Preference Amplification in Federated Recommender Systems

Mark Lacay[1], Carlo Uhl[1,2], Sophia Ludewig[1,3]

[1]University of California, Berkeley
[2]Technical University Berlin
[3]Technical University Munich

mlacay@berkeley.edu, carlo.uhl@berkeley.edu, sophia.ludewig@berkeley.edu

Abstract

Recommender systems have become central to how users discover content, offering highly personalized suggestions that enhance engagement. However, this personalization often narrows user interests, leading to preference amplification and phenomena such as echo chambers, filter bubbles, and polarization. Building on the theoretical framework of preference amplification in recommender systems by Kalimeris et al.[15], we extend it to a federated setting, applying the standard FedAvg algorithm to integrate local user models via a variational autoencoder (VAE-based) recommender trained on MovieLens 1M. We introduce novel divergence metrics—local-global divergence and personalization degree—to quantify global misalignment and preference clustering. Our results show that increasing the number of nodes and aggregation rounds reduces preference amplification, while more local epochs intensify it. These findings suggest that federated learning parameters serve as "dials" to maintain a healthier balance between personalization and diversity, thereby mitigating echo chamber formation.

1 Introduction

Recommender systems play a pivotal role in shaping user interactions with vast online content, spanning e-commerce, streaming platforms, and social media [2]. By analyzing user data—clicks, behaviors, and interactions—RSs deliver personalized recommendations that drive engagement and business outcomes [6, 11]. However, the mechanisms that power these systems often come with significant trade-offs.

A key concern is preference amplification, where feedback loops reinforce user preferences over time, narrowing their content exposure [15]. While this can increase user engagement, it can also lead to echo chambers and filter bubbles, limiting content diversity and exposing users primarily to like-minded perspectives [22, 3]. These dynamics not only reduce the richness of the user experience but also have societal consequences. For instance, during the COVID-19 pandemic, platforms like YouTube and Facebook disproportionately amplified divisive or sensationalist content, accelerating the spread of health-related misinformation and undermining societal trust [23, 5]. Additionally, the bias toward popular content often sidelines niche creators, further reducing diversity in recommendations [8, 12].

Addressing these challenges requires overcoming the difficulty of defining and measuring amplification effects, as well as identifying solutions that balance personalization with societal well-being [4, 20]. Existing mitigation strategies for centralized systems, such as re-ranking and diversity regularization, show promise but leave open questions about how these approaches might translate to decentralized settings.

This paper explores the potential of federated learning (FL) to mitigate these challenges. Unlike centralized systems, FL decentralizes user data by keeping it on individual devices while sharing only aggregated model, thereby enhancing privacy and reducing risks of data breaches [21, 14]. However, localized feedback loops in FL can still propagate biases [26]. By addressing the following research questions, this study seeks to uncover whether FL can balance personalization, privacy, and societal concerns:

- **RQ1 Understanding Amplification:** How does preference amplification develop in federated recommender systems?
- **RQ2 Breaking Feedback Loops:** Can federated learning prevent feedback loops and echo chambers?
- **RQ3 Engagement vs. Generalization:** Are federated systems less prone to overfiting for increased engagement, and can parameters like nodes, rounds, or epochs help moderate polarization?

By answering these questions, we aim to provide actionable insights into designing recommender systems that align user-centric goals with broader societal values.

2 Related Work

2.1 Filter Bubbles and Mitigation Strategies

Filter bubbles, as described by Pariser [22], arise when algorithms personalize content based on user behavior, restricting exposure to diverse perspectives. These effects are closely tied to echo chambers and group polarization, where algorithmic curation reinforces pre-existing preferences and limits engagement with alternative viewpoints [7, 17]. Studies have shown that such phenomena can skew user behavior and contribute to polarization in online ecosystems [3, 10].

Efforts to address these issues focus on balancing personalization with diversity in recommender systems. Some approaches aim to introduce mechanisms that enhance exposure to less popular or unique content, thereby improving diversity without significantly compromising accuracy [12]. Others incorporate diversity constraints into recommendation algorithms, encouraging exposure to varied content categories [16]. Techniques such as re-ranking reorder recommendations to balance relevance with diversity [9], while adaptive frameworks dynamically adjust content exposure to mitigate polarization while maintaining user engagement [24]. However, these strategies focus on centralized systems, leaving gaps in how decentralized approaches like federated learning can address similar concerns.

2.2 Preference Amplification

Kalimeris et al. [15] identified preference amplification as a critical problem in recommender systems, where feedback loops progressively reinforce users' initial preferences. This process drives the formation of echo chambers and filter bubbles, reducing content diversity and restricting exposure to alternative perspectives. Using real-world data from Facebook's video recommender system, they developed a theoretical model illustrating the evolution of user preferences, validated through empirical evidence. Their findings highlight the risk that even content with low initial prevalence can gain prominence in recommendations if users exhibit some initial engagement. To mitigate this, their work suggests that hyperparameter tuning of down-ranking and learning rate terms may moderate preference amplification, albeit potentially at the expense of overall recommender performance.

Building on the theoretical framework of preference amplification in recommender systems [15], we extend this analysis to a federated setting. Applying the standard FedAvg algorithm [18], we integrate local user models through a variational autoencoder (VAE-based) recommender system trained on the MovieLens 1M dataset [1]. This approach introduces two novel divergence metrics—local-global divergence and personalization degree—that quantify global misalignment and preference clustering in federated systems.

Kalimeris et al. [15] laid the groundwork for understanding preference amplification through foundational metrics, which we adapt and expand upon for decentralized contexts:

- **User Preference Vector Norm (ut_norm):** Quantifies the intensity of user preferences over time, capturing how interactions amplify specific behaviors.
- **Likable vs. Non-Likable Probability Separation:** Measures the system's tendency to recommend engaging ("likable") over ignored ("non-likable") content, reinforcing satisfaction at the expense of diversity.
- **Probability of Well-Correlated Items**

(Correlation Mass): fraction of recommendations allocated to items whose latent vectors are closely aligned with an initial user embedding.

Given the decentralized nature of federated learning, we introduce novel divergence metrics to assess the progression of preference amplification:

- **Local-Global Divergence:** Measures how much local user-specific distributions deviate from the aggregated global model.
- **Personalization Degree (Local-Local Divergence):** Quantifies the variance among different local user groups, indicating the formation of distinct echo chambers.

These metrics collectively provide a robust framework for assessing preference amplification and its dynamics in federated environments. By integrating measures of user preference intensity, recommendation concentration, and personalization variability, this approach enables a comprehensive evaluation of trade-offs between user engagement, content diversity, and fairness in decentralized systems. This extension offers critical insights into mitigating preference amplification while leveraging the privacy-preserving and personalization benefits of federated learning.

2.3 Federated Learning

Federated learning (FL) is a decentralized approach where user data remains on local devices, and only model updates are shared with a central server. This method enhances privacy by ensuring personal data isn't transmitted or stored centrally, reducing the risk of breaches [13]. In the context of recommender systems, FL allows for personalized updates through localized training. Each user's device trains the model on their unique preferences, leading to recommendations that better reflect individual tastes. This personalization is achieved without compromising privacy, as raw data stays on the user's device [19]. Additionally, FL mitigates the risks associated with centralized data aggregation, such as data breaches and biased global models. By keeping data decentralized, FL reduces the potential impact of a single point of failure. Moreover, it addresses biases that can arise from over-representation of certain user groups in centralized datasets, promoting more equitable model performance across diverse user bases [25].

3 Modeling Preference Amplification in Federated Settings

3.1 Recommender Setup

We consider a collaborative filtering model based on a standard matrix factorization, enhanced by a variational auto-encoder for learning latent representations. Each user is represented as a latent vector $\mathbf{u}$, and each item as a latent vector $\mathbf{x}$. The system predicts the probability that a user likes an item via a softmax function applied to $\mathbf{u} \cdot \mathbf{x}$.

Parameters like temperature adjust exploration and diversity, while popularity penalties bias the model away from mainstream items, and negative penalties reduce exposure to disliked items. Lower temperature or higher negative penalty can restrict exposure, intensifying preference clustering.

3.2 Federated Learning Parameters

In a federated setup, we assume a set of N local nodes (participants), each holding a subset of user data. Training proceeds in R rounds globally. In each round, the global model is broadcast to nodes, each node trains locally for E epochs on its own data, and then sends updates back. The server aggregates these updates (e.g., using FedAvg [18]) to form the updated global model.

- **N (Nodes):** More nodes means more heterogeneity and potentially more diverse feedback signals.

- **R (Rounds):** Additional communication rounds can gradually homogenize model parameters, or at least continually integrate diverse signals.

- **E (Epochs):** More epochs per round intensifies local overfitting, reinforcing each node's own preferences.

3.3 Measuring Divergence

To capture preference amplification and potential echo chambers:

- **Local-Global Divergence (LGD):** Formally, if P_u represents the local preference distribution for user u, P_g the global distribution implied by the global model, and D_{KL} refers to the Kullback–Leibler divergence as a measure of how one probability distribution diverges from another, then:

$$\mathrm{LGD}(u) = D_{\mathrm{KL}}(P_u \,||\, P_g)$$

Averaging Local-Global Divergence over all users gives a sense of the system-wide local-global misalignment.

- **Personalization Degree (PD):** Consider two clusters of users U_1 and U_2. If P_{U1} and P_{U2} are their aggregated distributions, we measure:

$$\mathrm{PD}(U_1, U_2) = D_{\mathrm{KL}}(P_{U1} \,||\, P_{U2})$$

Tracking Personalization Degree over time reveals whether users converge into distinct preference clusters.

4 Federated Training Procedure

4.1 FedAvg Algorithm

Our experiments rely on a standard Federated Averaging (FedAvg) strategy, adapted for a recommendation model. In each round r:

1. The server sends the global model parameters w_r to all N nodes.
2. Each node i trains its local model w_r^i starting from w_r for E epochs on its local data D_i.
3. Each node returns updated parameters w_r^i.
4. The server aggregates the updates to produce the new global model:

$$w_{r+1} = \frac{1}{N} \sum_{i=1}^{N} w_r^i$$

Pseudo Code for Federated Update:

```
Initialize w_0 randomly
for r in range(R):  # R rounds
    W_updates = []
    for i in nodes[1...N]:
        w_local = copy(w_r)
        for e in range(E): # E local epochs
            w_local = LocalTrain(
                w_local, D_i, params)
        W_updates.append(w_local)
    w_{r+1} = avg(W_updates) # (1/N)*sum(w_local_i)
    # Evaluate LGD, PD and metrics
```

4.2 User Preference Evolution

By incrementally summing weighted recommendation scores with a user embedding, we build on previous work that simulates preference evolution in traditional centralized systems.[15] We then demonstrate that preference amplification also emerges in a federated model, where each federated node simulates its own user that interacts with recommendations. Locally, if a node's user tends to like certain items, repeated exposure further biases their local model. Over epochs, local training steepens preference distributions.

Mathematically, consider the latent factor of a node user u_t. In a non-federated setup, repeated gradient updates according to observed user feedback cause u_t to drift towards stable but narrow configurations. In a federated scenario, after local training, the aggregated model parameter w_{r+1} balances many local node vectors u_t, potentially smoothing out extreme drifts.

5 Experimental Setup

5.1 Data and Model

We use the MovieLens 1M dataset, a popular benchmark for collaborative filtering and recomenders [1]. Each user's rating history is transformed into implicit feedback signals (like vs. not like). A VAE-based model encodes user and item embeddings.

We test different FL configurations:

- N $\in \{5, 20, 50\}$ nodes
- R $\in \{2, 10, 20, 30\}$ rounds
- E $\in \{5, 10, 20, 30\}$ local epochs

Throughout our experiments, we set common recommender parameters to be constants unless noted otherwise to isolate the effects of tuning FL parameters, where temperature=0.5, negative penalty=0.2, popularity penalty=0.1, learning rate=0.01, β=1.0, γ=0.5. All trials are repeated multiple times with different random seeds. We report means and std deviations for metrics like Local-Global Divergence and Personalization Degree over multiple runs.

5.2 Metrics and Analysis

We record:

- **Norm of user embedding (u_t norm):** Measures how strongly user preferences concentrate.
- **Likable vs. non-likable probabilities:** Probability that recommended items are liked. Higher polarization means a larger gap between these probabilities.
- **Correlation mass:** Fraction of probability mass concentrated on a narrow set of items highly correlated with initial preferences.

6 Results

6.1 Baseline Comparison with Centralized Models

Our findings show that the general trends observed in centralized systems, as outlined by Kalimeris et al., are also present in our federated model. Metrics such as the norm of user embeddings (u_t norm), likable versus non-likable probabilities, and the probability of well-correlated items exhibit similar behavior, with minor variations due to initialization and configuration differences when examining variations in system sensitivity (β) and user sensitivity (γ). These results (see Figure 7 and 8 in appendix) confirm that our federated baseline behaves comparably to the centralized setup.

To establish this baseline, we implemented the non-federated recommendation and user preference model described by Kalimeris et al. on a federated system configured with only one node, running a single global round with 75 local epochs. This setup mirrors their centralized model, allowing for direct comparisons. These results provide the foundation for analyzing the effects of federated parameters.

With the baseline established, we proceed to analyze how federated parameters—nodes (N), rounds (R), and epochs (E)—affect the progression of user preferences compared to results observed in centralized recommenders.

6.1.1 Effect of Rounds

More rounds gradually align local models with the global distribution. Increasing the number of rounds (R) decreases the concentration of user preferences and probability of recommending content already familiar to users, thus leading to less preference amplification. This effect is evident in the observed changes to the u_t norm and the Probability of Well-Correlated Items Figure 1.

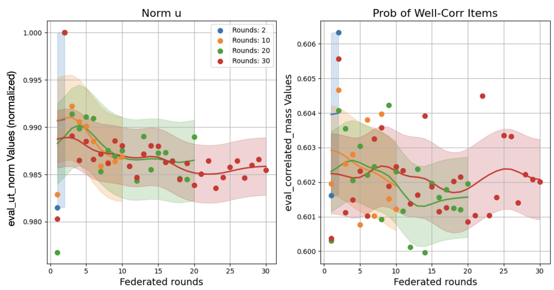

Figure 1: u_t norm and Probability of Well-Correlated Items vs. Federated Rounds. Shaded areas = ±std

As the number of rounds increases, the u_t norm decreases steadily, indicating reduced shifts in user preferences, while its variance becomes smaller, suggesting greater stability across users. A similar trend is observed for the Probability of Well-Correlated Items, where both the mean and variation decrease. These findings highlight how additional rounds contribute to balancing preference amplification through more frequent updates.

6.1.2 Effect of Nodes

Increasing N introduces more heterogeneity in global updates, leading to a lower overall concentration of individual user preferences. Changes in u_t norm are generally smaller for higher node counts compared to lower ones, with an overall downward trend. In the Probability of Well-Correlated Items graph, the standard deviation is notably smaller for higher node counts, indicating that the system becomes more stable as the number of nodes increases as can be seen in Figure 2.

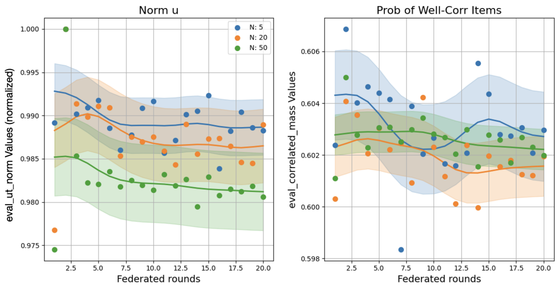

Figure 2: u_t norm and Probability of Well-Correlated Items vs. Federated Rounds with varying Nodes counts. Shaded areas = ±std

6.1.3 Effect of Epochs

We also find that increasing the number of epochs per round generally amplifies preference reinforcement. Users' local embeddings adapt more strongly to their current tastes, reducing the beneficial diversity introduced by federation. Because of this, we observe a significant increase in the user vector norm at the local level when epochs per round increase, as shown in Figures 3. The norm of user embeddings (u_t norm) is generally higher with higher epoch counts, highlighting the role of epochs in driving local adaptation. However, as rounds advance the effects of increased epochs weakens and we observed no significant changes to global aggregation.

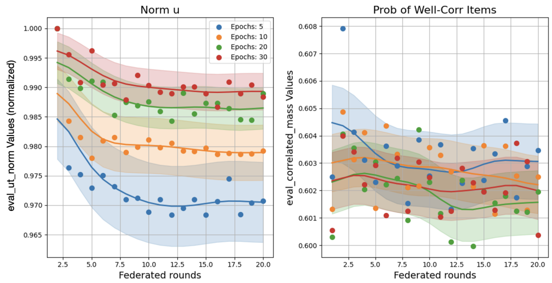

Figure 3: u_t norm and Probability of Well-Correlated Items vs. Federated Rounds with varying Epoch counts. Shaded areas = ±std

6.2 Evidence of Moderation from the Provided Data

We can now reintroduce the measures *Local-Global Divergence* and *Personalization Degree*, as defined in Section 2.3: Local-Global Divergence measures the deviations between local and global models, while Personalization Degree measures the variability between different nodes.

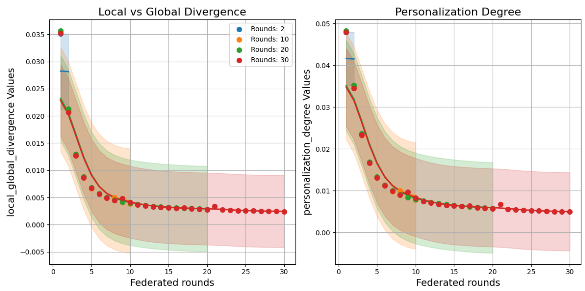

Figure 4: Local-Global Divergence vs. Federated Rounds for different round values. Shaded areas = ±std

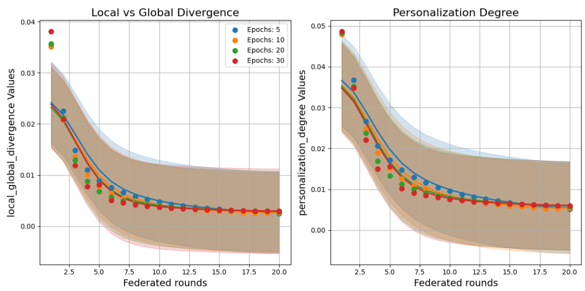

Figure 5: Local-Global Divergence vs. Federated Rounds for different Epochs. Shaded areas = ±std

Figures 4,5, and 6 show that increasing N and R, or reducing E, stabilizes Local-Global

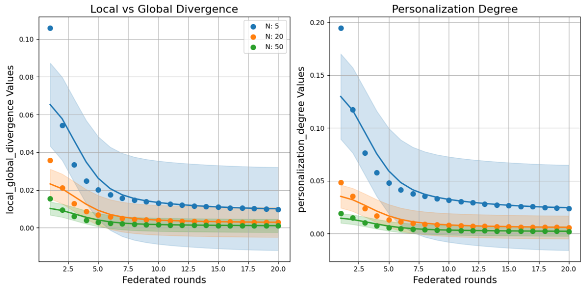

Figure 6: Local-Global Divergence vs. Federated Rounds for different Nodes. Shaded areas = ±std

Divergence as well as Personalization Degree, lowering the risk of forming echo chambers. Mean Local-Global Divergence curves flatten more quickly with larger N, and variances shrink. Reducing E moderates how quickly norm u and correlated mass grow, indicating slower preference amplification.

Higher round counts generally reduce the differences between local and global models and decrease variance across user groups. This trend is evident in the observed downward trajectory of both Local-Global Divergence and Personalization Degree, as well as the decreasing standard deviation with more rounds in Figure 4. When comparing this to Figure 5, where epochs vary but rounds are fixed at 20, we observe similar patterns. While increasing epochs locally amplifies user preference changes and reinforces existing preferences, these effects are largely mitigated globally through the aggregation of updates over rounds. Thus, the effects of increased epochs are smoothed out globally through the update aggregation process over rounds. Our results suggest that larger node counts dilute individual effects and promote greater consistency across the global model. When considering the same measures with rounds fixed at 20 and varying N, we observe that fewer nodes (e.g., $N = 5$) result in an increase in Local-Global Divergence and Personalization Degree. By contrast, with $N = 50$, the global model averages multiple distinct user profiles each round, producing a "dampening effect" on large differences between local and global models as can be seen in Figure 6.

7 Discussion

Our findings demonstrate that federated parameters, such as the number of nodes, rounds, and epochs, act as "dials" to moderate preference amplification. The federated structure itself allows this balance to be fine-tuned by dynamically adjusting parameters like rounds, nodes, and epochs. For example, increasing the number of participants and rounds helps slow the formation of echo chambers, while reducing the number of local epochs prevents excessive reinforcement of user preferences. These parameters complement other recommender hyperparameters, offering system designers a practical way to balance personalization and diversity. These promising results suggest Federated Learning's potential to act as an additional tool to address the trade-off between personalization and echo chamber formation. This flexibility enables the design of adaptable systems that can effectively mitigate preference amplification.

8 Conclusion

Building on the theoretical foundation of preference amplification [15] and incorporating federated learning paradigms [14], we analyzed how FL parameters interact with recommender system dynamics. Our main contributions:

1. Extended Theory of Preference Amplification: We integrate federated parameters (Nodes, Rounds, Epochs) into existing models of preference amplification, showing that increasing Nodes and Rounds can moderate the intensity and pace of echo chamber formation.
2. Novel Divergence Metrics: We introduce Local-Global Divergence and Personalization Degree to capture how local and global distributions evolve in tandem. These metrics serve as actionable indicators of polarization and preference clustering.
3. Practical Guidelines: By tuning N, R, and E, in addition to classical recom-

mender β, γ, temperature, and penalty factors, practitioners can balance personalization and diversity. More nodes and rounds slow preference amplification, while fewer epochs help avoid reinforcing narrow tastes.

8.1 Answers to Research Questions

- RQ1: Preference amplification persists but can be controlled. FL parameters shape how quickly and strongly user preferences consolidate.
- RQ2: Federated learning can delay or soften the formation of echo chambers by regularly introducing diversity through global aggregation. It does not eliminate them entirely but can serve as a mitigating force.
- RQ3: While federated systems can still overfit for engagement, higher N and R reduce this risk. Adjusting E and other model hyperparameters provides a further handle on controlling overfitting.

8.2 Future Directions

Our research leaves room for replication and validation on larger, more diverse datasets and broader experimental scales. Future work can refine these findings and explore their generalizability in different contexts and platforms. Additional studies may quantify divergence rates and examine more complex interaction models. Investigating adaptive federation strategies—where N, R, and E dynamically adjust to observed divergence metrics—could optimize long-term diversity and user satisfaction. Studies using user studies and real-world policy interventions would clarify the relationship between these theoretical findings and practical outcomes. Finally, integrating fairness criteria and ensuring that minority or niche user groups remain visible and well-served offers a promising avenue to prevent marginalization and maintain a healthy information ecosystem.

References

[1] MovieLens 1M Dataset, September 2015. URL `https://grouplens.org/datasets/movielens/1m/`.

[2] Charu C. Aggarwal. *Recommender Systems*. Springer International Publishing, Cham, 2016. ISBN 978-3-319-29657-9 978-3-319-29659-3. doi: 10.1007/978-3-319-29659-3. URL `http://link.springer.com/10.1007/978-3-319-29659-3`.

[3] Axel Bruns. Filter bubble. *Internet Policy Review*, 8(4), November 2019. ISSN 2197-6775. doi: 10.14763/2019.4.1426. URL `https://policyreview.info/node/1426`.

[4] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. Bias and Debias in Recommender System: A Survey and Future Directions, December 2021. URL `http://arxiv.org/abs/2010.03240`. arXiv:2010.03240 [cs].

[5] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. The COVID-19 social media infodemic. *Scientific Reports*, 10(1):16598, October 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-73510-5. URL `https://www.nature.com/articles/s41598-020-73510-5`.

[6] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, and Dasarathi Sampath. The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296, Barcelona Spain, September 2010. ACM. ISBN 978-1-60558-906-0. doi: 10.1145/1864708.1864770. URL `https://dl.acm.org/doi/10.1145/1864708.1864770`.

[7] Gregory Eady, Jonathan Nagler, Andy Guess, Jan Zilinsky, and Joshua A. Tucker. How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data. *Sage Open*, 9(1):2158244019832705, January 2019. ISSN 2158-2440. doi: 10.1177/2158244019832705. URL https://doi.org/10.1177/2158244019832705. Publisher: SAGE Publications.

[8] Daniel M. Fleder and Kartik Hosanagar. Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity, September 2007. URL https://papers.ssrn.com/abstract=955984.

[9] Zhaolin Gao, Tianshu Shen, Zheda Mai, Mohamed Reda Bouadjenek, Isaac Waller, Ashton Anderson, Ron Bodkin, and Scott Sanner. Mitigating the Filter Bubble While Maintaining Relevance: Targeted Diversification with VAE-based Recommender Systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2524–2531, Madrid Spain, July 2022. ACM. ISBN 978-1-4503-8732-3. doi: 10.1145/3477495.3531890. URL https://dl.acm.org/doi/10.1145/3477495.3531890.

[10] Yingqiang Ge, Shuya Zhao, Honglu Zhou, Changhua Pei, Fei Sun, Wenwu Ou, and Yongfeng Zhang. Understanding Echo Chambers in E-commerce Recommender Systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2261–2270, July 2020. doi: 10.1145/3397271.3401431. URL http://arxiv.org/abs/2007.02474. arXiv:2007.02474 [cs].

[11] Carlos A. Gomez-Uribe and Neil Hunt. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions on Management Information Systems*, 6(4): 1–19, January 2016. ISSN 2158-656X, 2158-6578. doi: 10.1145/2843948. URL https://dl.acm.org/doi/10.1145/2843948.

[12] Natali Helberger, Kari Karppinen, and Lucia D'Acunto. Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, 21(2):191–207, February 2018. ISSN 1369-118X, 1468-4462. doi: 10.1080/1369118X.2016.1271900. URL https://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1271900.

[13] Danish Javeed, Muhammad Shahid Saeed, Prabhat Kumar, Alireza Jolfaei, Shareeful Islam, and A. K. M. Najmul Islam. Federated Learning-Based Personalized Recommendation Systems: An Overview on Security and Privacy Challenges. *IEEE Transactions on Consumer Electronics*, 70(1): 2618–2627, February 2024. ISSN 1558-4127. doi: 10.1109/TCE.2023.3318754. URL https://ieeexplore.ieee.org/document/10262054?utm_source=chatgpt.com. Conference Name: IEEE Transactions on Consumer Electronics.

[14] Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G. L. D'Oliveira, Hubert Eichner, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaid Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konecný, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrède Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Özgür, Rasmus Pagh, Hang Qi, Daniel Ramage, Ramesh Raskar, Mariana Raykova, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun,

Ananda Theertha Suresh, Florian Tramèr, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, and Sen Zhao. Advances and Open Problems in Federated Learning. *Foundations and Trends® in Machine Learning*, 14 (1–2):1–210, 2021. ISSN 1935-8237, 1935-8245. doi: 10.1561/2200000083. URL http://www.nowpublishers.com/article/Details/MAL-083.

[15] Dimitris Kalimeris, Smriti Bhagat, Shankar Kalyanaraman, and Udi Weinsberg. Preference Amplification in Recommender Systems. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 805–815, Virtual Event Singapore, August 2021. ACM. ISBN 978-1-4503-8332-5. doi: 10.1145/3447548.3467298. URL https://dl.acm.org/doi/10.1145/3447548.3467298.

[16] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. Efficiency Improvement of Neutrality-Enhanced Recommendation. *Conference: The 3rd Workshop on Human Decision Making in Recommender Systems (Decisions)*, October 2013.

[17] Rune Karlsen, Kari Steen-Johnsen, Dag Wollebæk, and Bernard Enjolras. Echo chamber and trench warfare dynamics in online debates. *European Journal of Communication*, 32 (3):257–273, June 2017. ISSN 0267-3231. doi: 10.1177/0267323117695734. URL https://doi.org/10.1177/0267323117695734. Publisher: SAGE Publications Ltd.

[18] Jakub Konečný, H. Brendan McMahan, Felix X. Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated Learning: Strategies for Improving Communication Efficiency, October 2017. URL http://arxiv.org/abs/1610.05492. arXiv:1610.05492 [cs].

[19] Zhiwei Li, Guodong Long, and Tianyi Zhou. Federated Recommendation with Additive Personalization, February 2024. URL http://arxiv.org/abs/2301.09109. arXiv:2301.09109 [cs].

[20] Gabriel Machado Lunardi, Guilherme Medeiros Machado, Vinicius Maran, and José Palazzo M. De Oliveira. A metric for Filter Bubble measurement in recommender algorithms considering the news domain. *Applied Soft Computing*, 97:106771, December 2020. ISSN 15684946. doi: 10.1016/j.asoc.2020.106771. URL https://linkinghub.elsevier.com/retrieve/pii/S1568494620307092.

[21] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data, January 2023. URL http://arxiv.org/abs/1602.05629. arXiv:1602.05629 [cs].

[22] Eli Pariser. *The filter bubble: what the Internet is hiding from you.* Viking, London, 2011. ISBN 978-0-14-196992-3. OCLC: 748427853.

[23] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 131–141, Barcelona Spain, January 2020. ACM. ISBN 978-1-4503-6936-7. doi: 10.1145/3351095.3372879. URL https://dl.acm.org/doi/10.1145/3351095.3372879.

[24] Jianshan Sun, Jian Song, Yuanchun Jiang, Yezheng Liu, and Jun Li. Prick the filter bubble: A novel cross domain recommendation model with adaptive diversity regularization. *Electronic Markets*, 32(1):101–121, March 2022. ISSN 1019-6781, 1422-8890. doi: 10.1007/s12525-021-00492-1. URL

https://link.springer.com/10.1007/s12525-021-00492-1.

[25] Zehua Sun, Yonghui Xu, Yong Liu, Wei He, Lanju Kong, Fangzhao Wu, Yali Jiang, and Lizhen Cui. A Survey on Federated Recommendation Systems, March 2023. URL http://arxiv.org/abs/2301.00767. arXiv:2301.00767 [cs].

[26] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology*, 10 (2):1–19, March 2019. ISSN 2157-6904, 2157-6912. doi: 10.1145/3298981. URL https://dl.acm.org/doi/10.1145/3298981.

A notable distinction, however, is the lack of clear separation between different β values in our results. This is likely because of the ten times fewer timesteps in our experiments, which limits the granularity required to observe such divergences.

The results for varying γ values align with the observed trends. The separation between γ values is most evident in the Norm of User Embeddings (u_t norm), where higher γ values result in stronger growth over time. This underscores the role of user sensitivity in amplifying initial preferences and intensifying feedback loops.

An anomaly is observed in the Probability of Well-Correlated Items, where the values unexpectedly fall into the negative range instead of staying between 0 and 1 as expected. This is likely due to a configuration or calculation error.

A Appendix

A.1 Resources

GitHub: http://github.com/stoille/FedRecPolicyEval

Instructions about how to download appropriate libraries and execute the experiments in this paper can be found in the repo's README.md. A copy of this paper and our runtime data is publicly available there too.

A.2 In depth comparisons of base metrics γ and β

Our results confirm the same general trends observed by the Meta study. The u_t norm, likable probability, and the probability of well-correlated items all increase over time, while the non-likable probability decreases. This indicates that, similar to their system, our federated model influences user preferences, consistently prioritizing highly popular content and adapting closely to the initial preferences of the user. The system shows a strong bias toward reinforcing what the user initially finds engaging, while deprioritizing less popular content.

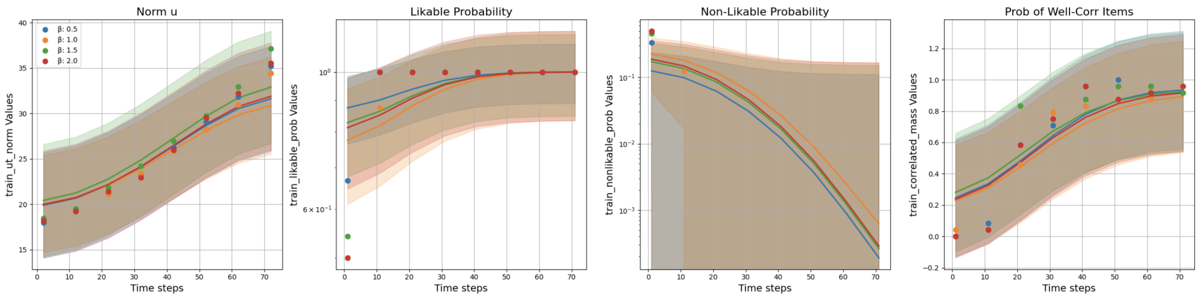

Figure 7: Effect of varying β per round on key metrics. The graphs show the trends in Likable and Non-Likable probabilities (log scale), user norms (u_t norm), and the Probability of Well-Correlated Items over time steps. Shaded regions represent the standard deviation across experiments.

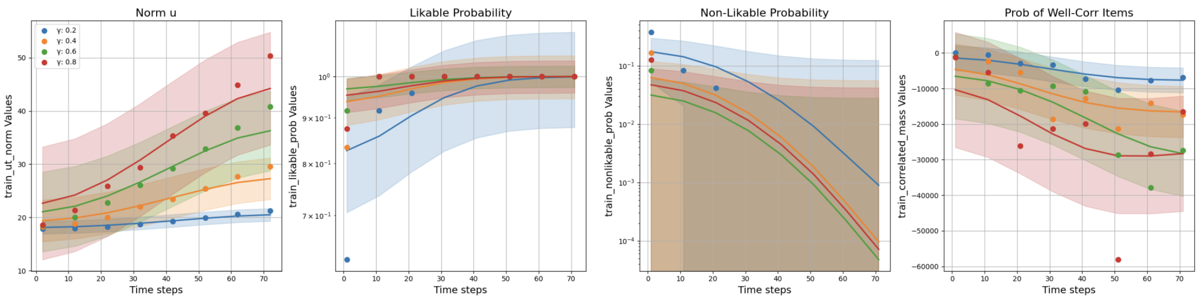

Figure 8: Effect of varying γ per round on key metrics. The graphs show the trends in Likable and Non-Likable probabilities (log scale), user norms (u_t norm), and the Probability of Well-Correlated Items over time steps. Shaded regions represent the standard deviation across experiments.