Algorithmic Bias in Echo Chamber Formation

Claudio Moroni

Table of contents

1 Goal

The research question is to assess algorithmic impact on Twitter echo chamber formation. In order to minimize biases, we observe the same user pool debating over the same topic (US Elections) before and after the introduction of the recommendation algorithm in 2014. Therefore, exploiting a complete dataset of 2012 tweets relative to 2012 U.S. elections, we extract a sample of users to initialize, calibrate and validate an agent based model (ABM) to describe a Twitter free of algorithms. The follow, retweet, favorite, mention (including replies) and hashtag emission of the same pool of users is being monitored in order to get the data needed to later model today's algorithm-biased Twitter echo chamber dynamics and compare it with what would predict the validated algorithm-free model.

2 Literature Review

Twitter can be considered a "media for breaking news in a manner close to omnipresent CCTV for collective intelligence" [1] since 2010, and it didn't but grow in size and information volume since then, being able to count on around 450 million monthly active users today versus around 40 million in 2010. From the beginning, Twitter served as an information sharing platform at least as much as a social network, given its low levels of follow reciprocity and somewhat comparable performance in freshness of "trends" to Google keywords or CNN headlines [1]. It would then be useful to characterize information flow on Twitter, and, particularly, what shapes it and in what measure, with particular regard to how Twitter's recommender systems impact the public debate and the formation of so-called "echo-chambers".

In fact, information flow is an essential aspect of modern democracies, and misinformation on social media is a complex yet widespread phenomenon where individual psychological traits, automated bots led by partisan organisations and populist political parties, confirmation bias and limited attention span combine to make it very hard to counter [2]. Yet, misinformation automation techniques have been widely exploited, contributing to the diminishment of trust towards public institutions via an inconcluding, polarized, partisan and disinformed debate. Debunking techniques have proven inferior w.r.t. impartial institutional communication in neutralizing these effects [2].

Moreover, we know that the hierarchy of news presentation is a moral question, a matter of normative importance because of its shaping influence on public attention, and news prioritization has historically been associated with human deliberations, not machine determinations [3]. It is particularly feared that recommender systems may enhance visibility of just an handful of sources while, at the same time, suggesting to users only pieces of information that agree with their view, in order to increase engagement [3]. Although [3] rules out the possibility that the Google News algorithm had a siloing effect on news recommendations (actually the importance of legacy newspapers born before the digital age is preserved by Google news), and that users, regardless of political leaning, were fed with homophilic articles potentially leading to algorithm-induced echo-chambers (see below), similar results have not yet been proven on Twitter.

The following two sections review the literature about echo-chambers and recommender systems in the context of our work.

2.1 Echo Chambers

An interesting and quite controversial topic about information flow is that of "echo chambers": intuitively speaking, they are social environments in which the opinion, political leaning, or belief of users about a topic gets reinforced due to repeated interactions with peers or sources having similar tendencies and attitudes [4]. Through confirmation bias, echo chambers may serve as "resonance chambers" where people's beliefs are reinforced via repeated exposure to like-mined individuals rather than engaging in healthy discussions with users of opposing views. Many studies have tried to address the echo chamber phenomenon on social media.

From a data analysis point of view, [4] considers four social networks (Twitter, Facebook, Reddit and Gab) and tries to characterize and compare the information flow between their users. The *opinion* (or *ideal point*) of user i over a topic is defined as a real number $x_i \in [-1, 1]$, and an operational definition of echo chamber is given, based on the visual features of the correlograms depicting an user's opinion versus that of its neighbors. Neighborhoods are calculated on "interaction networks" which are constructed with different procedures depending on the social medium in question (Twitter, Reddit and Gab: links to news outlets, Facebook: likes to posts of pages), and the opinion of each user is also estimated in non-comparable ways (Twitter: follower network, Facebook and Gab: co-reply network, Reddit: reply network). Indeed, [4] makes four questionable design choices:

1. No explicit definition of echo chamber is used (echo chambers are just "pointed out" as visual features of the correlograms);
2. The "interaction network" do not take into account the actual information flow (i.e. they are not a substitute for the actual feed of the users) and could be biased (depending on the case) towards capturing only homophilic or heterophilic interactions;
3. Performing a weighted average on tweets containing links to news outlets may be an ineffective way to estimate ideal points: in fact, the content of the tweet could be ironic;
4. Opinion estimation and interaction network retrieval is performed in non-comparable ways across the four platforms.

An intermediate task of this research effort will be to overcome these issues by designing a more robust approach.

From a probabilistic modeling perspective, [5] proves that, before taking into account some biases, Twitter's follower network is a good predictor of ideal points, both of politicians and general population. The author selects a group of US politicians and extracts their Twitter follower network during the 2012 US elections. Then, a simple probabilistic model is developed whose latent parameters include the ideal points of all the politicians and the users extracted. After addressing identifiability issues, the model is calibrated via the use of MCMC algorithms w.r.t. the follower network thus obtaining an estimate for all ideal points, both of politicians and general population. Successive validations and visual comparisons of outputs with existing literature show that the estimate of such ideal points is indeed credible. Finally, the author considers the retweet network over a certain period of time, concerning tweets that either contain the work "Obama" or "Romney". From the visual segregation of the binned correlogram between the ideal points of users that post the tweets and the users that retweet such tweets, the author infers that Twitter is polarized. Again, we raise similar issues with this article's approach:

1. No explicit definition of echo chamber is used (echo chambers are just "pointed out" as visual features of the correlograms);
2. The interaction network considered (the retweet network) is *by construction* an homophilic network: very few people would retweet a post which they don't agree with.

Generally, the phenomena deemed responsible for the formation of echo chambers are *Confirmation Bias* (i.e. the tendency of each individual to consider only information that is in agreement with own beliefs) and *Social Influence* (i.e. the degree to which one's opinion is affected by that of another). Taking a explicit modeling approach, [6] defines three variations of the Bounded Confidence Model (BCM) [7], and compares their phase space w.r.t. the asymptotic number of peaks of the opinion distribution. Briefly, the BCM is a model of opinion dynamics where N agents are described by N opinions $x_i \in [-1, 1]$ on a binary topic are distributed over a graph of N vertices. Two more parameters ϵ and μ (both in $[0, 1]$), respectively model Confirmation Bias and Social Influence. At each iteration of the model, every user's opinion is influenced by that of its neighbors. So considering the i-th user, for every user j within the neighborhood of i, $N_G(i), if $|x_i - x_j| < \epsilon$ we modify the respective opinions according to:

$$\begin{cases} x_i = x_i + \mu(x_j - x_i) \\ x_j = x_j + \mu(x_i - x_j) \end{cases} \tag{1}$$

The "Rewire" variants apply a rewiring step before the BCM step described above, where if $|x_i - x_j| > \epsilon$ the link between i and j is severed and i is randomly rewired to another user. The "Unbounded" variants modify the above system so that "repulsion" is also modelled (the RUCM applies the repulsion step before the rewiring step). We refer the interested reader to [6] for more information. The key findings of this article are:

1. The three new models are able to predict the asymptotic co-existence of two opinions, as opposed to the BCM;
2. The three new models are sensitive to μ in asymptotic number of opinions.

The opinion dynamic sub-model of the ABM that will be later described will be grid-searched among the aforementioned variants.

Another approach to explicit modeling is undertaken in [8], where an opinion dynamics model is defined that also models the user activity a_i sampled from a power distribution $a^{-\gamma}$, the social influence K, the homophily β and the topic controversialness α. Also, differently from [6], the interaction dynamics is

$$\frac{dx_i}{dt} = -x_i + K \sum_{j}^{N} A_{ij} \tanh(\alpha x_j) \tag{2}$$

Where A_{ij} is the interaction network (resulting from random activation of user based on their activity and the probability of users i and j of interacting is $p_{ij} = \frac{|x_i - x_j|^{-\beta}}{\sum_j |x_i - x_j|^{-\beta}}$). It is to be noted that, differently from most models of [6], the latter equation implies a convergence of opinions when two dissenting users interact. The model is shown to replicate the characteristic U-shaped relation between user opinion and activity, and to provide some degree of segregation in correlograms between the opinion of the users and that of their neighborhood. Just as in [4] (that we classified as data analysis paper) and in [5] (that we characterized as being representative of probabilistic modeling approaches) also this article:

1. Does not explicitly define an echo chamber;
2. Models an *a priori* homophilic interaction network;
3. The data used estimate ideal points via links to news outlets, just as [4] does;

In conclusion, all the analyzed papers do not explicitly model echo chambers, measure/define interaction networks in ways that are more ore less explicitly *a priori* biased towards homophily and use disputable means to perform ideal point/opinion estimation.

2.2 Recommender Systems

Recommender systems (or recommendation systems or recommender algorithms) are a subclass of information filtering systems that provide suggestions for items that are most pertinent to a particular user (see Wikipedia).

These are an essential components of modern social media, and in particular of Twitter, as it should be evident from the multiple RecSys challenges that were centered on Twitter (RecSys2020, RecSys2021) . Recommender systems come in various formats and with various tasks: [9] develops an early taxonomy of recommender systems on Twitter, explaining how they can be used to suggest followees, followers, hashtags, tweets, retweets and news. Recommender systems are first classified in *personalized* and *non-personalized*, and further stratified in being examples of *collaborative filtering* or being *content-based*. It is already argued that the most modern recommender systems should probably be endowed with features from both paradigms.

For each recommendation task, the authors compile a quick summary of every system invented so far.

[10] studies the effect of simple "tweet" recommender systems on networks with different values of degree heterogeneity, clustering coefficient and small-world features. A pool of user (each occupying a node in such networks) is simulated to have a debate on a binary topic (where opinions may be A or B). At each time step, every user has a probability to activate and perform two operations:

1. Broadcast its opinion to its neighbors by posting in their reading lists;
2. Update its opinion by reading its timeline (a subset of Q elements the reading list organized by the recommender systems) and acquiring the opinion of the majority of those who posted.

The investigated networks are:

1. Configuration model with power law distribution (heterogeneous degrees, small clustering coefficient, small average path length, briefly named "CM");
2. Watts-Strogatz model with $k = 6$ and $p = 0$ (homogeneous degrees, large clustering coefficient, large average path length, briefly named "WS0");
3. Watts-Strogatz model with $k = 6$ and $p = 0.01$ (homogeneous degrees, large clustering coefficient, small average path length, briefly named "WS001");
4. Watts-Strogatz model with $k = 6$ and $p = 1$ (homogeneous degrees, small clustering coefficient, small average path length, briefly named "WS1");
5. Regular lattice with periodic boundary conditions and $k = 4$ (homogeneous degrees, zero clustering coefficient, large average path length, briefly named "LA");

While the investigated recommender systems are:

1. Random selection of Q elements of the reading list ("reference method" REF);
2. Selection of the first Q elements of the reading list ("recent method" REC);
3. Selection of the oldest Q elements of the reading list ("oldest method" OLD);
4. Selection of Q elements of the reading list so that all elements that express the same view as the user are picked first ("preference method" PR)

After simulating the system for 50 steps, and averaging over $\tau = 10$ instantiations of networks that have a stochastic component, it is found that:

1. Given a balanced initial opinion distribution where half of the user believe A ($P_A = 0.5$) and half in B ($P_B = 0.5$) no algorithm can modify this balance on whatever network. Anyway, on WS0 and WS001 (large clustering coefficient i.e . large topological correlations) the PR method is capable of increasing the fraction of neighborhoods with the majority of user either believing A or B. A similar effect of similar size is not observed on the other networks and/or using other algorithms;

2. Given an imbalanced initial opinion distribution where $P_A = 0.2$ and $P_B = 0.8$, only the PR method is capable of further reducing P_A. On CM and WS1 (small clustering coefficient and small average path length) the PR algorithm reduces the fraction of neighborhoods where the majority of users believe A, in favor of those that believe B, while the remainder of the picture is similar to what is observed in point 1. ;
3. If an opinion, say A, is forcefully introduced in the reading lists of users in a capacity that is a fraction z of all posts, the system will converge to all users believing in A, no matter the recommender system and the underlying network. It is observed that time taken to convergence depends on the sorting algorithm and on z, where OLD is found to be the slowest to converge, while high for high values of z the dependence of the time to converge T^* on the recommender system and on z itself becomes weak.

It should be noted that, again, an explicit definition of echo chambers is not given, as the authors rely on "reader intuition" when looking at specific plots. Nonetheless, a potential influence of recommender algorithms on echo chambers (yet to be defined) cannot be ruled out.

It can be said that Twitter moved away from a purely chronological timeline (i.e. the set of tweets showed to a user) in 2014 (Wikipedia). Unfortunately, little information is available regarding current Twitter's tweet recommender system, while more information is available about Twitter's user recommender system in the form of [11]. The algorithm proposed in [11] should have been up at least until 2017 [12] (to the best of our knowledge, since it or part of it may still be active today).

As [13] notes, little effort has been produced to construct and test new tweet recommender systems, and we add that even less research is so far concerned with learning as many features as possible of the existing Twitter's tweet recommender system, although the need for such inquiries is known since a few years [3]. Particularly, no one has ever tried to quantify the impact of Twitter's suite of recommender systems onto echo chambers dynamics. In this work, we aim to fill this gap by introducing an agent based model (ABM) on Twitter's multilayer graph, upon which a parametric algorithm will act. Each agent will represent a user (in a pool of around 10^4) that took part in both the debate about 2012 US presidential elections and 2020 US presidential elections. Intuitively, we aim to fit the ABM using data from a period where no algorithm was present (2012), and then, keeping the ABM's parameters fixed, we will fit a parametric recommender system upon the ABM w.r.t. a period where Twitter's algorithms were present (2020). Finally, differences in echo chamber dynamics wil be assessed.

The ABM's purpose is to model the (collective) social behaviour (follow, retweet, like, mention, tweet) and opinion dynamics about 2012 and 2020 US presidential election of a group of user in such a way that algorithmic influences are filtered out. To this end, we downloaded as many tweets as possible from 2012 and the retweet and mention networks (which are temporal networks) from those 10'000 users dating back to 2012 (when no sophisticated algorithm was present), and intend to use these data to fit and validate the ABM. Later, we downloaded the retweet and mention networks from those 10'000 users dating back to 2020 (when more sophisticated algorithms were present, for a period that includes the presidential debates and elections), and this time we also extracted temporal information about follow and like activity (which is not possible when using historical API). Knowing the whole temporal activity during the elections will enable us to fit a parametric recommender system by imposing the measured 2020 activity onto the ABM (with the user parameters as calibrated from 2012) and calibrating the parameters of the parametric algorithm that best match the two.

By giving an an explicit definition of echo chamber, based on thresholding some clustering metrics on the multi-layer graph together with a measure of the statistical properties of the ideal points within those clusters, we are able to assess whether Twitter's suite of recommender systems may indeed have a measurable effect on echo chamber dynamics.

In the remainder, we will further detail the experimental setup, the data collected and the developed software.

3 Data and Data Wrangling

The data retrieval and processing is outlined below, starting with 2012 data and then presenting 2020 data. Both sets of data begin with the first presidential debate and end with the last one.

3.0.1 2012 Data

1. We downloaded all tweets from the two sources:

- "Twitter Stream Grab" from ArchiveTeam 2012 (link);
- Microsoft's 38M tweets dating back to specifically the 2012 election (link), from July 1, 2012 through November 7, 2012.

Please note that the "Twitter Stream Grab" accounts for roughly one one-hundredth of all tweets produced in 2012 [14], while Microsoft's data comprises al tweets that include the hashtags "Romney" or "Obama" produced between August 1, 2012 and November 6, 2012, inclusively [15].

1. From those tweets, we extracted the hashtags that were relevant to the 2012 elections: we started by selecting al tweets containing at least one of the hashtags "election2012","vote","voteblue","votered","obama" and "romney". Then considered all the hashtags that were presents in those tweets. From this set, we only kept the hashtags that contained the tokens "democrat","republican","obama","romney","votered" and "voteblue". The initial hashtags and tokens were selected by taking two general hashtags, two specific hashtags and two polarized hashtags. We repeated the process another time and saved all relevant hashtags.

2. We then proceeded to select only those tweets and users that:

    - Used at least one relevant hashtag in 2012;
    - Still existed throughout 2020;

We obtained roughly 160’000 users. From these, we only kept the ones that had at least interacted once (retweet or mention) with another user from the same set. This led to retaining around 10’000 users.

4. We further expanded the set of tweets produced by the 10’000 users by implementing the methodology in [14], retrieving a few more million tweets from those 10’000 users specifically produced during the election period. The number of tweets retrieved with this method might seem large, but it is expected since these users are the ones that not only participated to the debate but also stayed around for longer and are thus more invested and active.
5. All the hashtags retrieved in point 1. (around 4000) were manually labelled as +1 (republican) or -1 (democrat) by reading the tweets they were used in and decide upon their average leaning.

The last step allows us to initialize the modelled users’ opinions.

3.0.2 2020 Data

The follower, retweet, favorite and mention networks were downloaded before the first presidential debate. During the considered period, we ran a sophisticated multithreaded algorithm capable of retrieving all activity of those 10’000 users every 15 minutes. The algorithm made exclusively use of the official Twitter API. Thus, we have a temporal follower, retweet, mention and favorite network spanning the whole period.

4 Methods

We aim to measure (if any) the effects of Twitter’s suite of black box recommender systems on echo chamber dynamics. To this end, we need to model separately the activity the 10’000 users would perform in absence of any sophisticated algorithm and the sophisticated algorithm (suite) itself, so that later we may compare the activity of the group of user with and without the influence of the modelled algorithm and measure the differences in echo chamber dynamics.

The activity of the 10’000 users will be modelled via an ABM, implemented in Agents.jl [16] . We chose to model and measure the same users both in 2012 and 2020 in order to minimize biases due to user selection. The ABM’s base space will be Twitter’s multilayer graph, implemented using MultilayerGraphs.j1 [17]. Parametric algorithms will be taken from literature.

4.1 Echo Chamber Definition

As it is outlined in the Section 2, no explicit definition of "echo-chamber" has been given so far. We propose the following:

Def: Consider a graph G, a measure of node attribute homogeneity $\xi : 2^G \to \mathbb{R}$, a clustering algorithm C and a real number p in the target of C. We define a (C, p)-echo chamber as a subgraph $H \subseteq G$ such that:

1. H is s C-cluster;
2. $\xi(H) \geq p$

In short, this definition combines the notion of topological clustering with that of opinion homogeneity. Results will be tested for robustness w.r.t. C and p.

The features of the simultaneous evolution of the two variables (i.e. $\xi(H)$ and the loss associated to C), and/or an aggregation of them, will be the main quantities of interest for this inquiry.

4.2 Agent Based Model

4.2.1 Base Space and Dynamics

The agent based model has been implemented using Agents.jl [16], a pure Julia framework for agent-based modeling. Being part of the JuliaDynamics ecosystem, Agents.jl was designed with performance, scalability and ease of use in mind. The model aims to incorporate all the essential features of the Twitter community when not influenced by recommender algorithms, while remaining as simple as possible. In order to incorporate all kinds of interactions, the base space of the ABM is a multilayer graph ([18], [19]) implemented using MultilayerGraphs.jl [17] . The latter is a library that bridges the JuliaDynamics and JuliaGraphs ecosystems, thus enabling agent based modelling on (general) multilayer graphs. The effectiveness of the integration is proved in a dedicated script of the test suite. Each node of the multilayer graph represents an user out of the 10'000. The layers are the follow, retweet and like networks. Every user is represented in each layer, and interlayers are diagonal, making it the perfect use case for a `SynchroinizedEdgeColoredDiGraph`. Users update their opinion via a Bounded Confidence Model based on [6]. Static variables shared by all users are:

- ϵ: the "confirmation bias" parameter, that regulates whether an interaction between two users has a (potentially) reinforcing or (potentially) weakening effect on the edges between them (see below for details). It is a real number in $[0, 1]$, although different choices of ϵ may be explored [6];
- μ: the "social influence" parameter, which quantifies the effect of an user's opinion on another one's [6] .

The i^{th} user has the following fields:

- id: unique integer identifier;
- pos_i: the id of the node where the agent sits;
- `feed`: the ordered selection of tweets presented to the user by the suite of recommender systems. It is updated at each time step;
- a_i: the activity of the user i.e. its probability to be activates at each time step;
- x_i: the time-varying opinion of the agent, a real number between -1 (democrat) and $+1$ (republican).
- Tw_i: the probability that a user will tweet its opinion if activated;
- Rt_i: the probability that a user will retweet a tweet in its feed if activated and if its opinion is within ϵ from hers/his;
- Fv_i: the probability that a user will like a tweet in its feed if activated and if its opinion is within ϵ from hers/his;
- Fl_i:the probability that an activated user will follow a user whose tweet is in her/his feed and if its opinion is within ϵ from hers/his;
- Uf_i: the probability that an activated user will unfollow a user whose tweet is in her/his feed if its opinion is further than ϵ from hers/his;

A tweet T has the following properties:

- `id`: an unique integer identifier;
- `user_id`: the `i` field of the user who tweeted this tweet;
- `retweeted_tweet`: the tweet this tweet retweets. If this tweet os not a retweet, this field will be blank;
- `content`: the opinion expressed in this tweet (thus it is a real number within $[-1, 1]$). Same as `T.retweeted_tweet.content` if it is a retweet;
- `favorite_count`: the number of likes this tweet received;
- `retweet_count`: the number of retweets this tweet received.

Time is discrete. At each time step t_i, each user x has a probability to activate given by a_x . If activated, for each tweet T in its feed, the user decides whether the tweet agrees with her/his opinion by evaluating the condition $|x.opinion - T.content| < \epsilon$ [6], and if so, the user updates its opinion according to Equation 1, decides whether or not to retweet (with probability Rt_x), whether or not to like (with like Fv_x) and whether or not to follow the user who wrote it (with probability Fl_x). If instead the condition is not satisfied, the user decides whether or not to unfollow the user who wrote it (with probability Uf_x). All users' update opinion is simultaneous, meaning that users at time t_i update their opinion based solely on tweets produced at times $t < t_i$. This choice enforces consistency and allows for more freedom in the choice of the internal Agent.jl's scheduler, unlocking even faster performance.

To assess modeling impact on results, we will repeat the simulations substituting Equation 1 with Equation 2 , and deciding on whether a tweet's opinion agrees with that of the user who reads it by judging the concordance of the signs of the user's opinion and the derivative calculated via Equation 2 . In this case the static ϵ parameter would be substituted by the controversialness α as modelled in [8] .

At this stage, the user x's feed only contains tweets and retweets from its neighbors, and they are shown according to reversed chronological order. So the "recommender system" is simple and topological.

4.3 Model Initialization

The following sections provide some possible initialization strategies that could benefit calibration. Since the latter should be performed w.r.t. aggregate metrics (to be determined), a perfect initialization is not needed thus the following techniques will be applied in the measure deemed necessary.

4.3.1 2012 Model Initialization

The following model fields and parameters need to be initialized for the 2012 model:

- Follower network;
- Users' activity, tweet rate, retweet rate, favorite rate, follow rate and unfollow rate distributions;
- Users' opinions.

The 2012 data do not provide timestamped follower networks. These will be initialized according to three strategies:

- Random initialization using a directed configuration model (power-law from literature);
- Random initialization using correlations between follow and retweet activity over large user samples and/or derived from literature.
- Initialization using the retweet network up to a certain point in time (to be determined);

Results' robustness will be checked by performing a sensitivity analysis.

The retweet network will be initialized using the collected 2012 data. Users' activity distribution will be modelled according to a power law [8], while tweet, retweet, favorite, follow and unfollow rates will be extracted from a distribution to be looked for in the literature or from the data if they're found to be representative (i.e. we extracted a significant proportion of all tweet using the methodology taken from [14]).

2012 users' opinions have been calculated by averaging the opinions expressed in their tweets. Each tweet has been given an opinion by averaging the leanings associated to each of its hashtags (see Section 3.0.1).

4.3.2 2020 Model Initialization

The following model fields and parameters need to be initialized for the 2020 model:

- Follower network;
- Retweet network;
- Favorite network;
- Users's opinions.

For the 2020 US presidential elections we have downloaded the temporal follow, retweet, mention and favorite activity together with their initial conditions (before the first presidential debate took place), thus we are able to perfectly initialize all networks before calibrating the recommender system. Users' opinions will be initialized with a procedure analogous to that of 2012 (hashtag selection followed by manual hashtag classification).

5 Limitations

The above formulation assumes that the change in users' activity rates from 2012 to 2020 is totally due to the recommender system suite. This may cause undetectable identifiability issues to be measured and addressed in future works. Anyway, since the scope of this effort is to evaluate aggregate quantities (i.e. the differences of the time evolution of (C, p)-echo chamber metrics over time), rather than investigating latent variables of the ABM or the parametric recommender system, this issue shouldn't diminish the value of the results.

References

[1] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th international conference on world wide web - WWW '10*, 2010. doi: 10.1145/1772690.1772751.

[2] M. Cinelli *et al.*, "(Mis) information operations: An integrated perspective," *arXiv preprint arXiv:1912.10795*, 2019, Available: https://arxiv.org/pdf/1912.10795

[3] E. Nechushtai and S. C. Lewis, "What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations," *Computers in Human Behavior*, vol. 90, pp. 298–307, Jan. 2019, doi: 10.1016/j.chb.2018.07.043.

[4] M. Cinelli, G. D. F. Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, "The echo chamber effect on social media," *Proceedings of the National Academy of Sciences*, vol. 118, no. 9, Feb. 2021, doi: 10.1073/pnas.2023301118.

[5] P. Barberá, “Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data,” *Political Analysis*, vol. 23, no. 1, pp. 76–91, 2015, doi: 10.1093/pan/mpu011.

[6] M. D. Vicario, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, “Modeling confirmation bias and polarization,” *Scientific Reports*, vol. 7, no. 1, Jan. 2017, doi: 10.1038/srep40391.

[7] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, “Mixing beliefs among interacting agents,” *Advances in Complex Systems*, vol. 3, no. 01n04, pp. 87–98, Jan. 2000, doi: 10.1142/s0219525900000078.

[8] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, “Modeling echo chambers and polarization dynamics in social networks,” *Physical Review Letters*, vol. 124, no. 4, Jan. 2020, doi: 10.1103/physrevlett.124.048301.

[9] S. M. Kywe, E.-P. Lim, and F. Zhu, “A survey of recommender systems in twitter,” in *Lecture notes in computer science*, Springer Berlin Heidelberg, 2012, pp. 420–433. doi: 10.1007/978-3-642-35386-4_31.

[10] N. Perra and L. E. C. Rocha, “Modelling opinion dynamics in the age of algorithmic personalisation,” *Scientific Reports*, vol. 9, no. 1, May 2019, doi: 10.1038/s41598-019-43830-2.

[11] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, “WTF,” in *Proceedings of the 22nd international conference on world wide web - WWW '13*, 2013. doi: 10.1145/2488388.2488433.

[12] R. H. Nidhi and B. Annappa, “Twitter-user recommender system using tweets: A content-based approach,” in *2017 international conference on computational intelligence in data science(ICCIDS)*, Jun. 2017. doi: 10.1109/iccids.2017.8272631.

[13] L. Belli *et al.*, “Privacy-preserving recommender systems challenge on twitter’s home timeline,” *ArXiv*, vol. abs/2004.13715, 2020.

[14] D. Kergl, R. Roedler, and S. Seeber, “On the endogenesis of twitter’s spritzer and gardenhose sample streams,” in *2014 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2014)*, 2014, pp. 357–364. doi: 10.1109/ASONAM.2014.6921610.

[15] F. Diaz, M. Gamon, J. M. Hofman, E. Kıcıman, and D. Rothschild, “Online and social media data as an imperfect continuous panel survey,” *PLOS ONE*, vol. 11, no. 1, p. e0145406, Jan. 2016, doi: 10.1371/journal.pone.0145406.

[16] A. Vahdati, “Agents.jl: Agent-based modeling framework in julia,” *Journal of Open Source Software*, vol. 4, no. 42, p. 1611, Oct. 2019, doi: 10.21105/joss.01611.

[17] C. Moroni and P. Monticone, “MultilayerGraphs.jl: Multilayer network science in julia,” *Journal of Open Source Software*, vol. 8, no. 83, p. 5116, Mar. 2023, doi: 10.21105/joss.05116.

[18] M. De Domenico *et al.*, “Mathematical formulation of multilayer networks,” *Phys. Rev. X*, vol. 3, p. 041022, Dec. 2013, doi: 10.1103/PhysRevX.3.041022.

[19] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, “Multilayer networks,” *Journal of Complex Networks*, vol. 2, no. 3, pp. 203–271, Jul. 2014, doi: 10.1093/comnet/cnu016.