



Uncovering the Deep Filter Bubble: Narrow Exposure in Short-Video Recommendation

Nicholas Sukiennik
Department of Electronic
Engineering, Beijing National
Research Center for Information
Science and Technology
Tsinghua University
Beijing, China
sukiennikn10@mails.tsinghua.edu.cn

Chen Gao*
Beijing National Research Center for
Information Science and Technology,
Tsinghua University
Beijing, China
chgao96@gmail.com

Nian Li
Shenzhen International Graduate
School, Tsinghua University
Shenzhen, China
linian21@mails.tsinghua.edu.cn

ABSTRACT

Filter bubbles have been studied extensively within the context of online content platforms due to their potential to cause undesirable outcomes such as user dissatisfaction or polarization. With the rise of short-video platforms, the filter bubble has been given extra attention because these platforms rely on an unprecedented use of the recommender system to provide relevant content. In our work, we investigate the deep filter bubble, which refers to the user being exposed to narrow content within their broad interests. We accomplish this using one-year interaction data from a top short-video platform in China, which includes hierarchical data with three levels of categories for each video. We formalize our definition of a "deep" filter bubble within this context, and then explore various correlations within the data: first understanding the evolution of the deep filter bubble over time, and later revealing some of the factors that give rise to this phenomenon, such as specific categories, user demographics, and feedback type. We observe that while the overall proportion of users in a filter bubble remains largely constant over time, the depth composition of their filter bubble changes. In addition, we find that some demographic groups that have a higher likelihood of seeing narrower content and implicit feedback signals can lead to less bubble formation. Finally, we propose some ways in which recommender systems can be designed to reduce the risk of a user getting caught in a bubble.

CCS CONCEPTS

• Information systems → Personalization.

KEYWORDS

Filter Bubble; Short-Video Recommendation; Recommender Systems

ACM Reference Format:

Nicholas Sukiennik, Nian Li, Chen Gao. 2024. Uncovering the Deep Filter Bubble: Narrow Exposure in Short-Video Recommendation. In *Proceedings of*

*Corresponding author.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WWW '24, May 13–17, 2024, Singapore, Singapore

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0171-9/24/05.

<https://doi.org/10.1145/3589334.3648159>

the ACM Web Conference 2024 (WWW '24), May 13–17, 2024, Singapore, Singapore. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3589334.3648159>

1 INTRODUCTION

Automatic recommendation has been the standard way to make sense of the unprecedented quantities of online information since their implementation on social media platforms like Facebook as early as 2010 [14]. While recommender systems have solved the problem of providing relevant content to users and creating a more user-friendly online experience, they have also given rise to some problems, namely, the filter bubble. The filter bubble was first described by Eli Pariser in 2011 [26] as the phenomenon where the content seen on social media platforms becomes more and more personalized over time so that users eventually lack exposure to the full variety of content that exists.

More recently, the rise of short-video platforms has revolutionized both the way content is consumed and recommended on social media. Due to the single-column video layout ("Featured", or "For You Page"), users simply have to swipe up to enjoy an endless feed of unique content tailored to their interests. However, due to the lack of additional interest signals, the recommender must make use of largely implicit feedback on recommended content to learn users' preferences. Because on-platform interests are both learned and formed in the same space, the short-video recommender potentially has a larger risk of leading to the formation of a filter bubble. In other words, since there is no user-specified information such as search or browsing activity, the system has little information other than the feedback on existing videos to improve its ability to capture users' interests. This means it is possible that the recommender will only learn a small portion of user interests rather than all of them, and because most feedback is positive, it does not feel the need to learn more.

It is important to mention some concepts that are similar and often mentioned in the same context as the filter bubble: echo chamber and information cocoon. The concept of echo chamber is distinct from a filter bubble in that it refers to a scenario in which a user is embedded in a community where content of a certain standpoint or bias is continuously viewed [23] [16]. The echo chamber could be seen as a potential negative outcome of the filter bubble, but the echo chamber typically refers to the entire set of information seen by a person, and not that coming from one individual platform [6] [7]. As such, we do not address the echo

chamber explicitly in this work, leaving it to those in the fields of social science.

Information cocoon, on the other hand, refers to a user's exposure to information on a specific platform that often covers similar topics, regardless of whether it was caused by user action or by the recommender system [18]. In this way, we consider a filter bubble as a special case of information cocoon that is specifically related to the recommender system of a given platform. In this work, we address the filter bubble, as it is our primary interest to understand how the recommender system adapts to users' deeper interests.

Since its initial discovery, there have been many works addressing the filter bubble problem from several different perspectives, either to understand how it is formed or to alleviate it with novel recommendation models. As for the former, [27] develops a model of the adaptive information dynamics between the users and the recommender algorithms, showing that user feedback is the essential factor in leading the recommender towards more or less diverse content. The presence of filter bubbles on short-video platforms was explored in [18]. On the other hand, [20] design a simulation framework to model the recommender system pipeline of data collection, model learning, and item exposure in a loop, and propose a reinforcement learning (RL) method to select cross-community item exposures. Other works attempt to resolve the filter bubble via specialized model design [8], including those that are designed to increase diversity [36][13][34]. These models increase diversity while maintaining high accuracy.

1.1 User Decision Process

User decision processes in online environments, influenced by behavioral science and recommender systems, form the basis for our investigation of the deep filter bubble. Central is the exploration vs. exploitation concept, indicating users either stick to familiar choices or explore new ones based on their desire to reduce uncertainty or maximize coherency, respectively [29]. Coherency maximization reflects a drive for choices consistent with past ones, and can significantly influence political preferences [11].

Studies on online music platforms reveal users' preferences evolve over time, with more diverse music consumption leading to higher engagement [1, 30]. This diversity increase may represent online coherency maximizing exploration, contrasting with offline contexts where constant exploitation may lead to user boredom. Several other studies have emphasized the importance of diversity in recommendation [17][21][15][13].

The mere exposure effect, where increased exposure to a stimulus enhances positive feelings towards it [35], can impact recommender systems. Overemphasis on coherency can result in a filter bubble, with too much sameness leading to user boredom. Furthermore, the mere exposure effect could forge content-interests based on the recommender's offerings, potentially leading to less diverse content and a deep filter bubble.

Understanding these behaviors, and the role of recommenders in shaping them, has social significance. Filter bubbles can undermine informed decision-making, essential in democratic societies [32] [31]. Moreover, online segregation, exacerbated by algorithmic bias and personalized feeds, can fuel inter-community tensions [12][10].

However, up until now, no works have investigated the hierarchical nature of the filter bubble. That is, rather than studying the

formation of a bubble as subset of categories in relation to all the topics on the platform, no works have considered the narrowing of exposure to subcategories below those top-level interests. For example, a traditional filter bubble may mean that a user is seeing content mostly related to sports, rather than news, humor, pets, lifestyle, dance, etc. But a deep filter bubble would mean that within the topic of sports, a user's exposed content narrows over time from a variety of sports, to only soccer, and within soccer, only a given team.

In the current work, we aim to address this gap by validating the presence of a "deep" filter bubble, namely the narrowing of exposure to subcategories over time. To do so, we leverage a dataset from one of the top short-video platforms in China, wherein each video is labeled with three-level hierarchical categories. We conduct an extensive exploration of the data to provide a novel understanding of the depth dimension of the filter bubble.

The contributions of this paper can be summarized as follows.

- We develop a robust methodology for evaluating the presence and extent of the deep filter bubble.
- We conduct a comprehensive long-term analysis of the deep filter bubble and its correlating factors on a large real-world dataset.
- We provide insights as to how recommender models can prevent or alleviate the formation of the deep filter bubble and create more informative and gratifying content experiences for users.

The remaining part of this paper is organized as follows. We first formulate the methodology of our analysis in Section 2, followed by the overall filter bubble analysis in Section 3 and the factors analysis in Section 4. Then, we present the literature review in Section 5 and discuss important implications in Section 6. Finally, we conclude our paper in Section 7.

2 DATASET AND METHODOLOGY

2.1 Dataset

We use a large dataset of anonymized user-item interaction data from the single-column feed of a popular short-video application consisting of 400 million interactions, 400 thousand users, and 20 million items over a one-year period where all the users registered on the first day of the year. The data is filtered down to include only the videos that were viewed by the users and not those that were skipped immediately. The dataset includes three levels of hierarchical categories for each item and different forms of user feedback such as like, collect, and comment.

2.2 Hierarchical Structure

As our work heavily revolves around the hierarchical category structure of the data, we first motivate our investigation by presenting some basic statistics of this tree-like structure.

We first note the number of total unique categories per level across the whole dataset, depicted in figure 2a: 58, 326, and 512 in level-ascending order. The number of categories increases with level, which is intuitive, as we would expect the tree to have more branches than roots. However, in figure 2b, we note that the number of unique category children per level is, on average, higher for level 1 than for level 2. This is a counter-intuitive finding, as we would

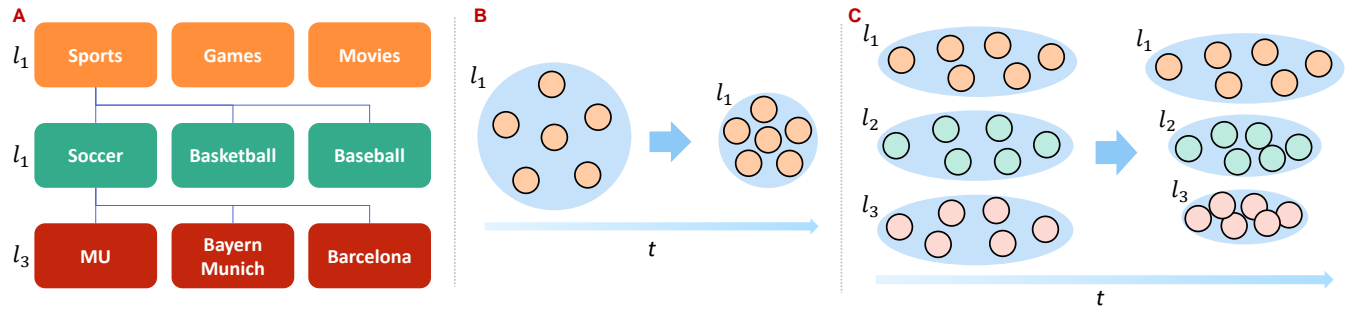


Figure 1: A conceptual overview of the traditional filter bubble, where A represents an example tree structure of hierarchical categories, and B and C represent the shallow and the deep filter bubbles, respectively, wherein the distance between items represents similarity. The items in B are assumed to be all uniform in level, whereas in C they are broken down into three levels. Figure C assumes that deeper-levels converge on more similar items over time.

expect that there are more branches at the deeper levels considering that there are more deeper categories overall.

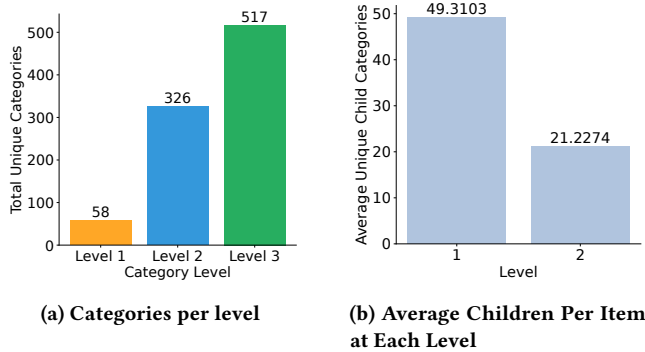


Figure 2: Data Investigation

2.3 Defining the Deep Filter Bubble

Different works use different metrics to quantify the filter bubble depending on their data or the problem they are trying to solve. In our case, we are concerned with the decrease in exposure of deeper-level categories over time, so we must devise a metric that accurately reflects this phenomenon. To quantify this, we use a normalized form of coverage, where coverage is defined as the number of unique categories seen by a user over a specified time period.

We formally define the metric for the filter bubble at each level by taking the coverage as a ratio of categories seen versus the total possible categories seen for a given level.

$$C_l = \frac{n_{\text{seen},l}}{n_{\text{total},l}}, \quad (1)$$

where C_l is the coverage at a given level, $n_{\text{seen},l}$ is the number of categories seen at that level, and $n_{\text{total},l}$ is the total number of possible categories at that level.

We can also compute an overall coverage by summing the seen categories and total categories over all the levels:

$$C = \sum_{l=1}^N \frac{n_{\text{seen},l}}{n_{\text{total},l}}. \quad (2)$$

These equations allow us to quantify the conceptual phenomenon of the deep filter bubble, as illustrated in Figure 1, using the coverage of categories to represent the similarity of items exposed to each user.

3 OVERALL DEEP FILTER BUBBLE CHARACTERIZATION

3.1 Diversity Investigation

In order to present results in an easily interpretable format, we split our interaction data into 52 time windows representing the weeks of the year, where user interaction statistics are consolidated into weekly statistics by taking their averages. Using the definition of low diversity as being a strong indicator of filter bubble formation, we conduct a preliminary analysis of the diversity, in terms of coverage ratio, for all users. This metric is provided on an overall and per-level basis in Figure 3.

From the overall coverage in Figure 3a, we can see an obvious decrease in the overall coverage ratio towards the end of the year, despite some fluctuations throughout. In Figure 3b, we can start to determine the reasons for the trend and fluctuations. Qualitatively, we can state that the coverage decrease is caused primarily by a decrease in top-level category exposure, whereas the second level counteracts that decrease to some extent, and also causes a spike at the end. At the same time, the third level is mostly constant over time, with some rapid fluctuations towards the end. From this perspective, we can define three distinct possible filter bubbles: one for each category level. Based on the quantitative analysis, we can conclude that there is an obvious filter bubble being formed on the first level, no filter bubble formation on the second level, and minimal filter bubble formation on the third level.

3.2 Evolution over Time

In a platform of millions of users, it is possible for the recommender to respond differently according to different users' needs and interests. With this in mind, we design a criterion to label a user as

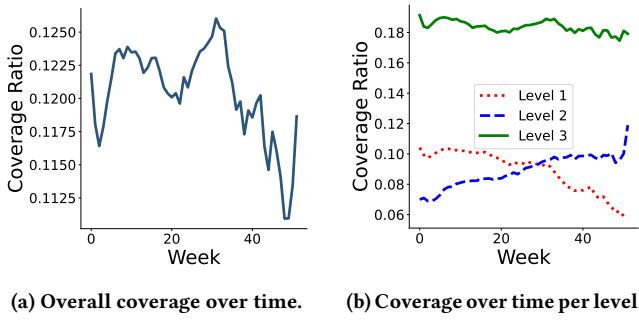


Figure 3: Coverage over time.

"in" or "out" of a filter bubble at each time window. To do this, we specify a coverage threshold: if the categories seen by a user are less than the median number of categories seen by all users on a given level, then we consider that user to be "in" the filter bubble at that level:

$$\mathcal{B}_{u,t} = \begin{cases} \text{"in"}, & \text{if } C_{u,t} < \text{Median}(C_{.,t}) \\ \text{"out"}, & \text{otherwise} \end{cases}, \quad (3)$$

where $\mathcal{B}_{u,t}$ is the status of user u at time window t , indicating whether they are "in" or "out" of the filter bubble, $C_{u,t}$ is the number of categories seen by user u during time window t and $\text{Median}(C_{.,t})$ is the median number of categories seen by all users during time window t . On this basis, we proceed to calculate the proportion of users who are in the filter bubble at each level for each week in the year.

From Figure 4, we can observe that as time progresses, the proportion of users in the filter bubble at each level varies in unique ways. For level 1, while remaining constant for most of the year, the number of users in the filter bubble increases within the final 20 weeks of the year. Level two, on the other hand, dominates the deep filter bubble ratio at the beginning, whereas by the end of the year, it significantly decreases, having the lowest filter bubble ratio. Over the same period, the proportion of users in the bubble for level two decreases significantly and then pans out by the end of the year before a final drop. Interestingly, the filter bubble for the deepest level remains relatively constant most of the year, even decreasing slightly before fluctuating during the final 20 weeks of the week. Both levels 2 and 3 have a substantial drop in the last three weeks of the year, resulting in an end state where filter bubble of level 1 is dominant, followed by levels 3 and 2 respectively.

The conclusions we can gather here are that upon first registering for the app, a user is more likely to see narrower content at a medium depth level, but over time, the content of the medium and deep levels grow more diverse, and the content in the top level gets narrowest. To use an intuitive example, if we consider a just-registered user who enjoys watching videos about games, movies, and music, the videos that they watch within each of them will span a variety of genres: for games, puzzle, sports, and strategy games; for movies, action, comedy, and romance movies; and for music, rock, electronic, and pop. But by the end of the year, the recommender will learn that they prefer games more than movies and music, but within games, they will now see four or five different

game genres. Meanwhile, the specific games, film actors/directors, and music artists under each genre will remain equally diverse over time. Assuming that the recommender is accurately learning users' real interests, we can conclude that a user who likes a specific form of media would be more interested in various genres within that form of media with more likelihood than they would be interested in two entirely distinct forms of media. Expanding to the general scenario, we can state that diversity of second-level categories is more important than diversity of top-level categories *once* the users' top-level interests are well understood, but not before.

However, if we consider that a filter bubble can give rise to an information cocoon or echo chamber, we should also pay attention to whether the narrowing of top-level categories is truly beneficial to the user over the long term and consider the trade-off between platform engagement and user satisfaction.

3.3 Distribution over Users

Because our metric for filter bubble presence is relative to the number of categories seen by other users, we should also define an absolute metric to determine the tendency for category-levels themselves to have a stronger tendency to form filter bubbles, regardless of the relative behaviors of the users viewing those category-levels. To do so, we normalize the coverages per user using equation 1, and then we calculate the probability density of each coverage ratio. Figure 5 depicts the resulting probability density distributions for both the overall coverage and the coverage per level. By decomposing the overall distribution into a separate distribution for each category level, we can understand the reasons for the contours of the probability density. In Figure 5a, there is a spike in probability density at a very low coverage ratio, and there is also a seemingly separate Gaussian distribution centered around approximately 0.35. By looking at Figure 5b, we can see that the initial spike is due to the fact that the probability densities for both levels 2 and 3 are extremely sharp at a low value between 0 and 0.1. Meanwhile, although there is a small initial spike for level 1, the distribution mostly centers in a normal manner around 0.2.

From this, we can conclude that, although the relative filter bubble was the strongest for level 1, as determined by the results of section 3.2, it also turns out that level 1 is the level with the most equitable distribution of coverages over all users. On the other hand, it turns out to be levels 2 and 3 that display a distribution of coverages that are restricted to a narrow low range. This result could be partially explained by the results in Figure 2b, which we recall reveals that level 1 categories, on average, have more children than their level 2 counterparts. At the same time, each level-category's children only make up a fraction of the categories for the following level. This tree structure means that a user who only sees a few upper-level categories has a specific limited number of deeper categories that are *possible* to see, given that the other categories are children of parents whom they also did not see. In this way, it is intuitive that our distribution of coverage will decrease the deeper the category level. We can generalize this phenomenon by saying that a user who has seen a certain fraction of categories at an upper level will necessarily see a smaller absolute fraction of categories at the lower levels, assuming that children are evenly dispersed.

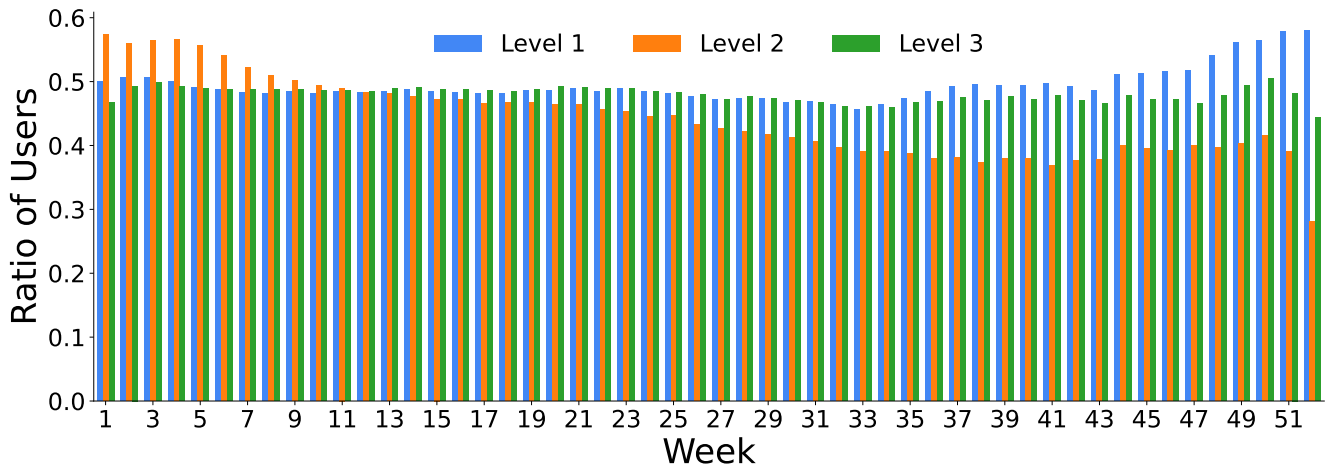
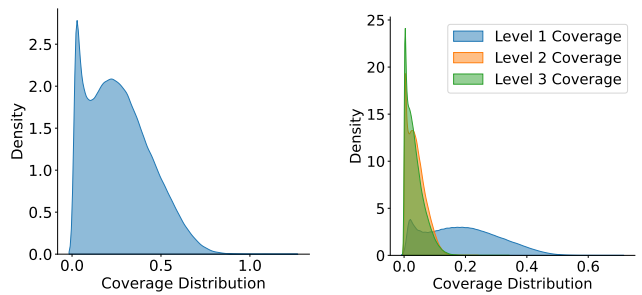


Figure 4: Ratio of users in bubble for each level over time.



(a) Distribution of User Coverage over all Levels

(b) Distribution of User Coverage Per Level

Figure 5: Coverage probability density distribution overall and for each level.

From a user standpoint, it is natural that users’ interests would be more narrow as they get deeper. But from a platform perspective, it is important to broaden upper-level category exposure so that users do not get stuck in deeper-level absolute bubbles based on their incomplete upper-level exposure.

Now that we have an understanding of the overall state of the deep filter bubble for our dataset, we proceed with an analysis of the factors that could potentially contribute to the formation of the deep filter bubble from three angles: specific categories, user characteristics, and feedback.

4 FACTORS ANALYSIS

4.1 The Effect of Specific Categories

To determine whether certain categories have a "pulling" effect, i.e. cause the filter bubble to be pulled into the filter bubble more so than other categories, we perform an analysis of the probabilities with which users switch from category to category over time. Since each user sees potentially hundreds of videos per week, we must devise a concise way of classifying the important categories for each

user for which to compute the probabilities. To do so, we select the 50 dominant categories per user for each level. Then we calculate the probability of each of these dominant categories changing to another category from week to week. Finally, we average the probabilities over all users and all weeks, providing us with a general picture of the pulling effect of the different categories for each level.

The results, in Figure 6, tell us that for level 1, users are more likely to remain viewing the same categories from week to week, which is in line with our expectations. We would assume that a user has a stronger opinion on top-level, broad, categories, than deeper sub-categories. This tells us that for the top level, there is a potential of being pulled into a filter bubble based on initial interests. In Figure 6b, we note that there is no specific pattern of transition, telling us that a user may have a wide variety of medium-depth interests and that none of them has a particular pulling effect. A similar conclusion can be drawn from Figure 6c for the third-level categories. From this analysis, we conclude that there are essentially no specific categories that have an overwhelming pulling effect on users’ viewing activity. However, we also conclude that for top-level interest, there is a chance of a filter bubble forming based on the categories viewed initially.

4.2 Effects of User Characteristics

We conduct an extensive analysis of user characteristics with respect to category coverage at each level in Figure 7, with the goal of determining if there are any personal factors that could lead to a user being more or less prone to get trapped in a filter bubble. To do so, we plot an empirical cumulative distribution function to depict the strength of the relationship between category coverage and each of the four demographic characteristics: age, city level, phone price, and gender. The first thing we note is that for all demographics and levels, there is virtually no variation in probability for any of the variables after a coverage of 0.5. This is because the vast majority of users’ coverages are in the lower half of the coverage range. Thus, we will focus our analysis on the first half of the coverage ranges to determine if there are any significant correlations between users’ characteristics and coverage.

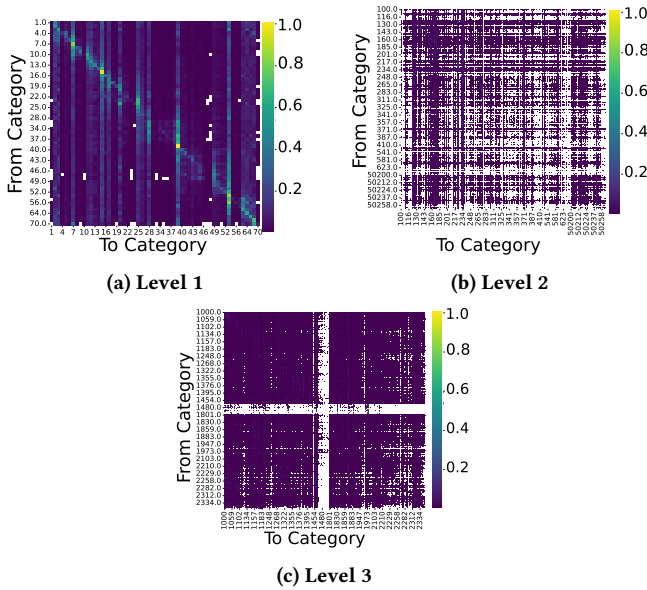


Figure 6: Probability of switching dominant categories at any point in time for each level.

4.2.1 Age. First, we look at the correlations between age and coverage ratio. Looking closely, we notice that there is a slight relationship between age and coverage, the one in level 3 being the most salient. The trend shows us that younger users have a higher probability of having a lower coverage ratio. This could pose a risk for younger users to get stuck in a deep filter bubble more easily than older users.

4.2.2 City Level. Since our data was collected in China, we classify the cities of each anonymized user by their city-tiers, a classification that roughly corresponds to the size and economic strength of a city in China. Although the relationship between city-tier and coverage is not especially strong, the figures for all three levels do show that users in first-tier cities have the highest relative coverage compared to users in other cities, and slightly more so in level 3, suggesting that users in lower tier-cities could be more likely to get stuck in a deep filter bubble.

4.2.3 Phone Price. Phone price has been used in prior works as a proxy metric for income [18]. In our dataset, there is no exceedingly evident trend between phone price and coverage.

4.2.4 Gender. From the figures, we can gather that gender indeed plays a part in users’ exposure to diverse categories, but only on levels 1 and 3. For these two levels, there is a tendency for males to have lower coverage than females, implying that males could be more likely to get caught in a shallow or deep filter bubble.

4.3 Influence of Feedback

A user’s feedback represents their level of satisfaction or dissatisfaction with the content provided by the recommender on a social media platform. However, different types of feedback can have different meanings. Namely, implicit and explicit feedback are used

in different ways by the recommender. Implicit feedback is much more abundant than explicit feedback and is therefore much more commonly used to train recommendation models. However, there are still questions as to how to use different forms of feedback to optimally learn user preferences [25]. For example, if a user finishes watching a video but doesn’t tap the like button, does she like that video? Or does "liking" override the fact that a video was not watched entirely through?

In this section, we will consider the ways in which different forms of feedback inform the recommender in different ways, and investigate in which ways if any, these different forms of feedback lead to different trends in the formation of the filter bubble. If there is a difference in trend between implicit and explicit feedback in the formation of the filter bubble, this could be seen as the recommender making inferences about which signals have more or less power to explain user interests. As such, by conducting the analysis of feedback type with respect to filter bubble formation, we aim to determine whether the recommender itself could be in part responsible for the creation of the filter bubble.

To conduct our analysis, we first filter down interactions for all users based on what form of feedback they gave to the viewed item. We consider implicit and explicit forms of positive feedback. For implicit feedback, we make use of the "click" metric, which, contrary to its name, simply means that the user has watched a video for more than half of its duration. On the other hand, for explicit feedback we consider "like", "follow", "share", and "comment". Interactions are preserved in the explicit data subset if they include one or more of the explicit forms of feedback. For implicit interactions, we specifically filter out those that have any of the explicit forms of feedback, whereas explicit does not necessarily imply a non-click.

Next, we recalculate the coverage at each time window for each user and then use the criterion defined in equation 3 to label users as "in" or "out" of the filter bubble at each time window. In Figure 8, we plot the resulting ratios of users in the bubble for each level. The ratios for all three levels are stacked on top of each other for space considerations.

Let us first assess the results for Figure 8a, explicit feedback. Immediately we notice that there is a drastic and nearly even decrease in filter bubble for all levels in the first three months of the year. Afterward, the ratio for all levels once again goes up, and within 10 weeks surpasses where it was initially. From then on, the ratios remain practically constant until the very end of the year, at which point they increase, being lifted up primarily by the increase in level 1. Overall, this is a very different trend from the overall trend we saw in Figure 4, aside from the increase of level 1 at the end. This tells us that the explicit feedback does indeed pose a relationship with the formation of the deep filter bubble. In order to define the nature of that relationship, we shall proceed to compare this trend with the one for implicit feedback as seen in Figure 8b.

The figure relating the progression of filter bubble across three layers over time using only the subset of interactions with implicit feedback is shown in Figure 8b. This figure shows yet another trend, distinct from both explicit feedback interactions in Figure 8a and the overall interactions in Figure 4. This figure shows an almost consistent downward trend in filter bubble ratio for all three category-levels from the start to week 48, at which point the filter bubble ratio for all three levels jumps. Indeed, this final jump is the

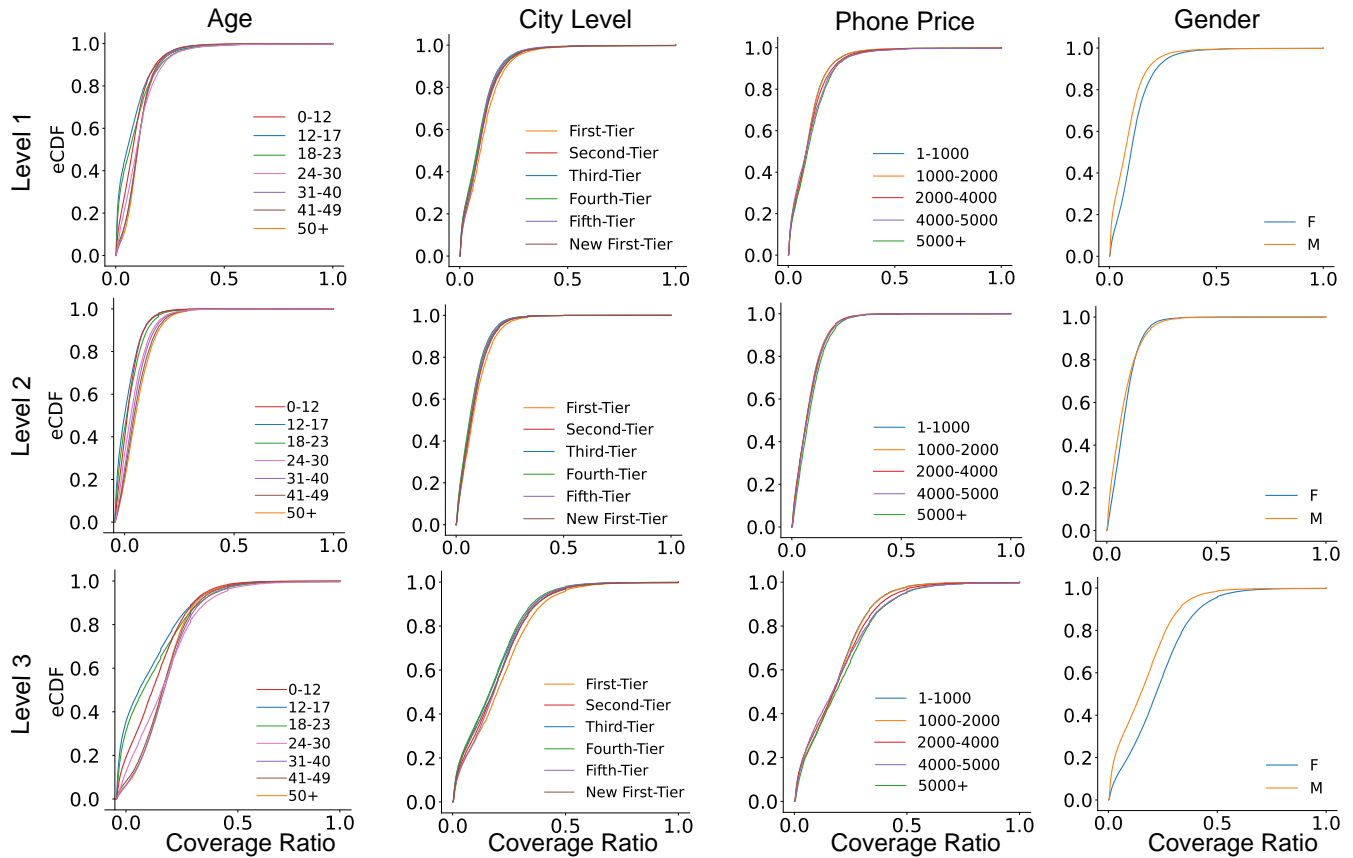


Figure 7: Empirical cumulative distribution function (eCDF) of average coverage ratio over all users at each level for four different demographic factors.

only similarity in trend across all three figures. This spike could be a result of user sampling bias due to the fact that all the users registered at the beginning of the year.

Most important for our current analysis, however, is the comparison between the trends in implicit and explicit feedback-induced deep filter bubbles. What we can conclude is that interactions with implicit feedback can lead to a decrease in the filter bubble over time in comparison with those explicit feedback interactions. This implies that the recommender is potentially leveraging explicit feedback too strongly in its modeling of the users' interests over time. While explicit feedback can clearly represent stronger preferences than implicit feedback, it is important for the recommender to moderate the impact of strong preferences so that they do not dominate the exposed content and push out items from more diverse categories. The overemphasis on modeled interests represents one of the ways that the recommender itself could be responsible for the formation of the filter bubble.

5 RELATED WORK

5.1 Filter Bubble

The concept of the filter bubble has been growing in prominence since it was first proposed in Eli Pariser's 2011 book [26], detailing

how over-personalization can lead to intellectual isolation. A subsequent study by Nguyen *et al.* [24] examines the temporal impacts of a collaborative filtering recommender on the diversity of items to better quantify the filter bubble effect. More recently, McKay *et al.* [22] attempts to shine light on the mechanisms of filter bubble from a social perspective by conducting studies on users' needs for diverse perspectives. Meanwhile, several works approach the filter bubble from a technical perspective, designing recommender models specifically to alleviate the problem. Li *et al.* [20] proposes a reinforcement learning framework for controllable recommendation to alleviate filter bubbles by adaptively selecting diverse item connections, whereas Gao *et al.* [8] proposes an interactive recommender system using reinforcement learning and causal inference to model item overexposure effects on user satisfaction. Some works attempt to explain the underlying mechanisms of filter bubbles. Li *et al.* [18] is the first one to conduct an exploratory study of the filter bubble created by short-video platforms, while [27] creates a structural model to explain their formation as a result of human and AI interactions. Several works examine the potentially harmful societal effects of filter bubbles. Bakshy *et al.* [2] examines the extent of ideological diversity in social media users' online news consumption, Bozdag *et al.* [3] explores the ethical implications of algorithmic bias in personalization, and finally, Nguyen

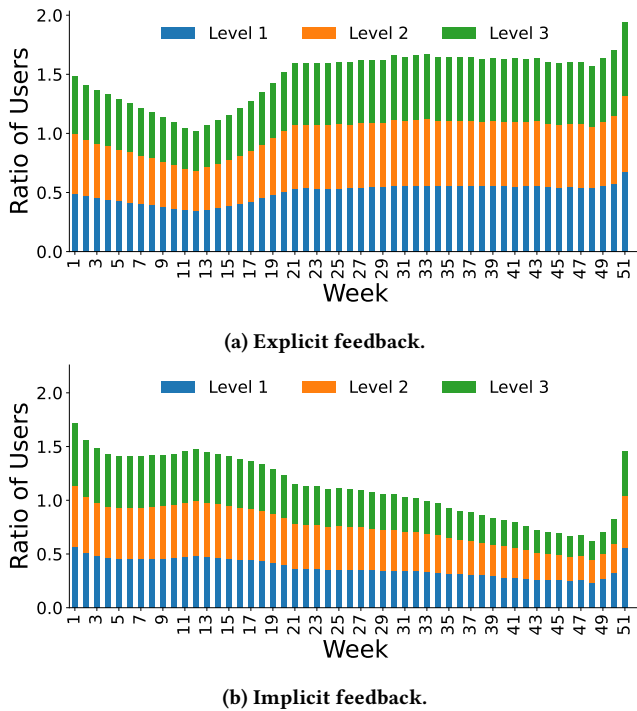


Figure 8: Progression of three-level filter development over time using only interactions with explicit and implicit feedback, respectively.

[23] explores the distinction between epistemic bubbles and echo chambers and their implications on post-truth and fake news.

5.2 Diverse Recommendation

Diversified recommendation aims to improve the recommendation list’s diversity besides the main goal of recommendation accuracy. This is usually seen as a way to counterbalance filter bubbles since the latter are characterized by low item diversity. An early but significant work is by Zhou *et al.* [37] who define an algorithm based on heat conduction to increase diversity, effectively resolving the trade-off between diversity and accuracy. Some works propose post-processing methods that rely on a backbone recommendation model. MRR [4] proposed a greedy method that can re-rank the output of any recommendation model, serving as a plug-in strategy to improve recommendation diversity. Qin *et al.* [28] use regularization terms based on backbone recommendation to ensure the entropy. DPP [5] is a popular method that uses a greedy strategy for accelerating the determinantal point process, whereas [9] and [33] further accelerate it via tree-search or factorization techniques. Li *et al.* [19] enhance diversity by increasing unexpectedness in the latent space. In contrast, DGCN, proposed by Zheng *et al.* [36] was the first to design a diversified recommendation strategy for graph neural network-based models. More recently Yang *et al.* [34] uses a concoction of techniques such as submodular neighbor selection, layer attention, and long-tail loss reweighting to outperform the state-of-the-art in diverse recommendation for GNN.

6 DISCUSSION

In the above exploratory study of the deep filter bubble on a popular short-video platform, we undergo a comprehensive study of the presence of variations in category coverage over the set of users based on time and category level. First, we define a metric and criterion that defines whether a user is “in” or “out” of a filter bubble based on whether they have been exposed to less than the median number of categories per level. Using this criterion, we discover trends in the evolution of the three-level filter bubble: the top level bubble starts out narrow and expands over time, whereas the middle level narrows and the deepest level remains mostly invariant. We also depict the probability distribution of coverages for the three levels, showing that despite having a lower relative filter bubble, deeper layers are prone to having lower coverage overall. This is due to the tree structure of the data which naturally causes narrowing of diversity per level as the categories get deeper. After qualifying the evolution of the filter bubble at each later over time, we analyze three correlating factors that could be partially responsible for the formation of the deep filter bubble: specific categories, user characteristics, and feedback type. By analyzing changes in dominant categories, we find that users may be led into a bubble based on top-layer categories identified at the beginning of their usage. Concerning user characteristics, we find that age and gender pose the largest risk of deep filter bubbles. Finally, we note that implicit feedback can serve as a mediator to the filter bubble, whereas explicit feedback can exacerbate it.

7 CONCLUSION AND FUTURE WORK

In light of the above findings, we suggest some ways that recommender systems can be designed to alleviate the deep filter bubble problem. First of all, the recommender models should not over-value preferences learned early, and instead maintain top-level diversity over time so that the user is not caught in a deep bubble due to narrow early top-level exposure. Secondly, concerning the demographic findings, we suggest that platforms introduce measures to prevent users of certain groups from being more susceptible to the deep filter bubble, namely, younger people, females, and those from lower-tier cities. Finally, we suggest that recommenders should bestow an appropriate amount of importance on various forms of feedback. For example, while explicit feedback usually signifies stronger interest, implicit feedback can also be used to learn diverse user preferences and alleviate the deep filter bubble.

For future work, we propose to delve deeper into the effect of the partial exposure of each level’s categories on the deep filter bubble. We also propose to delve into the dynamics between user-platform interactions to understand how different types of behaviors could result in different treatment by the recommender.

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China under 62272262, U23B2030, and 72342032, the National Key Research and Development Program of China under 2022YFB3104702, and the Guoqiang Institute, Tsinghua University under 2021GQG1005.

REFERENCES

- [1] Ashton Anderson, Lucas Maystre, Ian Anderson, Rishabh Mehrotra, and Mounia Lalmas. 2020. Algorithmic Effects on the Diversity of Consumption on Spotify. In *Proceedings of The Web Conference 2020 (WWW '20)*. Association for Computing Machinery, New York, NY, USA, 2155–2165. <https://doi.org/10.1145/3366423.3380281>
- [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] Engin Bozdog. 2013. Bias in algorithmic filtering and personalization. *Ethics and Information Technology* 15, 3 (Sept. 2013), 209–227. <https://doi.org/10.1007/s10676-013-9321-6>
- [4] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. 335–336.
- [5] Laming Chen, Guoxin Zhang, and Eric Zhou. 2018. Fast Greedy MAP Inference for Determinantal Point Process to Improve Recommendation Diversity. In *Advances in Neural Information Processing Systems*, Vol. 31. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2018/hash/dbbf603ff0e99629dda5d75b6f75f966-Abstract.html
- [6] 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), e2023301118.
- [7] Seth Flaxman, Sharad Goel, and Justin M. Rao. 2016. Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly* 80, S1 (2016), 298–320. <https://doi.org/10.1093/poq/nfw006>
- [8] Chongming Gao, Shiqi Wang, Shijun Li, Jiawei Chen, Xiangnan He, Wenqiang Lei, Biao Li, Yuan Zhang, and Peng Jiang. 2023. CIRS: Bursting Filter Bubbles by Counterfactual Interactive Recommender System. *ACM Transactions on Information Systems* (April 2023). <https://doi.org/10.1145/3594871> Just Accepted.
- [9] Jennifer Gillenwater, Alex Kulesza, Zeldia Mariet, and Sergei Vassilvitskii. 2019. A Tree-Based Method for Fast Repeated Sampling of Determinantal Point Processes. In *International Conference on Machine Learning*. 2260–2268.
- [10] Ehsan-Ul Haq, Gareth Tyson, Tristan Braud, and Pan Hui. 2022. Weaponising Social Media for Information Divide and Warfare. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media*. ACM, Barcelona Spain, 259–262. <https://doi.org/10.1145/3511095.3536372>
- [11] Adam N. Hornsby and Bradley C. Love. 2020. How decisions and the desire for coherency shape subjective preferences over time. *Cognition* 200 (July 2020), 104244. <https://doi.org/10.1016/j.cognition.2020.104244>
- [12] Homa Hosseinmardi, Amir Ghasemian, Aaron Clauset, Markus Mobius, David M. Rothschild, and Duncan J. Watts. 2021. Examining the consumption of radical content on YouTube. *Proceedings of the National Academy of Sciences* 118, 32 (Aug. 2021), e2101967118. <https://doi.org/10.1073/pnas.2101967118> Publisher: Proceedings of the National Academy of Sciences.
- [13] Elvin Isufi, Matteo Pocchiari, and Alan Hanjalic. 2021. Accuracy-diversity trade-off in recommender systems via graph convolutions. *Information Processing & Management* 58, 2 (March 2021), 102459. <https://doi.org/10.1016/j.ipm.2020.102459>
- [14] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2010. *Recommender systems: an introduction*. Cambridge University Press.
- [15] Marius Kaminskis and Derek Bridge. 2016. Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiIS)* 7, 1 (2016), 1–42.
- [16] Brent Kitchens, Steve L. Johnson, and Peter Gray. 2020. Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption. *MIS Quarterly* 44, 4 (Dec. 2020), 1619–1649. <https://doi.org/10.2530/MISQ/2020/16371>
- [17] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems—A survey. *Knowledge-based systems* 123 (2017), 154–162.
- [18] Nian Li, Chen Gao, Jinghua Piao, Xin Huang, Aizhen Yue, Liang Zhou, Qingmin Liao, and Yong Li. 2022. An Exploratory Study of Information Cocoon on Short-form Video Platform. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. ACM, Atlanta GA USA, 4178–4182. <https://doi.org/10.1145/3511808.3557548>
- [19] Pan Li and Alexander Tuzhilin. 2020. Latent Unexpected Recommendations. <http://arxiv.org/abs/2007.13280> arXiv:2007.13280 [cs].
- [20] Zhenyang Li, Yancheng Dong, Chen Gao, Yizhou Zhao, Dong Li, Jianye Hao, Kai Zhang, Yong Li, and Zhi Wang. 2023. Breaking Filter Bubble: A Reinforcement Learning Framework of Controllable Recommender System. In *Proceedings of the ACM Web Conference 2023*. ACM, Austin TX USA, 4041–4049. <https://doi.org/10.1145/3543507.3583856>
- [21] Feng Lu and Nava Tintarev. 2018. A Diversity Adjusting Strategy with Personality for Music Recommendation. In *InRS@ RecSys*. 7–14.
- [22] Dana McKay, Kaipin Owyong, Stephann Makri, and Marisela Gutierrez Lopez. 2022. Turn and face the strange: investigating filter bubble bursting information interactions. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*. 233–242.
- [23] C. Thi Nguyen. 2020. ECHO CHAMBERS AND EPISTEMIC BUBBLES. *Episteme* 17, 2 (June 2020), 141–161. <https://doi.org/10.1017/epi.2018.32>
- [24] 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 (WWW '14)*. Association for Computing Machinery, New York, NY, USA, 677–686. <https://doi.org/10.1145/2566486.2568012>
- [25] Yunzhu Pan, Chen Gao, Jianxin Chang, Yanan Niu, Yang Song, Kun Gai, Depeng Jin, and Yong Li. 2023. Understanding and Modeling Passive-Negative Feedback for Short-video Sequential Recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*. ACM, Singapore Singapore, 540–550. <https://doi.org/10.1145/3604915.3608814>
- [26] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.
- [27] Jinghua Piao, Jiazhen Liu, Fang Zhang, Jun Su, and Yong Li. 2023. Human-AI adaptive dynamics drives the emergence of information cocoons. *Nature Machine Intelligence* (Oct. 2023), 1–11. <https://doi.org/10.1038/s42256-023-00731-4> Publisher: Nature Publishing Group.
- [28] Lijing Qin and Xiaoyan Zhu. [n.d.]. Promoting Diversity in Recommendation by Entropy Regularizer. ([n.d.]).
- [29] Peter S. Riefer, Rosie Prior, Nicholas Blair, Giles Pavey, and Bradley C. Love. 2017. Coherency Maximizing Exploration in the Supermarket. *Nature human behaviour* 1 (Jan. 2017), 0017. <https://doi.org/10.1038/s41562-016-0017>
- [30] Francesco Sanna Passino, Lucas Maystre, Dmitrii Moor, Ashton Anderson, and Mounia Lalmas. 2021. Where To Next? A Dynamic Model of User Preferences. In *Proceedings of the Web Conference 2021 (WWW '21)*. Association for Computing Machinery, New York, NY, USA, 3210–3220. <https://doi.org/10.1145/3442381.3450028>
- [31] Fernando P. Santos, Yphtach Lelkes, and Simon A. Levin. 2021. Link recommendation algorithms and dynamics of polarization in online social networks. *Proceedings of the National Academy of Sciences* 118, 50 (Dec. 2021), e2102141118. <https://doi.org/10.1073/pnas.2102141118>
- [32] Vitor V. Vasconcelos, Sara M. Constantino, Astrid Dannenberg, Marcel Lumkowsky, Elke Weber, and Simon Levin. 2021. Segregation and clustering of preferences erode socially beneficial coordination. *Proceedings of the National Academy of Sciences* 118, 50 (Dec. 2021), e2102153118. <https://doi.org/10.1073/pnas.2102153118>
- [33] Romain Warlop, Jérémie Mary, and Mike Gartrell. 2019. Tensorized Determinantal Point Processes for Recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1605–1615.
- [34] Liangwei Yang, Shengjie Wang, Yunzhe Tao, Jiankai Sun, Xiaolong Liu, Philip S. Yu, and Taiqing Wang. 2023. DGRec: Graph Neural Network for Recommendation with Diversified Embedding Generation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. 661–669. <https://doi.org/10.1145/3539597.3570472> arXiv:2211.10486 [cs].
- [35] Robert B. Zajonc. 1968. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology* 9, 2, Pt.2 (1968), 1–27. <https://doi.org/10.1037/h0025848>
- [36] Yu Zheng, Chen Gao, Liang Chen, Depeng Jin, and Yong Li. 2021. Dgen: Diversified recommendation with graph convolutional networks. In *Proceedings of the Web Conference 2021*. 401–412.
- [37] Tao Zhou, Zoltán Kuscik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences* 107, 10 (March 2010), 4511–4515. <https://doi.org/10.1073/pnas.1000488107>