ABMs modelling the spread of misinformation and echo chambers on social networks. Misinformation is a complex term to define but it is used to encompass "harmful, false and/or misleading information online" [35]. Echo chambers describe the scenario where individuals are surrounded by information that is aligned with their own beliefs or opinions [2].

This section will review related works in regard to five questions to identify a gap in existing literature. The questions are specific to the proposed model in this paper and offer an extension to the review in [17]. These are: (1) Does the

ABM model misinformation? (2) Is the ABM of a social media network? (3) Does the analysis use real social media data for calibration and validation? (4) Does the ABM model agent beliefs? (5) Does the ABM model newsfeed curation algorithms? The review can be seen in Table 1. A model that fulfills all five of these review questions is required to address the research questions in Section 1.

Table 1. Review of Related Works

Related Works	Q1	Q2	Q3	Q4	Q5
Alvarez-Galvez [2016]	Y				
Fan et al. [2018]		Y			
Fränken and Pilditch [2020]	Y	Y		Y	
Jin et al. [2013]	Y	Y	Y		
Kaligotla et al. [2016]	Y	Y	Y		
Kozitsin and Chkhartishvili [2020]		Y		Y	Y
Madsen et al. [2018]		Y		Y	
Onuchowska and Berndt [2020]	Y	Y			
Sathanur et al. [2015]	Y	Y			
Serrano and Iglesias [2016]	Y	Y	Y		
Tang et al. [2014]		Y			
Proposed Model	Y	Y	Y	Y	Y

This review of related work highlights the diversity in the types of ABMs that have been developed in this field. There are a number of different taxonomies of agents that have been employed. In addition, there are a variety of mechanisms that have been developed for agents to communicate and different phenomena being measured. A large number of the reviewed papers, such as Serrano et al. [37], focus on modelling misinformation. Fewer papers tackle agent beliefs and echo chamber formation. Madsen et al. [29] show that large networks of rational agents, meaning agents with no cognitive biases, will still result in the formation of echo chambers. A similar paper by Franken et al. [16] also found ABMs could capture the formation of echo chambers with information cascades, by comparing social and asocial agents. Kozitsin et al. [24] is the only reviewed paper that examines the relationship between recommendation algorithms and echo chamber formation. However without grounding in real data or a significant population size ($N = 50$), it cannot provide a convincing analysis of the relationship.

From this review, it is clear that there is a gap in existing literature which is bridged by this paper: an ABM to model the relationship between curation algorithms, agent beliefs and misinformation on social networks, which is calibrated with real social media data and is empirically validated. The proposed model will be novel in addressing all five questions shown in Table 1. In addition, it will have the ability to compare and evaluate the impact of multiple newsfeed curation algorithms on both belief purity of agents' newsfeeds and information diffusion, which has not been seen in existing literature.

3 PROPOSED MODEL

This section describes the model proposed by this paper. It contains a detailed description of the ABM, the implementation of the curation algorithms and justifications for the model assumptions.

3.1 Model Description

The model proposed in this paper is an ABM based on Twitter where the agents represent social media users, the connections between agents represent a follower/followee relationship, and agents can see information captured by tweets, shared by their neighbours. The information the users see is controlled by a newsfeed curation algorithm, which is programmed with four different and independently implemented objectives. The model is inspired by epidemiology modelling based on the conceptual similarity between how misinformation spreads across a social network and how a disease spreads through a population [21] [32]. Due to the scale and pace that misinformation can spread online, it is helpful to model misinformation in a similar way to a viral pathogen [45].

Figure 1 illustrates how the ABM is programmed. The proposed ABM models three related aspects of a social network: the newsfeed curation algorithm, the spread of information and polarization or echo chamber formation. The newsfeed curation algorithm logic represented by "get curated newsfeed" in Figure 1. The spread of information is represented by the "tweet" and "retweet" functionality in the figure. Finally, the echo chamber formation is captured when the model "updates state and beliefs" of the agents and calculates belief purity in the "agent reporter". More details on the modelling and measurement of echo chamber formation can be found in Section 3.2 on belief modelling and Section 4.2 on the belief purity metric. The network has a Barabasi-Albert network structure [47]. The model has four agent states in relation to the story being studied: susceptible (agents who are susceptible to the story), believe (agents who have seen and believe the story), deny (agents who have seen and not believed the story) and cured (agents who had previously believed the story but no longer do). These agent states have been used in existing literature [17].

To initialise the model: the population, average number of followers/followees and curation objective are set. Then the agents are generated, including the connections to their neighbours, their state and their initial beliefs, which are sampled between 0 and 1. Each agent has the following attributes: followers/followees, time of most recent post, retweet number, beliefs, whether their account is verified and number of tweets they have posted about the story of interest.

At each timestep, the agent will sample a probability to determine if they are online. If they are online, they sample probabilities to determine whether they post a tweet and whether they read N_{posts} tweets from their newsfeed and potentially retweet. If they post a tweet, this tweet will be based on their current state. If the agent views their newsfeed, they will read the top N_{posts} tweets, where N_{posts} is sampled each time. The way in which the N_{posts} tweets are chosen out of all of their neighbours' tweets will based on the specified newsfeed curation objective logic. The content of the tweets viewed in the agents' newsfeed can be retweeted and integrated into the agent's beliefs. This logic is only analysed for agents who are susceptible to the story. For agents who believe and have tweeted about the story, there is a probability they may later reject this belief or stop sharing the story. If an user is verified, they have greater influence meaning the the probability of others sharing their tweets is higher. At the end of each timestep, the agent reporter records a set of metrics which are tracked through the duration of the simulation, as seen in Figure 1.

The behaviour of this model is controlled by a set of probabilities, which are seen in Figure 1: (1) Mean probability that the agent is online P_{online}. (2) Mean probability that the agent reshares information they have seen on their newsfeed $P_{reshare}$. (3) Mean probability that the agent rejects their previously held beliefs or stop sharing tweets about them P_{reject}. The behaviour is also controlled by the standard deviation of these probability distributions that each agent samples from. The mean and standard deviation of the distributions of each probability are found through calibration with real Twitter datasets, discussed in Section 4.3. Other network level values such as population size, average node degree, proportion initially infected, proportion of verified users and average number of posts seen are taken from analysis of real data on both false and non-false stories, see Section 4.5 for more details.

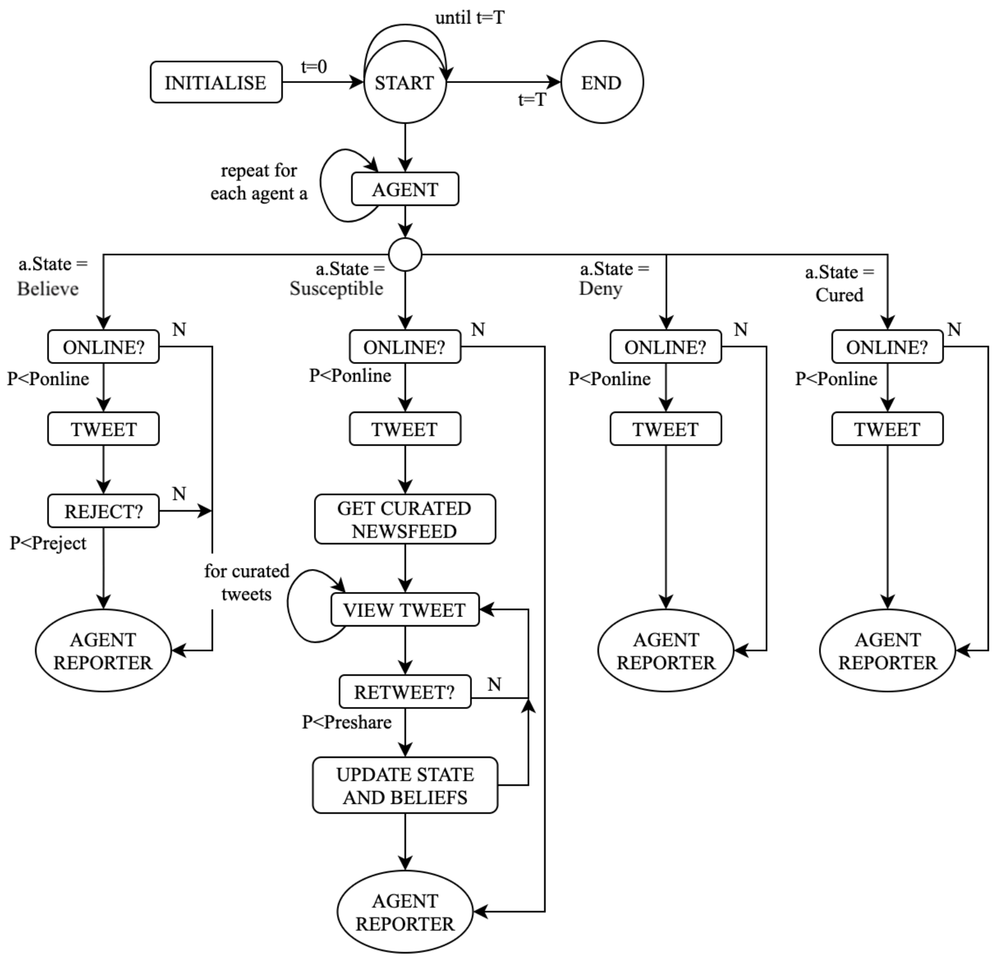


Fig. 1. Diagram describing Proposed Model logic.

3.2 Belief Modelling

The model is populated with Bayesian agents. The agents' beliefs are influenced by the information they see from their neighbours through Bayesian belief updating, meaning that there is information integration and their beliefs are updated without cognitive bias [29]. Bayesian belief updating is applied using Bayes Theorem [1]. So $P(H|E)$, which is the probability of an agents own beliefs given the new evidence, is calculated as:

$$P(H|E) = P(H) * P(E|H)/P(E) \tag{1}$$

where $P(H)$ is the agents' prior belief and $P(E)$ is the probability that evidence occurs irrespective of an agents' own beliefs. This is calculated as:

$$P(E) = \frac{1}{\sigma_T * \sqrt{2\pi}} * e^{\frac{(X-\mu_T)^2}{2\sigma_T^2}} \tag{2}$$

where σ_T and μ_T are the standard deviation and mean of the true distribution and X is a piece of new evidence seen by the agent. This evidence X is the belief of a neighbouring agent, where $0 \leqslant X \leqslant 1$. Then $P(E|H)$, the probability of the evidence considering the agents' own beliefs, is calculated as:

$$P(E|H) = \frac{1}{\sigma_A * \sqrt{2\pi}} * e^{\frac{(X-\mu_A)^2}{2\sigma_A^2}} \tag{3}$$

where σ_A and μ_A are the standard deviation and mean of the agents own belief distribution.

3.3 Curation Algorithm Implementation

This proposed model is novel as it attempts to simulate the curation algorithms that sort users' newsfeeds. The curation algorithms control which of its neighbours' posts an agent sees. As mentioned earlier the number of posts N_{posts} is sampled from a distribution with a mean $\bar{N}_{posts}$ and a standard deviation $\bar{N}_{posts}/2$, where the value of $\bar{N}_{posts}$ comes from analysis of user activity on Twitter [4] discussed in Section 4.5. There are four control algorithm objectives compared in this paper:

(1) Chronological: this means the agent sees the most recent N_{posts} tweets posted by its neighbours, this is given by the agent attribute: time of most recent post.
(2) Beliefs: this means the agent sees the N_{posts} tweets from its neighbours with the most similar beliefs, which could be a proxy for interest. Beliefs are given by the agent attribute: current beliefs.
(3) Popularity: this means the agent sees the most popular N_{posts} tweets from its neighbours. Popularity is calculated by accounting for the number of retweets of that tweet and the number of followers the neighbour has.
(4) Random: this means that the agents sees a random group of N_{posts} tweets posted by its neighbours.

Each of these curation algorithms are applied independently and will be uniform for all agents in the network for a given simulation. This is a simplification of actual recommendation algorithms, which are driven by multiple objectives. The algorithms will most likely curate tweets based on how new, popular and aligned with the user's interests they are. However, due to the opaqueness of these algorithms it is not possible to directly approximate their behaviour. So this study will treat each objective independently to provide an understanding of the impact each has on information cascades and newsfeed content diversity.

3.4 Model Assumptions

ABMs provide a simplified approximation of the real population they are modelling. The proposed model in this research contains the following assumptions:

- Connections between agents are kept constant for the duration of the simulation.
- Bayesian agents on a Barabasi-Albert Network.
- The tweets being analysed are not aligned to a belief.
- False and non-false stories travel with the same logic.

The simplification of a constant network structure and the choice of Bayesian agents are seen as appropriate for the analysis in this paper due to the short time duration and the focus of the analysis. This paper is trying to evaluate the impact of changing the objective of the newsfeed curation algorithm; by maintaining the connection between nodes and having agents with no cognitive biases means these additional layers of complexity do not interfere with the results. The decision not to align the tweets (i.e. the infected tweet) with a belief was made because often this relationship is assumed due to a US-centric focus where partisanship and misinformation have been seen to be correlated [13]. For the analysis presented in the body of this paper, the two case-studies analysed are related to popular culture and are not aligned with belief. Keeping tweets separate from belief makes the analysis more globally applicable however in some cases this will be a limitation. Finally, the parameter values and probability distributions that control the propagation structures of the tweets are what differentiate between false and non-false tweets. A number of existing papers, most notably from Vosoughi et al [46], have shown that there are distinct differences in the propagation structures of false and true news.

4 METHODOLOGY

This section describes the methodology used to calibrate, validate and run the proposed model for the experimental results. The proposed model was calibrated and empirically validated with real Twitter data on one false and one non-false case-study. Then it was run for each newsfeed curation objective across 5 simulations and the results were compared. The two metrics measured and compared for each story were the average and standard deviation of the proportion of posts about the story and the belief purity of the agents' newsfeeds over the simulations.

4.1 Simulation

For each set of results the ABM was parameterised with the fixed model parameter values in Table 2. This table provides the parameter values used across both runs and the parameters that are specific to each case-study. The population size was chosen as it is large enough to capture the values of the proportion of tweets per a timestep Φ (Equation 4) seen in the real data. For example, a population of 100 agents would not be able to produce $\Phi < 1$ and would therefore be unable to capture the behaviour seen in the real world data. It should be noted that the simulated population size of $N_{agents} = 1,000$ is not realistic and future work into accelerating ABMs could enable a more realistic population size to be modelled. The choice of $\bar{N}_{posts}$ is justified in Section 4.5. The story-specific parameters: the proportion of users initially sharing the tweet of interest, and the total number of timesteps, in hours, were calculated from the real dataset [28], also discussed in Section 4.5. The model was calibrated for each story in the real dataset [28]. The model was calibrated to provide the mean probability values and their standard deviation, presented in Table 3. The calibration process is discussed in detail in Section 4.3.

For each story the model is run with the parameters in Table 2 and the calibrated probabilities in Table 3 for 5 simulations for each of the four curation algorithm objectives. Justification for the choice of the number of simulations can be found in Appendix A. For each simulation, the proportion of tweets about the story and average belief purity were recorded at each timestep. The proportion of tweets shared about the story at each timestep provides a measure of the spread of the information and, in the case of the false story, the misinformation. The proportion of tweets Φ is calculated as:

$$\Phi = \frac{N_{tweets}}{N_{agents}} \tag{4}$$

Table 2. Model Parameter Values: the fixed values provided to the ABM

Run	Parameters	Values
Both	N_{agents}	1000
	$N_{followers}$	100
	N_{posts}	40
	N_{sims}	5
False	Initially Inf	0.7
	Timesteps	80
Non-False	Initially Inf	0.5
	Timesteps	40

Table 3. Calibrated Parameter Values: the calibrated mean of the probabilities that are sampled during the ABM

Probabilities	False Story	Non-False Story
$P_{reshare}$	0.03669678	0.0770166
P_{reject}	0.01459229	0.06335449
P_{online}	0.10714111	0.06335449

where N_{tweets} is the number of tweets posted about the story in each timestep and N_{agents} is a proxy for the population in the simulated or real data, depending on the calculation. For the real data, the population proxy was calculated by multiplying the ratio between the simulated population and the simulation average node degree by the average number of followers in the real world data. The proportion of tweets about the story provides a way of scaling the number of tweets between the real-world data and the simulated data. This adjustment is required as in the real-world data, the number of users and number of connections is far greater than in the ABM.

4.2 Belief Purity

The belief purity (ϵ) is a measure of the similarity in beliefs between an agent and the posts on their curated newsfeed. The belief purity of an agent's newsfeed is the sum of the difference in beliefs between the agent and each neighbour on their newsfeed, over the total number of neighbours on their newsfeed, given by:

$$\epsilon = 1 - \frac{\sum_0^{N_{posts}} \Delta}{N_{posts}} \tag{5}$$

where N_{posts} is number of posts and Δ is the difference between an agents own beliefs and the belief of the tweet they are viewing. This is calculated as:

$$\Delta = |P(H)_{agent} - P(H)_{neighbour}| \tag{6}$$

where $P(H)_{agent}$ is the agents current belief and $P(H)_{neighbour}$ is the neighbour's current belief. This is a measure of potential for echo chamber formation and polarization through belief purification. This metric is used as a proxy for the echo chamber effect. Echo chambers are often characterised by two attributes, firstly homophily of the network and secondly bias towards similar agents in information diffusion [11]. The metric of belief purity does not account for this second attribute. Future work could try to estimate this metric as part of the evaluation of echo chamber formation. The higher the belief purity of an agent's newsfeed, the greater the echo chamber effect.

For the full set of simulations, the average and standard deviation of the belief purity over the time period were plotted. These results where plotted for each newsfeed curation algorithm and can be found in Section 5.

4.3 Calibration and Validation

The behaviour of the model is controlled by three probabilities: mean probability of re-sharing content, mean probability of rejecting content and mean probability of being online. In addition to these probabilities is their standard deviation. These parameters are calibrated to simulate the behaviour of real-world data using a combination of a machine learning surrogate and optimised parameter choice. The machine learning surrogate [26] enables a broad parameter space to be explored with lower computational cost. Then the parameter selection is further refined by finding the choice that minimises the mean-square error (MSE) between the simulation output and the real-world values of Φ.

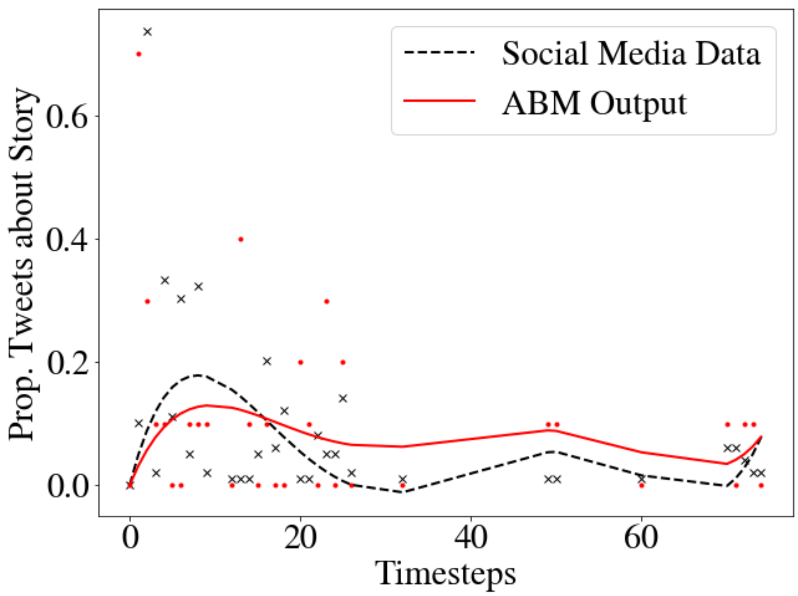


Fig. 2. Graph comparing the output of the proposed model with chronological newsfeed (red) to social media dataset (black) for the false tweet. The crosses represent the real data points and the dots represent the ABM output. The RMSE is 0.18 and NRSME 0.25.

The chronological curation objective was used for the baseline ABM for calibration. This is a valid baseline, as at the time the two case-studies were posted Twitter had a chronological newsfeed. Comparing the output of the calibrated model to the Twitter data-points enables one to measure how well the model captures the real user behaviour, as seen in Figure 2. This enables validation of the assumptions and simplifications built into the model. To quantify this, the root-mean-square-error (RMSE) and normalised root-mean-square-error (NRMSE) are calculated.

$$NRMSE = \frac{RMSE}{\Phi_{max} - \Phi_{min}} \tag{7}$$

where Φ_{max} and Φ_{min} are the maximum and minimum of the proportion of tweets across all time steps in a simulation.

Once the model has been empirically validated with the baseline setting of a chronological newsfeed, each of the other newsfeed curation objectives can be applied independently with the same calibrated parameter values and their output is compared. These results are presented in Section 5.

4.4 Case-Studies

Each ABM model represents a case-study and the parameters are tuned to match the behaviour in that case-study. This approach is commonplace in existing ABM literature [37] [21]. Additional case-studies provide additional examples for comparison but they do not enrich the calibration or validation processes, which are case-study dependent. The results therefore only reflect the impact of a given case study and are not expected to be understood as representing all examples of true or fake news. A potential future avenue would be to look at average propagation patterns of true and fake news and set the ABM parameters to these average values. This could enable analysis on a characterisation of the propagation of a false tweet. The case-studies in this paper were chosen due to the data coverage, number of tweets, timings of posting and fact-checked labels; more information is provided in Section 4.5. An additional case-study is presented in Appendix B for comparison and to confirm that it supports the findings presented in this paper.

4.5 Data

Three different data sources were required to initialise, calibrate and empirically validate the proposed model. The most important pieces of data required to calibrate and empirically validate the model were the cascades of two tweets on Twitter, one false and one non-false. These tweets had to be posted when Twitter had a strictly chronological newsfeed, before 2016, so that the model could be calibrated with a chronological newsfeed as the baseline. The dataset used for this was from Ma et al [28]. Two tweets about popular culture were chosen as case-studies for this analysis, however there is scope to analyse further stories as future work to understand if the conclusions hold. The false tweet is: "tupac was hiding for 18 years URL" and the non-rumour tweet is: "justin bieber must be loving himself right now...". For each tweet the dataset maps the user and retweet structure through time.

Some additional datasets were required to initialise some of the model parameters. Firstly, an understanding of retweet activity for verified users compared to non-verified users was required. This was calculated from the FakeNewsNet dataset [38] [40] [39] based on the activity of 123,050 true and 67,877 false tweets. Secondly, user activity data were used to estimate the number of posts users view when online. Inspired by the logic in [4], in this model the average number of posts viewed by a user was 40 posts as this is the number that users are most likely to engage with.

5 RESULTS

This section shows the results of the simulation runs for the false and non-false case-studies. Tables 4 and 5 show the average change in three metrics compared to the chronological simulation run.

Table 4. Results for False Case-Study

Curation Objective	Change in Φ_{avg}	Change in Φ_{max}	Change in Belief Purity
Time	0.00	0.00	0.00
Popularity	6.04	1.11	-0.10
Beliefs	1.42	0.11	0.32
Random	-0.04	0.00	-0.07

These metrics are the change in the mean proportion of tweets about the story Φ_{avg}, the maximum proportion of tweets about the story Φ_{max} and the mean change in belief purity ϵ over the simulation. These are calculated as the difference between the chosen objective value and the chronological value, over the chronological value for that metric.

Table 5. Results for Non-False Case-Study

Curation Objective	Change in Φ_{avg}	Change in Φ_{max}	Change in Belief Purity
Time	0.00	0.00	0.00
Popularity	6.15	2.44	-0.35
Beliefs	0.63	0.10	0.58
Random	-0.37	-0.37	0.01

Figures 3 and 5 show the plots of the mean and standard deviation of the proportion of posts per timestep on the story of interest, across 5 simulations. These graphs compare the propagation of the story through the network for each of the four curation algorithms. The graph has timesteps on the x-axis and proportion of tweets about the story Φ posted in each timestep on the y-axis. On the graph, the points represent the real social-media data points from [28] and the solid line represent a polynomial fit for those real data points. The dotted lines on the graphs represent the polynomial fit for the mean simulation output and the shaded region represents the standard deviation across the runs. Figures 4 and 6 show plots of the mean and standard deviation of how the belief purity changed over the duration of the simulation, across 5 runs, for each of the four curation algorithms. The dashed line is the mean and the shaded region shows the standard deviation.

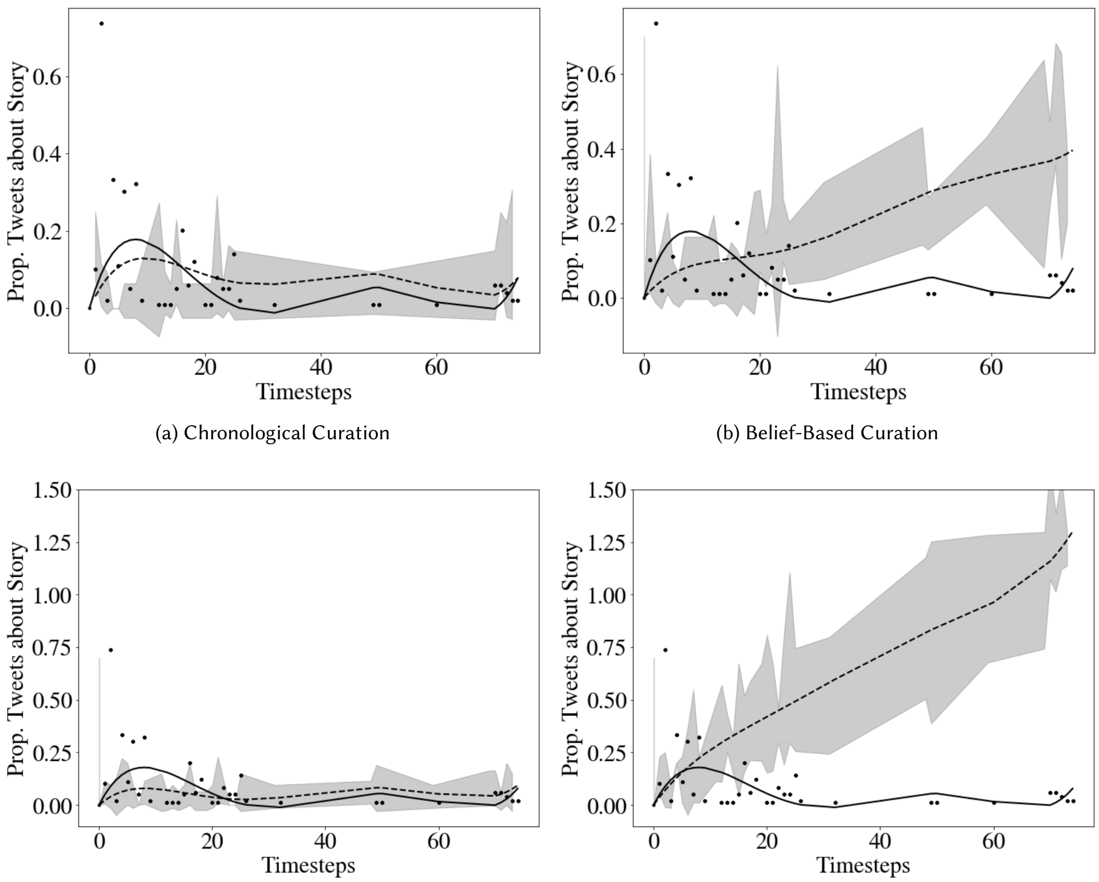
6 DISCUSSION

The results in Section 5 present how varying the objective of the newsfeed curation algorithm on a social network can impact both the spread of information and the diversity of content on a user's newsfeed. This has repercussions for the spread of misinformation, echo chamber formation and online polarisation. This was investigated through the development of an ABM based on Twitter and two case-studies of false and non-false tweets propagating on Twitter. The models were run with four different curation algorithm objectives and their behaviour was compared. Tables 4 and 5 show the comparison between the ABM results for each curation algorithm through a set of metrics. The following sections will discuss these results in relation to the three research questions raised in Section 1.

1. What impact do curation algorithms have on the spread of information on social media?

The first question this research poses is what impact do newsfeed curation algorithms have on the propagation of information, particularly misinformation, on social media networks. The propagation of information is quantified by measuring the proportion of tweets or retweets Φ about the chosen story in each timestep. In the case of the false story, this measures the propagation of misinformation. Figure 3 and 5 show graphically how the propagation of information changes for each newsfeed curation algorithm and will be discussed in more detail below.

The first set of results shown in Figure 3 show the propagation of a real false tweet about Tupac [28] (solid line) and compares it to the simulated propagation of the same tweet with different curation objectives (dashed line). The shaded region shows the standard deviation of the simulation results across each of the five runs. Figure 3 (a) shows the comparison between the real-world data and the simulated data with a chronological newsfeed. This is the calibrated baseline as the users in the data were experiencing a chronological newsfeed. The other three results, Figure 3 (b)-(d) compare the real-world propagation of the Tupac tweet (which had a chronological newsfeed) to the simulated propagation of the tweet with other curation objectives.



(a) Chronological Curation

(b) Belief-Based Curation

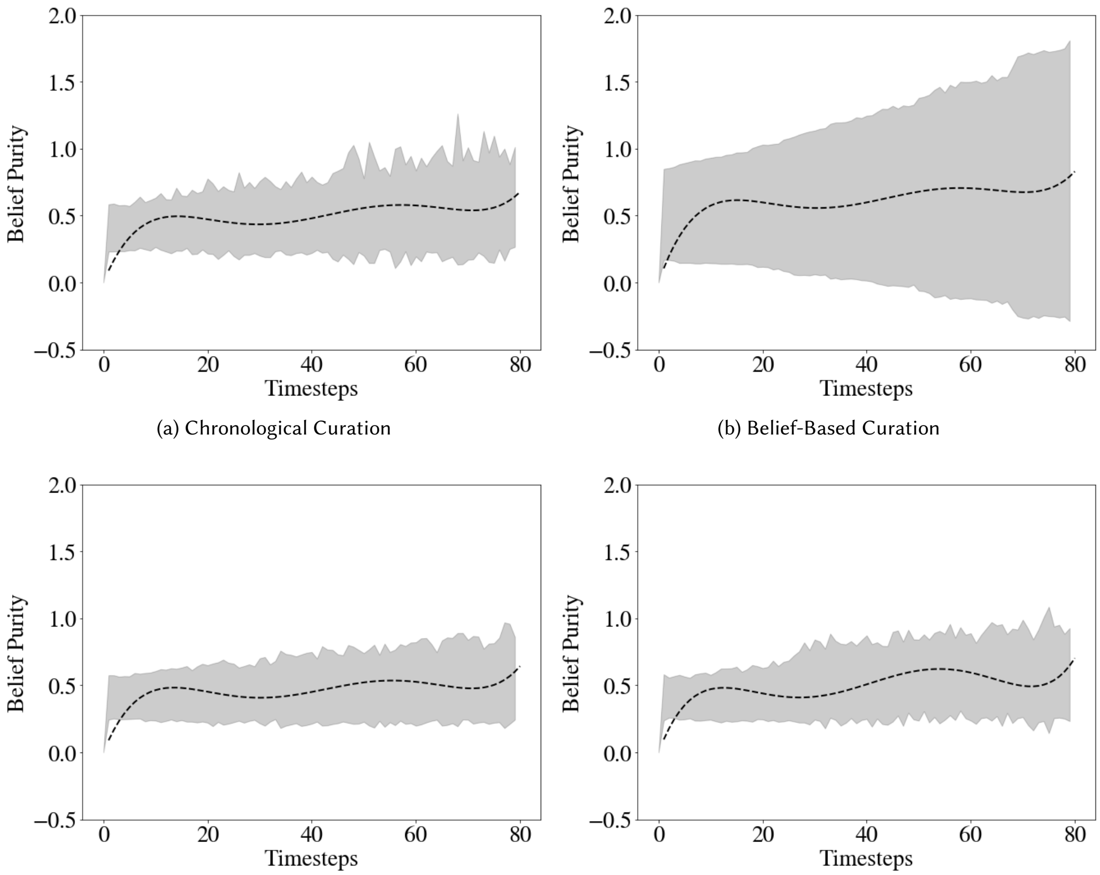
(c) Random Curation

(d) Popularity-Based Curation

Fig. 3. Figures (a)-(d) show the output of the ABM simulation for the false story about Tupac [28].

This presents interesting insights into how varying the objective of the curation algorithm can impact how a tweet spreads through the network. With a belief-based curation objective, the propagation of the tweets increases with every timestep. Whilst with the chronological newsfeed the number of tweets start to decrease after around 10 timesteps. This results in an overall increase of 142% in the number of tweets shared compared to a chronological newsfeed. This same behaviour is seen with the popularity-based curation but with an even steeper increase, it leads to a 604% increase in tweets shared about the story compared to a chronological newsfeed. In contrast, the model with a randomly curated newsfeed reduces the number of posts about the story by 4%.

The results in Figure 5 compare the propagation of a real-world non-false tweet about Justin Bieber [28] (solid line) to the simulated propagation of the tweet from the ABM with different curation objectives (dashed line). Similarly to the propagation of the false tweet, these results show that a belief-based or popularity-based newsfeed curation algorithm cause the number of tweets on the topic to increase every timestep, unlike a chronological newsfeed. The popularity-based curation algorithm leads to 615% more and the belief-based leads to 63% more tweets shared about the



(a) Chronological Curation

(b) Belief-Based Curation

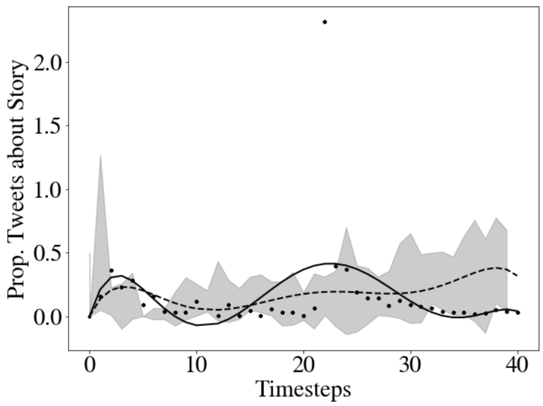
(c) Random Curation

(d) Popularity-Based Curation

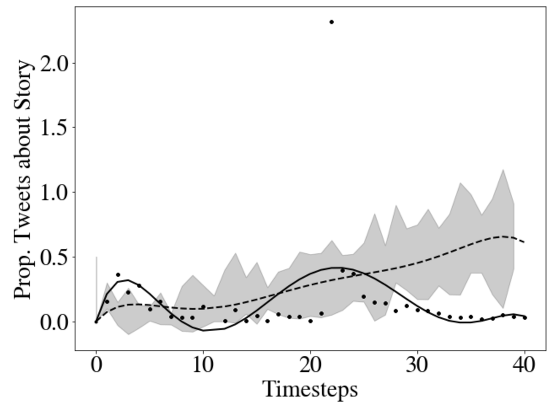
Fig. 4. Figures (a)-(d) show the change in belief purity of an agent's newsfeed for each curation objective for the false story.

story compared to a chronological newsfeed. Again, similarly the results show that the randomly curated newsfeed leads to a decrease of 37% in the number of tweets posted on the topic over the time period.

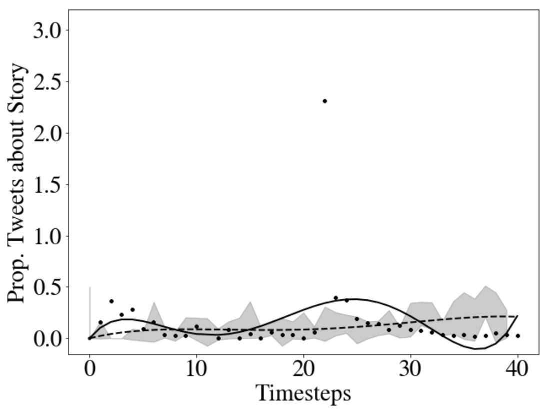
The two sets of results, in Figure 3 and 5, provide some interesting insights into the effects of and potential motivations for different types of curation algorithms. Most social media platforms have decided to move from a chronological newsfeed, also known as a timeline, to a curated newsfeed generated by a recommendation algorithm. This algorithm curates the newsfeed based on a set of objectives which include choosing posts the user will be most interested in (belief-based), timely posts (time-based) and posts that have a high-level of engagement (popularity-based). The results of the model support why social media companies would be incentivised to move to curated newsfeeds, as for both stories, moving to a belief-based or popularity-based curation algorithm results in a significant increase in engagement and tweets. For some news stories, such as the non-rumour about Justin Bieber, this algorithmic amplification is harmless. However, in cases where the news is false and potentially harmful, amplification of the content can be damaging. The results of the randomly curated newsfeed shows that when there is no underlying objective in how



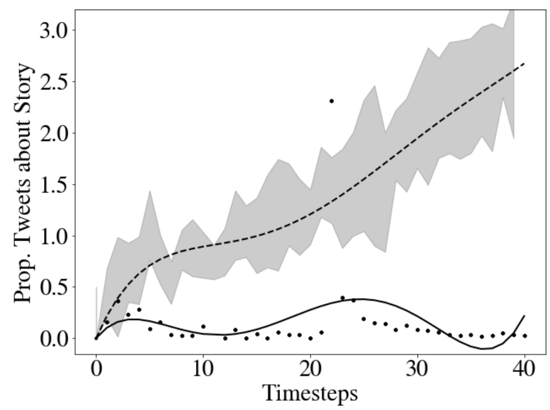
(a) Chronological Curation



(b) Belief-Based Curation



(c) Random Curation



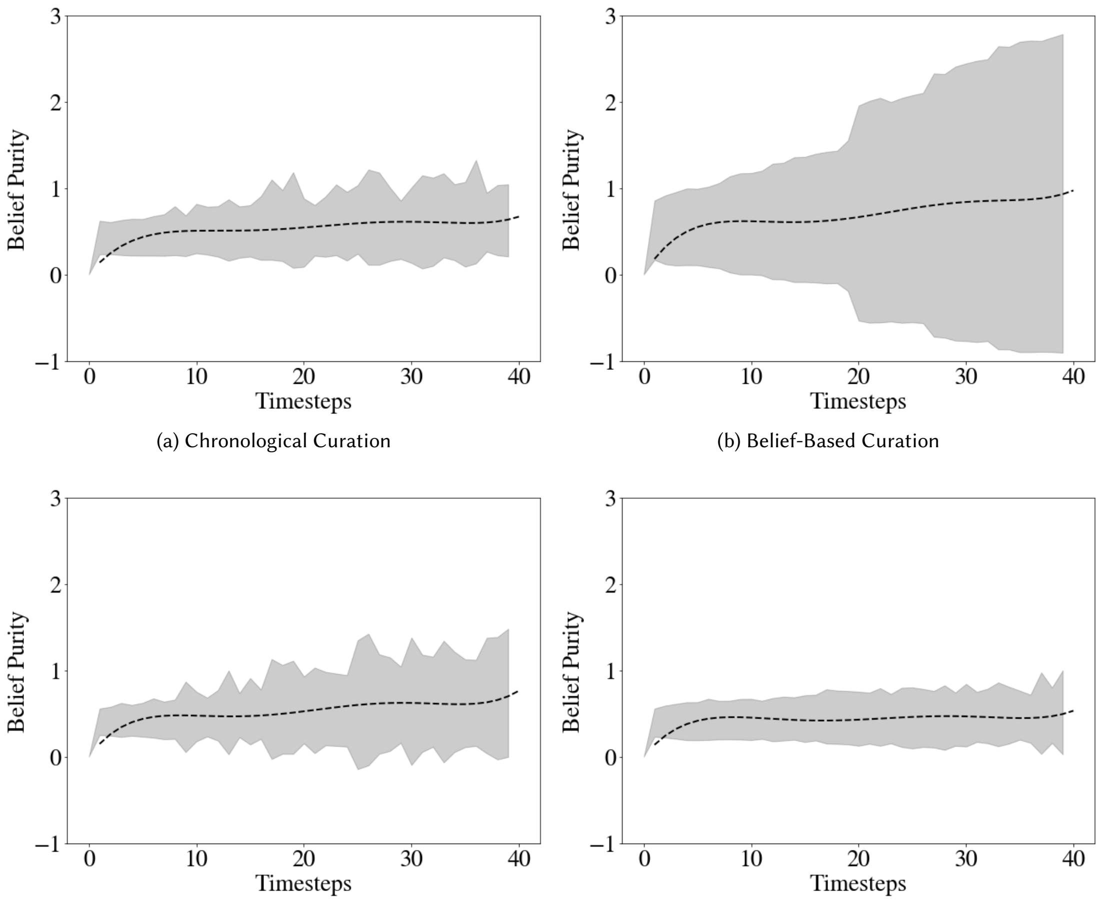
(d) Popularity-Based Curation

Fig. 5. Figures (a)-(d) show the output of the ABM simulation for the non-false story about Bieber [28].

certain tweets are being promoted, there is a reduction in posts about that story. This supports the suggestion that using random dynamical nudges could improve information quality by preventing algorithmic amplification [12]. To conclude, these results show that the chosen objective of the newsfeed curation algorithm significantly impacts how information, and misinformation, spread through the network. Decisions about how newsfeed curation algorithms are encoded should be explicit and balance the motivation for user engagement with considerations of information quality.

2. What impact do curation algorithms have on the formation of echo chambers and polarization on social media?

The second research question aims to understand the impact that different curation algorithm objectives have on the formation of echo chambers and polarization on a social network. This analysis focuses on the belief purity of the tweets that users see on their newsfeed. Belief purity of users' newsfeed is used as a measure to quantify potential for



(a) Chronological Curation

(b) Belief-Based Curation

(c) Random Curation

(d) Popularity-Based Curation

Fig. 6. Figures (a)-(d) show the change in belief purity of an agent's newsfeed for each curation objective for the non-false story.

echo chamber formation and polarization. The equation for belief purity can be found in Section 4. Figures 4 and 6 plot the mean and standard deviation of average user belief purity for each time step across 5 simulations.

The results in Figure 4 and Figure 6 present the impact that newsfeed curation algorithms have on belief purity on the social network. Intuitively, the results show that when users' newsfeeds are curated based on shared-beliefs the belief purity of their newsfeeds increase with time, as seen in Figure 4 (b). The standard deviation of this increase is far greater than for the other three curation algorithms. This shows that if algorithms curate users' newsfeed to show them tweets aligned with their own beliefs, the belief purity of what they see will increase and therefore content diversity will decrease. This has repercussions in terms of echo chamber formation and polarization.

3. What is the relationship between the content diversity of users' newsfeeds and the propagation of misinformation?

Finally, this research aims to identify whether there is a relationship between the lack of content diversity on users' newsfeeds, which can lead to echo chambers and polarization, and the spread of misinformation. This question will be explored by analysing the results of the false story shown in Figure 3 and Figure 4.

In the case of the belief-based curation algorithm, there is a positive correlation between increased belief purity of users' newsfeeds and an increase in the spread of misinformation with time. In comparison to a chronological newsfeed these increases were measured at 142% and 32% respectively. The random curation algorithm results in the same correlation; it causes a reduction of 7% in the belief purity of agents' newsfeeds and a reduction of 4% in the propagation of the misinformation, on average, when compared to the chronological newsfeed. However, this relationship is not identified for the popularity-based curation algorithm. In this case, the algorithm results in a 10% decrease in the mean belief purity of the agents' newsfeed with a 604% increase in the spread of the false tweet. Overall, these findings do not find a consistent relationship between misinformation and content diversity. This indicates that high belief purity of newsfeeds does not directly lead to amplification of misinformation on a network. However, these findings could be different in a model where the story is aligned to a belief. Carrying out this analysis for cases of political or ideological misinformation would be an interesting avenue for future work.

Limitations

Whilst discussing the results, it is worth highlighting the limitations of the proposed model. Firstly, the results and discussion presented in the body of this paper have only been applied to two case-studies. The results presented in Section 5 cannot be generalised to all true and false news stories. Secondly, the population size is not realistic, the simulation was carried out on a population size of $N = 1000$. Additionally, there are limitations in the validation. The proposed model was empirically validated using real Twitter data. This level of validation exceeds most related works, see Section 2, however implementing formal verification would increase confidence in the model output. Finally, the proposed approach has not been used in auditing multi-objective curation algorithms. The ability to combine curation objectives would enable more advanced analysis through a simple extension of the current model. This could be achieved by weighting different curation objectives and identifying the optimum set of weights that achieve these goals. These limitations could be addressed by future work.

Summary

The results of this research highlight the impact of newsfeed curation algorithms on how information spreads on social networks. These results simplify the analysis by showing independent implementations of different curation objectives. In reality, newsfeed curation algorithms are multi-objective. However, presenting these objectives independently provides a clear understanding of how each algorithm impacts both the propagation of information and content diversity. This research highlights that the way in which curation algorithms are encoded has significant impact on information propagation, echo chamber formation and polarization. Whilst social media platforms are motivated by increased user engagement, this should be balanced with promoting information quality and content diversity. By balancing different curation algorithm objectives, both engagement and a healthy social network environment should be achievable.

7 CONCLUSION

Misinformation and online polarization pose some of the greatest societal challenges of the current decade. Understanding how these challenges arise from the way social networks have been designed is critical to tackling them. This paper proposes a novel approach to understanding and comparing the impact of different newsfeed curation algorithms using an ABM of a social network inspired by Twitter. This research highlights how impactful the objective of a newsfeed curation algorithm is on information propagation and diversity. The model is calibrated and empirically validated by one false and one non-false case-study from Twitter [28]. This research shows that ABMs can provide useful insight into how altering the objective of newsfeed curation algorithms can impact the propagation of information, echo chamber formation and polarization on social networks.

The results show that using a popularity-based or belief-based newsfeed curation algorithm leads to significant increases in the number of tweets shared on a topic, which could be detrimental in the case of misinformation. The analysis also shows that belief-based curation algorithms result in increased belief purity in a user's newsfeed which could lead to echo chamber formation and polarization. The results indicate there is a clear incentive for social media platforms to use curation algorithms over chronological newsfeeds as they lead to much higher user engagement and activity. However when these algorithms are being designed there should be consideration of the impact they have on information quality and content diversity.

The proposed model does have a number of limitations, discussed in Section 6, which could be overcome through future work. Firstly, further case-studies could be analysed to ensure the main conclusions presented in this paper, based on two case-studies, persist in additional case-studies. Secondly, future work into computational acceleration of ABMs could allow more realistic and non-trivial populations of agents to be simulated. Additionally, further work into implementing formal verification of the model would increase confidence in the output. Finally, more advanced implementations of newsfeed curation algorithms could be achieved by increasing the complexity of both agent and tweet characteristics. This increased complexity would motivate the use of ABMs for scenario modelling and policy evaluation. This future work could enable this type of modelling to be used in optimising newsfeed curation algorithms to meet the objectives of enhancing information quality or content diversity. Since curation algorithms are becoming so central in shaping beliefs and access to information, we propose that their future design should involve the use of appropriate ABMs to assess their likely impact.

ACKNOWLEDGEMENTS

The first author is supported by a studentship from the UKRI Centre for Doctoral Training in Safe and Trusted Artificial Intelligence (EP/S023356/1). In addition, the support of UK EPSRC (grant number EP/S023356/1, EP/L016796/1 and EP/P010040/1), Intel and Xilinx is gratefully acknowledged.

REFERENCES

[1] Daron Acemoglu and Asuman Ozdaglar. 2011. Opinion dynamics and learning in social networks. *Dynamic Games and Applications* 1, 1 (3 2011), 3–49. https://doi.org/10.1007/s13235-010-0004-1

[2] Muhammad Al Atiqi, Shuang Chang, and Hiroshi Deguchi. 2020. Agent-based approach to resolve the conflicting observations of online echo chamber. In *International Conference on Soft Computing and Intelligent Systems and International Symposium on Advanced Intelligent Systems, SCIS-ISIS*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/SCISISIS50064.2020.9322696

[3] Javier Alvarez-Galvez. 2016. Network Models of Minority Opinion Spreading: Using Agent-Based Modeling to Study Possible Scenarios of Social Contagion. *Social Science Computer Review* 34, 5 (2016), 567–581. https://doi.org/10.1177/0894439315605607

[4] Jack Bandy and Nicholas Diakopoulos. 2021. Curating Quality? How Twitter's Timeline Algorithm Treats Different Types of News. *Social Media and Society* 7, 3 (2021). https://doi.org/10.1177/20563051211041648

[5] Jack Bandy and Nicholas Diakopoulos. 2021. More Accounts, Fewer Links. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (4 2021), 1–28. https://doi.org/10.1145/3449152

[6] Nathan Bartley, Andres Abeliuk, Emilio Ferrara, and Kristina Lerman. 2021. Auditing Algorithmic Bias on Twitter. In *ACM International Conference Proceeding Series*. Association for Computing Machinery, 65–73. https://doi.org/10.1145/3447535.3462491

[7] Engin Bozdag and Jeroen van den Hoven. 2015. Breaking the filter bubble: democracy and design. *Ethics and Information Technology* 17, 4 (12 2015), 249–265. https://doi.org/10.1007/s10676-015-9380-y

[8] Shea Brown, Jovana Davidovic, and Ali Hasan. 2021. The algorithm audit: Scoring the algorithms that score us. https://doi.org/10.1177/2053951720983865

[9] André Calero Valdez and Martina Ziefle. 2018. Human factors in the age of algorithms. Understanding the human-in-the-loop using agent-based modeling. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 10914 LNCS. Springer Verlag, 357–371. https://doi.org/10.1007/978-3-319-91485-5{_}27

[10] Matteo Cinelli, Emanuele Brugnoli, Ana Lucia Schmidt, Fabiana Zollo, Walter Quattrociocchi, and Antonio Scala. 2020. Selective exposure shapes the Facebook news diet. *PLoS ONE* 15, 3 (2020). https://doi.org/10.1371/journal.pone.0229129

[11] Matteo Cinelli, Gianmarco De, Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. [n. d.]. The echo chamber effect on social media. ([n. d.]). https://doi.org/10.1073/pnas.2023301118/-/DCSupplemental.y

[12] Christopher Currin, Sebastian Vallejo Vera, and Ali Khaledi-Nasab. 2021. Depolarization of echo chambers by random dynamical nudge. (1 2021). http://arxiv.org/abs/2101.04079

[13] William H Dutton, Bianca C Reisdorf, Elizabeth Dubois, and Grant Blank. 2017. Social Shaping of the Politics of Internet Search and Networking: Moving Beyond Filter Bubbles, Echo Chambers, and Fake News. (2017). http://dx.doi.org/10.2139/ssrn.2944191

[14] Ivan Dylko, Igor Dolgov, William Hoffman, Nicholas Eckhart, Maria Molina, and Omar Aaziz. 2018. Impact of Customizability Technology on Political Polarization. *Journal of Information Technology and Politics* 15, 1 (1 2018), 19–33. https://doi.org/10.1080/19331681.2017.1354243

[15] Rui Fan, Ke Xu, and Jichang Zhao. 2018. An agent-based model for emotion contagion and competition in online social media. *Physica A: Statistical Mechanics and its Applications* 495 (2018), 245–259. https://doi.org/10.1016/j.physa.2017.12.086

[16] Jan-Philipp Fränken and Toby Pilditch. 2020. Cascades across networks are sufficient for the formation of echo chambers: An agent-based model. (2020). https://doi.org/10.31234/osf.io/8rgkc

[17] Anna Gausen, Wayne Luk, and Ce Guo. 2021. Can We Stop Fake News? Using Agent-Based Modelling to Evaluate Countermeasures for Misinformation on Social Media. In *International AAAI Conference on Web and Social Media (ICWSM)*. https://doi.org/10.36190/2021.63

[18] Natali Helberger, Kari Karppinen, and Lucia D'Acunto. 2018. Exposure diversity as a design principle for recommender systems. *Information Communication and Society* 21, 2 (2 2018), 191–207. https://doi.org/10.1080/1369118X.2016.1271900

[19] Philip Howard. 2016. Is Social Media Killing Democracy? *Oxford Internet Institute* (2016). https://www.oii.ox.ac.uk/blog/is-social-media-killing-democracy/

[20] Philip N Howard, Samantha Bradshaw, Bence Kollanyi, Clementine Desigaud, and Gillian Bolsover. 2017. Junk News and Bots during the French Presidential Election: What Are French Voters Sharing Over Twitter? *Comprop Data Memo* May (2017), 1–5. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/04/What-Are-French-Voters-Sharing-Over-Twitter-v9.pdf

[21] Fang Jin, Edward Dougherty, Parang Saraf, Peng Mi, Yang Cao, and Naren Ramakrishnan. 2013. Epidemiological modeling of news and rumors on Twitter. *Proceedings of the 7th Workshop on Social Network Mining and Analysis, SNA-KDD 2013* (2013). https://doi.org/10.1145/2501025.2501027

[22] Chaitanya Kaligotla, Enver Yucesan, and Stephen E. Chick. 2016. An agent based model of spread of competing rumors through online interactions on social media. In *Proceedings - Winter Simulation Conference*, Vol. 2016-Febru. 3088–3089. https://doi.org/10.1109/WSC.2015.7408413

[23] Nicolas Koumchatzky and Anton Andryeyev. 2017. Using Deep Learning at Scale in Twitter's Timelines. *Twitter* (2017). https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines

[24] Ivan V. Kozitsin and Alexander G. Chkhartishvili. 2020. Users' Activity in Online Social Networks and the Formation of Echo Chambers. In *Proceedings of nternational Conference Management of Large-Scale System Development, MLSD*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/MLSD49919.2020.9247720

[25] Kyandoghere Kyamakya. 2006. Artificial intelligence in Transportation Telematics. *OGAI Journal (Oesterreichische Gesellschaft fuer Artificial Intelligence)* 25, 3 (2006), 2–4.

[26] Francesco Lamperti, Andrea Roventini, and Amir Sani. 2018. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control* 90 (5 2018), 366–389. https://doi.org/10.1016/j.jedc.2018.03.011

[27] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096. https://doi.org/10.1126/science.aao2998

[28] Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Vancouver, Canada, 708–717. https://doi.org/10.18653/v1/P17-1066

[29] Jens Koed Madsen, Richard M. Bailey, and Toby D. Pilditch. 2018. Large networks of rational agents form persistent echo chambers. *Scientific Reports* 8, 1 (12 2018). https://doi.org/10.1038/s41598-018-25558-7

[30] Mozilla. 2021. *YouTube Regrets A crowdsourced investigation into YouTube's recommendation algorithm.* Technical Report.

[31] Salman Bin Naeem, Rubina Bhatti, and Aqsa Khan. 2020. An exploration of how fake news is taking over social media and putting public health at risk. *Health Information and Libraries Journal* (2020). https://doi.org/10.1111/hir.12320

[32] Marily Nika, Gergana Ivanova, and William J Knottenbelt. 2013. *On Celebrity, Epidemiology and the Internet.*

[33] Agnieszka Onuchowska and Donald J. Berndt. 2020. Using agent-based modelling to address malicious behavior on social media. *International Conference on Information Systems, ICIS* (2020).

[34] Professor Vaira Vīķe-Freiberga, Professor Herta Däubler-Gmelin, Ben Hammersley, and Professor Luís Miguel Poiares Pessoa Maduro. 2013. *Final Report of the High Level Group on Media Freedom and Pluralism.* Technical Report. High Level Group on Media Freedom and Pluralism. https://ec.europa.eu/information_society/media_taskforce/doc/pluralism/hlg/hlg_final_report.pdf

[35] Jon Roozenbeek and Sander Van Der Linden. 2021. *Inoculation Theory and Misinformation.* Technical Report. NATO Strategic Communications Centre of Excellence. www.stratcomcoe.org

[36] Arun V. Sathanur, Miao Sui, and Vikram Jandhyala. 2015. Assessing strategies for controlling viral rumor propagation on social media - A simulation approach. *2015 IEEE International Symposium on Technologies for Homeland Security, HST 2015* (2015), 1–6. https://doi.org/10.1109/THS.2015.7225278

[37] Emilio Serrano and Carlos A. Iglesias. 2016. Validating viral marketing strategies in Twitter via agent-based social simulation. *Expert Systems with Applications* 50 (2016), 140–150. https://doi.org/10.1016/j.eswa.2015.12.021

[38] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2018. FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. *arXiv preprint arXiv:1809.01286* (2018).

[39] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.

[40] Kai Shu, Suhang Wang, and Huan Liu. 2017. Exploiting Tri-Relationship for Fake News Detection. *arXiv preprint arXiv:1712.07709* (2017).

[41] Statista Research Department. 2021. Number of daily active Facebook users worldwide as of 3rd quarter 2021. https://www.statista.com/statistics/346167/facebook-global-dau/

[42] Stefani Mans. 2016. In The Filter Bubble: How algorithms customize our access to information. *Masters of Media* (2016). http://mastersofmedia.hum.uva.nl/blog/2016/11/13/in-the-filter-bubble/

[43] Mingsheng Tang, Xinjun Mao, Shuqiang Yang, and Haibin Zhu. 2014. Policy evaluation and analysis of choosing whom to tweet information on social media. In *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, Vol. 2014-Janua. 45–50. https://doi.org/10.1109/SMC.2014.6973882

[44] Twitter. 2021. Twitter Moments guidelines and principles. https://help.twitter.com/en/rules-and-policies/twitter-moments-guidelines-and-principles

[45] Sander van der Linden. 2022. Misinformation: susceptibility, spread, and interventions to immunize the public. *Nature Medicine* (3 2022). https://doi.org/10.1038/s41591-022-01713-6

[46] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151. https://doi.org/10.1126/science.aap9559

[47] Soon-Hyung Yook, Hawoong Jeong, and Albert-László Barabási. 2002. Modeling the Internet's Large-Scale Topology. *Proceedings of the National Academy of Sciences* 99 (2002), 13382–13386. http://dx.doi.org/10.1073/pnas.172501399

8 APPENDICES

A SENSITIVITY ANALYSIS

This paper is mainly focused on the conceptual merits of the approach. Therefore detailed implementations and sensitivity analyses are outside of the scope of this paper and will provide interesting avenues for future work. However, the choice of number of simulations is important to consider to ensure the robustness and validity of results whilst balancing computational expense. For robust results, the number of simulations must be high enough to reach convergence. However, due to the computational expense of running each simulation, the number should not be unnecessarily high.

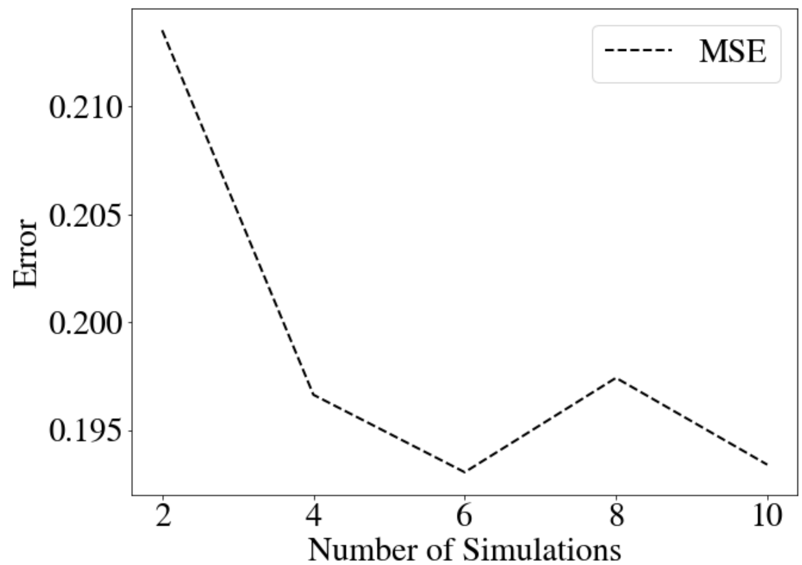


Fig. 7. Figure plotting model MSE against number of simulations N_{sim} for the false case-study.

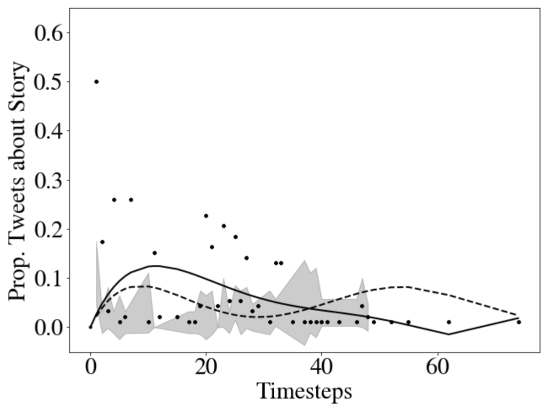
In order to identify a suitable number of simulations for the results in this paper, a coarse sensitivity analysis was carried out to find the number of simulations that minimise the computational expense whilst reaching convergence. For the false case-study, the model was run three times for N_{sims} simulations, which varied between 2 and 10. Then the average mean-square error MSE of the output of the simulations was plotted. Figure 7 presents the results of this analysis. This shows that running above 4 simulations leads to a significant decrease in the MSE of the model. For experiments presented in this paper, a value of $N_{sim} = 5$ was used as this provides a suitable MSE whilst minimising computational expense. A more detailed sensitivity analysis of N_{sim} and other important parameters, such as number of agents, should be presented in future work focused on implementation details.

B ADDITIONAL CASE-STUDY

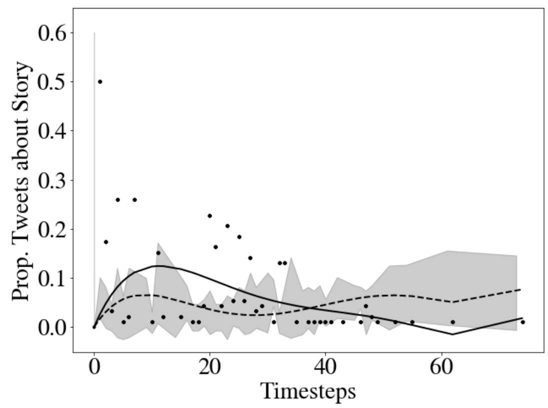
An additional case-study is provided for comparison. The case-study chosen is a false tweet from [28], "police chief helric fredou, one of the police officers investigating the case #charliehebdo, commits suicide". This case-study was generated with the same methodology outlined in Section 4. The probabilities outlined in Table 6 were calibrated using the method described in Section 4.3.

Table 6. Calibrated Parameter Values: the calibrated mean of the probabilities that are sampled during the ABM for the additional case-study

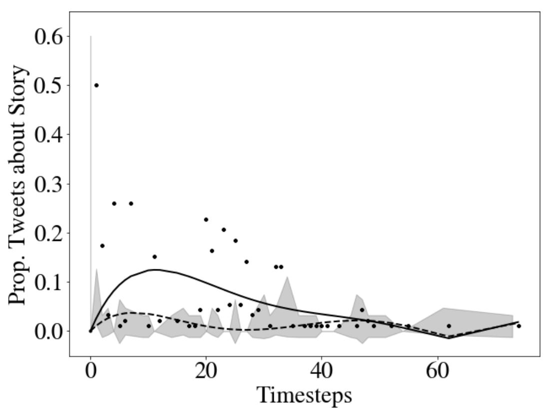
Probabilities	**Additional Case-Study**
$P_{reshare}$	0.02094727
P_{reject}	0.00696533
P_{online}	0.10001950



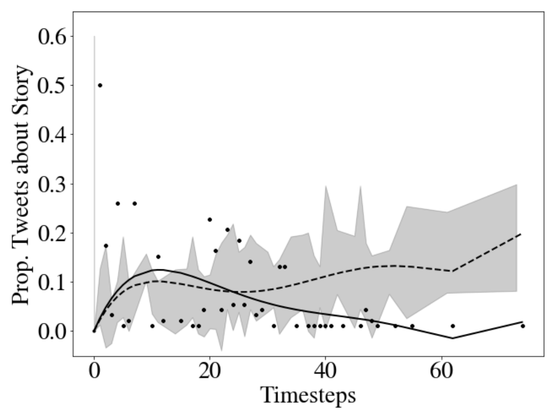
(a) Chronological Curation



(b) Belief-Based Curation



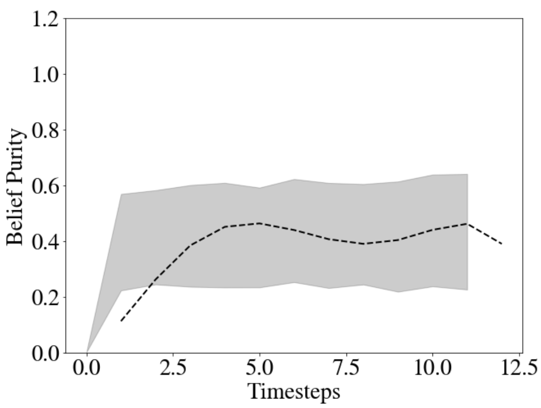
(c) Random Curation



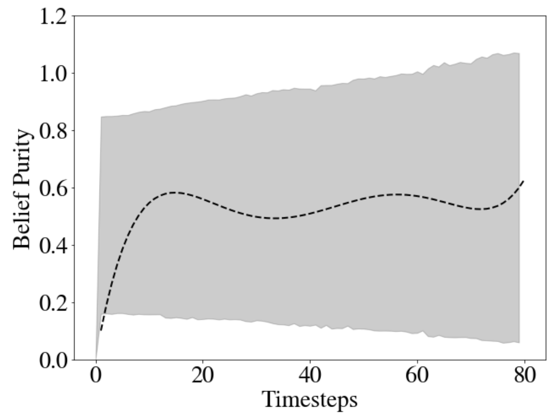
(d) Popularity-Based Curation

Fig. 8. Figures (a)-(d) show the output of the ABM simulation for the additional case-study [28].

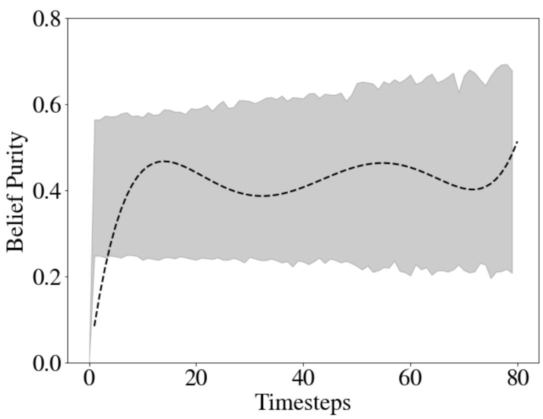
Running the ABM for this case-study generated the results outlined in Table 7. These results show agreement with the case-studies presented in the body of the paper, see Section 5. Figure 8 shows that both a popularity and a belief-based newsfeed curation objective result in an increase in the average number of tweets shared on a story, compared to a chronological newsfeed. It also shows that the random newsfeed curation algorithm results in a decrease in this metric. Figure 9 show that a belief-based algorithm leads to the greatest, 94%, increase in belief purity over the simulation which is a proxy for echo chamber formation.



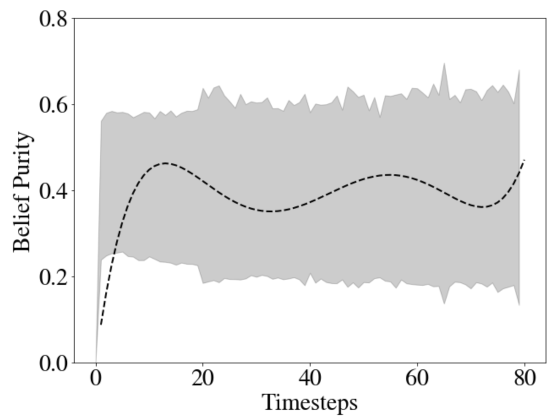
(a) Chronological Curation



(b) Belief-Based Curation



(c) Random Curation



(d) Popularity-Based Curation

Fig. 9. Figures (a)-(d) show the change in belief purity of an agent's newsfeed for each curation objective for the false story.

Table 7. Results for Additional Case-Study

Curation Objective	**Change in** Φ_{avg}	**Change in** Φ_{max}	**Change in Belief Purity**
Time	0.00	0.00	0.00
Popularity	1.23	0.00	0.31
Beliefs	0.52	0.00	0.94
Random	-0.43	0.00	0.47