Alleviating Filter Bubbles and Polarization in News Recommendation via Dynamic Calibration

Han Zhang
Texas A&M University
College Station, Texas, USA
zh89118877@tamu.edu

Ziwei Zhu
George Mason University
Fairfax, Virginia, USA
zzhu20@gmu.edu

James Caverlee
Texas A&M University
College Station, Texas, USA
caverlee@tamu.edu

ABSTRACT

Recent work in news recommendation systems has demonstrated that recommendation algorithms can over-expose users to articles that support pre-existing opinions. Such a filter bubble problem can intensify over time if users and the recommender form a closed feedback loop, eventually resulting in severe political polarization. While empirical work has uncovered this problem in a dynamic recommendation process, how to effectively break this cycle remains elusive. Hence, in this work, we propose a Dynamic Calibration method for new recommendation, which calibrates the recommendations from perspectives of both rankings and predicted scores. Extensive experiments demonstrate the strong performance of the proposed Dynamic Calibration algorithm and also illustrate the effectiveness of the two modules in the proposed method. Data and code can be found at https://github.com/lakers2/recommender_calibration.

1 INTRODUCTION

Recent studies [5, 14] have shown that personalized recommendations can create echo chambers and filter bubbles, where users are primarily exposed to information that aligns with their pre-existing beliefs and opinions. This is particularly concerning in the context of news consumption, where digital news apps increasingly rely on such recommendations to present articles to users [1, 6]. The resulting filter bubble phenomenon in news recommendations [13, 15] contributes to further intellectual and political segregation and polarization. Indeed, a recent work [18] empirically showed that in a dynamic news recommendation process, where users continuously interact with the recommender with a significant interplay of influence between them, a severe filter bubble problem emerges quickly as the recommender evolves with users becoming more extreme in their beliefs over time. These findings suggest that users can easily fall into filter bubbles and become more polarized in the dynamic recommendation process.

While these important studies have demonstrated the problem of filter bubble and political polarization in news recommendations, a solution that addresses this issue remains elusive. There are two key challenges in overcoming this problem:

- **[Challenge 1] how to dynamically eliminate bias in predicted scores from the model?**
- **[Challenge 2] given the debiased predicted scores, how to dynamically maintain the unbiasedness of the ranking lists exposed to users?**

Effectively addressing these challenges is crucial to dynamically alleviate the problem of filter bubble and political polarization in news recommendations. Hence, in this work, we propose the **Dynamic Calibration method (DC)**, which contains two components: the **Dynamic Ranking Calibration (DRC)** module that dynamically manages the unbiasedness of the ranking lists (targeting challenge 2); and the **Dynamic Score Calibration (DSC)** module that dynamically reduces the bias in model predicted scores (targeting challenge 1). We conduct extensive empirical studies to evaluate the performance of the proposed DC method as well as the effectiveness of the two modules. We show that the proposed method can lessen the effects of filter bubbles, leading to less polarized recommendations over time.

2 PRELIMINARIES

In this section, we first introduce the formalization of the dynamic recommendation process, and then, we introduce the Mean Political Stance (MPS) metric to evaluate the degree of polarization among users in the dynamic recommendation process.

2.1 Dynamic Recommendation Process

The dynamic news recommendation process can be summarized into two parts. In the first *Bootstrap* part, we randomly expose b articles to each user and collect their feedback as initial data to train the first version of the recommendation model. Then, in the second part, users visit the system one by one (the same user can visit the system multiple times) , and the recommender recommends k articles to each user using the current version of the model. After every L user visits (what we call one epoch), the recommendation model will be retrained with all feedback data collected up to now.

2.2 Polarization Evaluation

To study filter bubbles and polarization in terms of political stances in news recommendations, we assume that each article has a political stance value, labeled as one of $\{l_1, l_2, \ldots, l_n\}$, which spans the ideological spectrum from extreme liberal (l_1) to extreme conservative (l_n). Likewise, each user can be labeled as one of $\{v_1, v_2, \ldots, v_n\}$, indicating users' political ideology groups and spanning from solid liberal (v_1) to core conservative (v_n). For example, in Section 4, we have news articles with political stance values $\{-2, -1, 0, 1, 2\}$ (spanning the ideological spectrum from extreme liberal (-2) to extreme conservative ($+2$)) and five groups of users (ranging from solid liberals to core conservatives).

Then, to study how the recommendations influence users' news reading behaviors and opinions, we propose the metric Mean Political Stance (MPS) to show how the stance of articles a user reads changes over time. MPS measures that at a certain timestamp t when a user u_t visits the system and is provided by the system

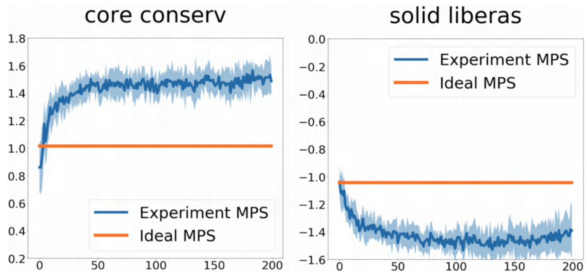

Figure 1: Empirical evidence of the polarization and filter bubble problem.

a ranked list of recommended news articles, what is the average political stance of articles being read by u_t:

$$MPS_t = \frac{\sum_{o=1}^{k} y_{u_t,o} \cdot stance(o)}{\sum_{o=1}^{k} y_{u_t,o}} \tag{1}$$

where we iterate the k recommended articles (from top position $o = 1$ to the end $o = k$), and if u_t clicks and reads the article at position o, $y_{u_t,o} = 1$, otherwise $y_{u_t,o} = 0$. We calculate the average political stance of articles read by the user at interaction t, and $stance(o)$ returns the political stance of an article at position o. We report the average MPS for each user group in each experiment epoch and show how it evolves.

2.3 Empirical Evidence

In [18], during the dynamic recommendation process, the political stance of articles users from different groups read will increasingly deviate from their initial positions, from which can be concluded that users quickly fall into filter bubbles and read increasingly more radical articles as the system evolves. Figure 1 shows when using a conventional matrix factorization model, how the MPS changes over time for two different types of users: a user group with a core conservative ideology and a user group with a solid liberal ideology. The x-axis shows the recommendation epochs, and the y-axis represents the MPS at each recommendation epoch. The 'ideal MPS' is the real stance of each type of user, which is unbiasedly evaluated by the reading history of users during the bootstrap step with random recommendations. This figure illustrates that for these two types of users, the political stance of articles users read will increasingly deviate from their initial positions, quickly falling into filter bubbles of extreme content, indicating the severe filter bubble and polarization problems in the dynamic recommendation process. More details about the experimental settings and results will be introduced in Section 4.

3 INTERVENTION APPROACHES

To address such filter bubble and polarization problems in dynamic recommendations, in this section, we introduce the Dynamic Calibration method (DC). The method is composed of two complementary components: the Dynamic Ranking Calibration module (DRC) and the Dynamic Score Calibration module (DSC).

3.1 Dynamic Ranking Calibration

Conventional recommendation models deliver recommendations with utility maximization as the only goal. Such a conventional method will only recommend content with the same political stance of users and will give up the possibility of exploring other aspects of the user's interests, thus leading to the filter bubble and polarization problem.

Hence, we need to calibrate the ranking lists shown to users to avoid radical recommendations. A straightforward idea is to directly conduct a re-ranking so that the re-ranked list contains a political stance distribution p that is as close as possible to the ideal distribution q. The ideal distribution indicates the real political preferences of users. Specifically, for each user, we calculate the probabilistic distribution of articles from each of the n political stances read by the user as the political stance distribution. In practice, ideal distribution q can be unbiasedly estimated by the feedback data collected during the Bootstrap step (in which articles are randomly exposed to users and feedback data is purely driven by users' interests), and p is calculated by all the feedback data after Bootstrap.

We use KL divergence to measure the difference between p and q, The smaller the KL divergence score is, the more similar the two distributions are. Then, we take this distribution difference with the recommendation utility together into consideration in our re-ranking process. To balance recommendation utility and calibration effect, we trade-off between them in the re-ranking criterion following [17] as:

$$S'_{u,i} = (1-\lambda) \cdot S_{u,i} + \lambda \cdot KL(p_{+i}|q), \tag{2}$$

where $S_{u,i}$ is the predicted preference score from an existing model for the user u and article i; p_{+i} calculates the reading history distribution if i is recommended and read by the user; λ is the trade-off parameter to control the strength of calibration; and $S'_{u,i}$ is the new score of i for re-ranking. During the re-ranking process, starting from an empty list, we iteratively calculate $S'_{u,i}$ and greedily pick the article with the highest score into the recommendation list.

In the dynamic recommendation process, the reading behaviors of users and the recommender influence each other, and both of them evolve over time. Thus, we further propose the dynamic version of this ranking calibration method: we dynamically change the strength of calibration by adjusting the λ in Equation 2 according to the current severeness of filter bubbles. When the current reading history distribution of a user group differs significantly from the ideal distribution, we strengthen calibration by increasing λ, and vice versa. Formally, we set $\lambda = \min(1, 0.5 + \alpha \cdot KL(p|q))$, where larger hyperparameter α indicates stronger strength of calibration. Note that λ in this work is designed to be user group dependent.

3.2 Dynamic Score Calibration

One strong premise of the introduced dynamic ranking calibration method is that the predicted preference score $S_{u,i}$ in Equation 2 is accurate and unbiased. However, the fact is that the recommendation model produces biased predictions and overestimates the scores for items matching the dominant interests of a user [19]. Hence, besides ranking calibration, it is also crucial to directly calibrate the predicted scores to evade overestimation of dominant interests and underestimation of minority interests (such as overestimating the score of a left-leaning article and underestimating the score of a right-leaning article for a left-leaning user).

A straightforward idea is to modify the loss function in the training process so that the learned model focuses more on how to deliver filter-bubble-free and polarization-friendly articles. In the original algorithm, we treat all training samples equally. However, the model will deliver biased results due to the overestimation of the dominant interests of a user. We need to mitigate this overestimation during the training process by assigning a weight for each training sample. If one political stance l is under-recommended, we add a large weight for these articles in the next training round. To minimize the loss, the recommender system will focus more on exposing more articles in stance l. The improved training loss can be formulated as:

$$\begin{aligned} \mathcal{L}' &= \sum_{u,i} w_{u,l(i)} \cdot \mathcal{L}(u,i), \\ w_{u,l(i)} &= (q(u,l(i))/p(u,l(i)))^{\beta}, \end{aligned} \tag{3}$$

where $\mathcal{L}(u,i)$ is the original loss of user-article pair (u,i); $l(i)$ is the political stance of the article i; $q(u,l(i))$ is the value of stance $l(i)$ in the ideal distribution of the user; $p(u,l(i))$ is the value of stance $l(i)$ in the current reading history distribution; and β is the hyper-parameter controlling the strength of calibration. Because $p(u,l(i))$ changes over time for each user group, this score calibration method is a dynamic approach.

When a user reads fewer articles from political stance l than expected indicated by the ideal distribution, the model could underestimate the scores for these articles. In this case, $w_{u,l(i)}$ will be larger than 1, which means that the loss function will push the model to correct its prediction bias towards these user-article pairs, thus addressing the underestimation.

Last, we combine these two approaches together to achieve a comprehensive solution for fighting filter bubbles and polarization – the Dynamic Calibration algorithm (DC).

4 EXPERIMENTS

In this section, we conduct experiments to answer three questions: (1) Can the dynamic calibration (DC) method reduce the effect of filter bubble and polarization? (2) What are the effects of the two components of the DC model? (3) What are the impacts of the two hyper-parameters α and β?

4.1 Experiment Setting

4.1.1 Dataset. We use a variation of the dataset from [13], which consists of a collection of 40,000 news articles and a set of 500 users. The 40,000 articles are with annotations of their topics and political stances. Specifically, there are 14 topics: *abortion, environment, guns, health care, immigration, LGBTQIA, taxes, technology, trade, Trump impeachment, US military, welfare, US 2020 election, and racism.* Each article can cover one or more topics. For political stance, each article is labeled as one of $\{-2,-1,0,1,2\}$, which spans the ideological spectrum from extreme liberal (-2) to extreme conservative $(+2)$. There are 8,000 articles for each political stance. We can use a binary utility matrix $\mathbf{A}_i \in \{0,1\}^{14\times5}$ to represent the topic and stance for

Algorithm 1: Dynamic News Recommendation Process

```
Bootstrap: Randomly expose 10 articles from each topic
  (140 in total) to each user, and collect initial clicks D,
  calculate ideal MPS for each user group, and train the first
  model ψ by D;
for t = 1 : 40,000 do
    Randomly choose a user u_t as the current visiting user;
    Re-rank candidates using dynamic ranking calibration;
    Recommend 5 articles to the current user u_t by ψ;
    Collect new clicks and add them to D;
    Update preference matrix of user u_t;
    if t%200 == 0 then
        Implement dynamic score calibration;
        Retrain ψ by D;
```

an article i. Figure 3(a) shows an example of an article related to abortion and immigration with a political stance of -2.

The user set is simulated based on the Pew survey of U.S. political typologies [7], which summarizes 9 political typologies in the U.S. and their opinions toward different topics. We consider the five most representative typologies: *solid liberal* (extreme liberal), *opportunity democrats* (lean toward liberal), *bystanders* (mild group), *market skeptic republicans* (lean toward conservative), and *core conservatives* (extreme conservative). For each typology, we generate 100 users, where each user can be represented by a preference matrix $\mathbf{U}_u \in \mathbb{R}^{14\times5}$ to represent the user's political stances toward different topics. The larger $\mathbf{U}_u(p,s)$ is, the more likely user u holds an opinion of stance s toward the topic p. Figure 3(b) shows an example preference matrix of a 'solid liberal' user and Figure 3(c) shows an example preference matrix for a 'core conservative' user.

With the utility matrices for news articles and preference matrices of users, we can determine the ground-truth preference of a user for an article by vectorizing their corresponding matrices and then taking the dot product to calculate the preference score between them. We can further determine user-article interaction behaviors by this preference score. The higher the preference score is, the more likely a user is to click and read the article.

4.1.2 Dynamic Recommendation Experiment. The detailed experimental process is presented in Algorithm 1. We first conduct a bootstrap step to collect initial click data from all users by randomly showing 140 articles (10 articles from each topic) and then training the first recommendation model with the initial click data. Then, we run the dynamic experiment for 40,000 iterations. At each iteration, a random user will come and ask for recommendations of 5 articles. The user will iterate all the 5 articles and determine whether click and read them. The interaction data will be stored for further model training. We retrain the model after every 200 iterations, resulting in 200 experiment epochs. In this work, we use the fundamental Matrix Factorization (MF) [13] model as the core approach to deliver recommendations.

Moreover, users' preferences can be influenced by recommendations exposed to them. If an article was recommended and read by a user, the corresponding opinions of the user will be reinforced,

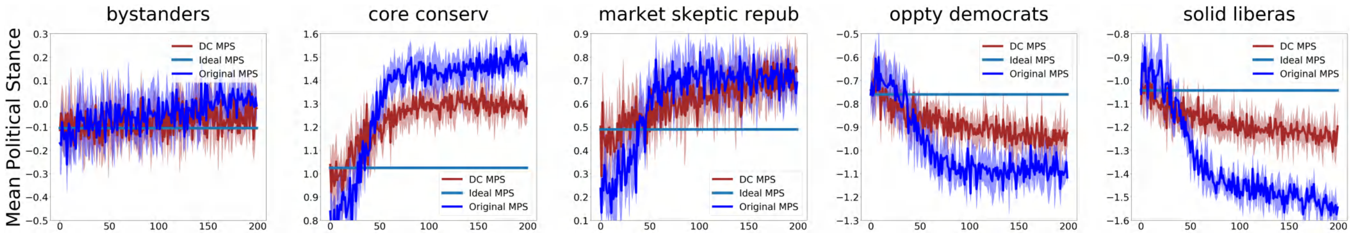

Figure 2: Results comparison with ideal MPS, Combined MPS, and Original MPS.

	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2
abortion	1	0	0	0	0	1.0	0.8	0.6	0.3	0.1	0.1	0.3	0.5	0.7	0.8
LGBTQIA	0	0	0	0	0	0.8	0.5	0.3	0.0	0.0	0.0	0.0	0.3	0.6	0.9
guns	0	0	0	0	0	1.0	0.8	0.5	0.1	0.0	0.0	0.1	0.4	0.6	0.9
⋮	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
immigration	1	0	0	0	0	0.9	0.6	0.3	0.0	0.0	0.0	0.4	0.4	0.6	0.9
			(a)					(b)					(c)		

Figure 3: (a) shows an article matrix. (b) shows a preference matrix for a 'solid liberal' user. (c) shows a preference matrix for a 'core conservative' user.

and the user is more likely to click articles with the same political stances and topics in the future. So, we model these dynamics by changing preference matrices of users corresponding to articles read by users. We first define an influence parameter c to determine to what degree users can be influenced by recommendations. Then, every time a user u is exposed to an article i, if u clicks and reads i, we update the preference matrix $\mathbf{U}_u$ of u by $\mathbf{U}_u \leftarrow \mathbf{U}_u + c \cdot \mathbf{A}_i$. In our experiment, we set $c = 0.03$.

4.2 Performance of Dynamic Calibration

First, we study: can the DC method reduce the effect of filter bubble and polarization? We set the hyperparameters $\alpha = 2$ and $\beta = 8$. The evolution of MPS of different user groups is shown in Figure 2:

- The Ideal MPS is the MPS calculated based on the bootstrap data and can be considered as an unbiased estimation of users' political preferences. The Original MPS is the result of conventional MF without any intervention. We can observe that the Original MPS value deviates from the Ideal MPS as the recommender system evolves, showing significant polarization among users.
- The DC MPS is the result of our proposed DC method, which is closer to the Ideal MPS compared to the Original MPS, depicting that the DC method can successfully alleviate the filter bubble and polarization problem.
- The DC method performs better on the 'solid liberal' group and the 'core conservative' group, indicating that we can produce a better intervention effect for groups with more extreme political stances.

4.3 Ablation Study

In this section, we aim to evaluate the effectiveness of the dynamic ranking calibration (DRC) module and the dynamic score calibration (DSC) module. In this experiment, we set $\alpha = 2$ for DRC and $\beta = 8$ for DSC. We show the MPS evolution on the 'solid liberal' group and the 'core conservative' group in Figure 4 for DRC and

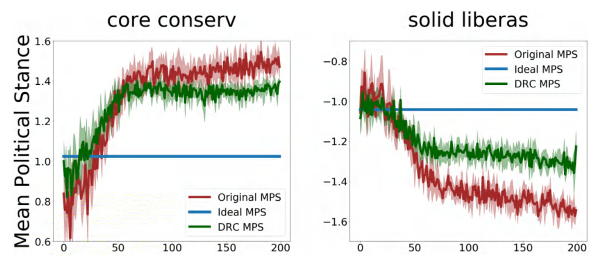

Figure 4: Results comparison among Ideal MPS, Original MPS, and DRC MPS.

Figure 5 for DSC. From these two figures, we can observe that both DRC and DSC can alleviate the filter bubble and polarization problem compared to the Original MPS result by the MF model, demonstrating the effectiveness of the two proposed modules. Moreover, comparing Figure 2 with these two figures, we can see that the DC model combining these two modules outperforms both DRC and DSC in terms of the filter bubble and polarization alleviation. This result illuminates the complementarity of these two modules.

4.4 Hyperparameter Study

Last, we study the impacts of the two hyperparameters that dynamically control the calibration strength of DRC and DSC. The α in the DRC module controls the trade-off between ranking calibration and ranking utility: the larger α the stronger calibration. In Table 1, we show the experimental results for solid liberals and core conservatives user groups of different α with $\beta = 8$, where we list the average Discounted Cumulative Gain (denoted as avg DCG) to show the overall recommendation utility, the higher the better. And we list the absolute difference between the average MPS of the experiment and the Ideal MPS (denoted as $|\Delta_{MPS}|$) to show the overall calibration performance, the lower the better. From the table, we can see that in line with our expectation, larger α leads to less severe filter bubble and polarization but lower utility.

Next, we study the impact of β in the DSC module: the larger β leads to stronger calibration for predicted scores. We show the experimental results of different β with $\alpha = 2$ in Table 2. The results depict that larger β results in less severe filter bubble and polarization but lower recommendation utility.

Hence, based on these experimental observations, we can conclude that both proposed dynamic modules deliver effective intervention performance, and the DC method can further improve the result. However, recommendation utility is sacrificed by these methods for alleviating filter bubbles and polarization.

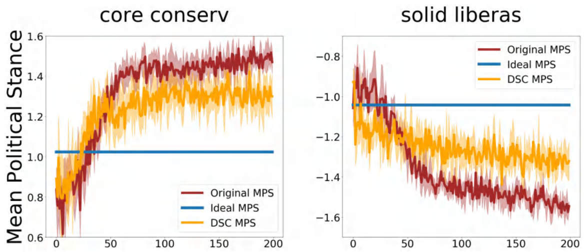

Figure 5: Results comparison among Ideal MPS, Original MPS, and DSC MPS.

Table 1: The influence of α with the $\beta = 8$. $|\Delta_{MPS}|$ calculates the absolute difference between avg MPS of the experiment and the Ideal MPS.

	metric	α=0	α=1	α=2	α=3	α=4
solid liberals	avg DCG	1.628	1.569	1.523	1.494	1.493
	$\|\Delta_{MPS}\|$	0.182	0.152	0.129	0.126	0.120
core conserv	avg DCG	1.705	1.748	1.717	1.715	1.673
	$\|\Delta_{MPS}\|$	0.226	0.230	0.212	0.204	0.187

Table 2: The influence of β with the $\alpha = 2$. $|\Delta_{MPS}|$ calculates the absolute difference between avg MPS of the experiment and the Ideal MPS.

	metric	β=2	β=5	β=8	β=11	β=14
solid liberals	avg DCG	1.621	1.565	1.523	1.518	1.482
	$\|\Delta_{MPS}\|$	0.174	0.165	0.129	0.120	0.103
core conserv	avg DCG	1.767	1.744	1.717	1.696	1.625
	$\|\Delta_{MPS}\|$	0.241	0.241	0.212	0.204	0.161

5 RELATED WORK

Filter bubbles and polarization in personalized recommender systems have been extensively studied across various platforms such as Facebook and YouTube [2, 3, 16]. Grosetti et.al [11] demonstrate the impact of a recommender system on user behavior and how the personalized recommendations amplify filter bubbles, revealing a decrease in recommendation diversity by 30%. The root cause of filter bubbles is the tendency of recommendation algorithms to prioritize content that users are more likely to click on for maximum utility [4, 8, 12]. This problem can negatively impact user experience and exacerbate intellectual segregation and polarization in society [9]. Specifically, a recent study [18] conducted simulation experiments to analyze the dynamic nature of political polarization in news recommendations and found that users become more radical and the polarization problem intensifies over time. While some mitigation strategies [10, 17] for filter bubble in recommendations have been proposed, most of them only focus on short-term and static scenarios, which cannot work effectively in real-world dynamic systems. Therefore, our work aims to fill the gap and address the filter bubble and polarization problem in a dynamic manner.

6 CONCLUSION AND FUTURE WORK

In this work, we aim to study how to effectively intervene in a dynamic news recommendation process – where users and the recommender form a closed feedback loop – to lessen filter bubbles and polarization. We propose a Dynamic Calibration method, which calibrates the recommendations from perspectives of both rankings and predicted scores in a dynamic manner. Extensive experiments show the encouraging performance of the proposed Dynamic Calibration model and also demonstrate the effectiveness of the two modules in the proposed method. In the future, we want to study more advanced approaches based on reinforcement learning algorithms to more effectively alleviate the filter bubble and polarization problems while preserving the recommendation utility.

REFERENCES

[1] Deepak Agarwal, Bee-Chung Chen, and Pradheep Elango. 2009. Explore/exploit schemes for web content optimization. In *2009 Ninth IEEE International Conference on Data Mining*. IEEE, 1–10.
[2] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.
[3] Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science* 26, 10 (2015), 1531–1542.
[4] Wei Chu and Seung-Taek Park. 2009. Personalized recommendation on dynamic content using predictive bilinear models. In *Proceedings of the 18th international conference on World wide web*. 691–700.
[5] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* 118, 9 (2021).
[6] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*. 271–280.
[7] C Doherty, J Kiley, and B Johnson. 2017. Political typology reveals deep fissures on the right and left: Conservative Republican groups divided on immigration,'openness'. *Pew Research Center* (2017).
[8] Susan Dumais, Thorsten Joachims, Krishna Bharat, and Andreas Weigend. 2003. SIGIR 2003 workshop report: implicit measures of user interests and preferences. In *ACM SIGIR Forum*, Vol. 37. ACM New York, NY, USA, 50–54.
[9] Robert Epstein and Ronald E Robertson. 2015. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences* 112, 33 (2015), E4512–E4521.
[10] Fabio Gasparetti, Giuseppe Sansonetti, and Alessandro Micarelli. 2021. Community detection in social recommender systems: a survey. *Applied Intelligence* 51, 6 (2021), 3975–3995.
[11] Quentin Grossetti, Cédric du Mouza, and Nicolas Travers. 2020. Community-based recommendations on Twitter: avoiding the filter bubble. In *International Conference on Web Information Systems Engineering*. Springer, 212–227.
[12] Christopher C Johnson. 2014. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems* 27, 78 (2014), 1–9.
[13] Ping Liu, Karthik Shivaram, Aron Culotta, Matthew A Shapiro, and Mustafa Bilgic. 2021. The Interaction between Political Typology and Filter Bubbles in News Recommendation Algorithms. In *Proceedings of the Web Conference 2021*. 3791–3801.
[14] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. 677–686.
[15] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think.* Penguin.
[16] Matthew A Shapiro and Han Woo Park. 2015. More than entertainment: YouTube and public responses to the science of global warming and climate change. *Social Science Information* 54, 1 (2015), 115–145.
[17] Harald Steck. 2018. Calibrated recommendations. In *Proceedings of the 12th ACM conference on recommender systems*. 154–162.
[18] Han Zhang, Ziwei Zhu, and James Caverlee. 2023. Evolution of Filter Bubbles and Polarization in News Recommendation. https://doi.org/10.48550/ARXIV.2301.10926
[19] Ziwei Zhu, Yun He, Xing Zhao, Yin Zhang, Jianling Wang, and James Caverlee. 2021. Popularity-opportunity bias in collaborative filtering. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 85–93.