

Filter bubbles created by collaborative filtering algorithms themselves, fact or fiction? An experimental comparison

Flore Vancompernelle Vromman

flore.vancompernelle@uclouvain.be

Louvain Institute of Data Analysis and Modeling in
economics and statistics, Université catholique de Louvain
Louvain-la-Neuve, Belgium

François Fouss

francois.fouss@uclouvain.be

Louvain Research Institute in Management and
Organizations, Université catholique de Louvain
Mons, Belgium

ABSTRACT

By recommending only, or mostly, items close to our *supposedly known* tastes, a recommender system could lead to the narrowing of the diversity/novelty of the recommended list, leading to the creation of so-called filter bubbles. Authors however disagree on the real existence of filter bubbles and, if real, on the responsibility of algorithms in the process of their creation. This paper aims to experimentally analyze the potential creation of such filter bubbles by testing various collaborative filtering algorithms on their intrinsic propensity of changing the diversity/novelty of recommended items, in configurations where users behave, with various intensities, in a way that could influence the evolution in diversity/novelty. Results show that both human and algorithmic components could have an impact on diversity/novelty, depending on the type of algorithm.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative filtering**; **Empirical studies in collaborative and social computing**.

KEYWORDS

Recommender System, Filter Bubble, Collaborative Filtering

ACM Reference Format:

Flore Vancompernelle Vromman and François Fouss. 2021. Filter bubbles created by collaborative filtering algorithms themselves, fact or fiction? An experimental comparison. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '21 Companion)*, December 14–17, 2021, ESSENDON, VIC, Australia. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3498851.3498945>

1 INTRODUCTION

For years now, the main aim of recommender systems has been to be as *accurate* as possible, as the evaluation of the quality of a recommender system was made through accuracy metrics (such a metric evaluates the relevance of the recommended items by taking a distance between predicted evaluations and available evaluations of users). By recommending only, or mostly, items close to our

supposedly known tastes (as determined by the available data, which however only represent a small part of the real tastes of a user), a recommender system could thus lead to the narrowing of the diversity/novelty of the recommended list, leading to the creation of so-called filter bubbles. In their work, [11, 15, 16], some authors have argued that the reliance on personalization and social media can lead people to being exposed to a narrow set of viewpoints, locking people, and preventing them from engaging with ideas different from their own [10]. Pariser argued in [11] that filter bubbles introduce three dynamics: first, you are alone in it, second, the filter bubble is invisible, and third, you do not choose to enter the bubble.

In addition to narrowing individual viewpoints, other consequences of such filter bubbles include the individual's loss of their ability to choose how they want to live [11], the lack of "bridging" social capital [12], a biased access to information [11, 14], the reduction of creativity and innovation [11], the creation of a fertile ground for polarization and misinformed opinions [13], or even discrimination (e.g., racial or gender discrimination) [14].

Authors however disagree on the real existence of filter bubbles and, if real, on the responsibility of algorithms in the process of their creation.

However, in [3], the authors analyze filtering processes on three different levels: the individual ("individuals are to different extents cognitively motivated to search for and add fitting bits of information and to ignore or deny conflicting ones"), the social ("human beings display a tendency to form friendships and other social network structures preferably with people with whom they share sociodemographic, behavioral, and intrapersonal characteristics"), and the technological ("stronger automatic filtering leads to a decreased variety of information that is offered to individuals") levels. In [16], Sunstein states that recommender systems contribute to fragmentation and polarization of society. Nguyen et al. [9] found that, by analyzing the effects of a collaborative filtering-based recommender system on users' range of interests, the average movie diversity decreased over time, but the effect was stronger for those users that did not usually follow the recommendations than for those who did it frequently. Also, in a study on the effects of a music platform's recommender system, Hosanagar and colleagues found little empirical evidence for fragmentation over time [5], but rather on an increase in the commonality between the users, showing that the recommender system helps users explore new items and expand their preferences.

To provide further insights on the existence of filter bubbles as well as on their emergence process, this paper aims to experimentally analyze the potential creation of such filter bubbles by

Permission to make digital or hard copies of all or part 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 components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WI-IAT '21 Companion, December 14–17, 2021, ESSENDON, VIC, Australia

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-9187-0/21/12...\$15.00

<https://doi.org/10.1145/3498851.3498945>

testing various collaborative filtering algorithms on their intrinsic propensity of changing the diversity/novelty of recommended items, in configurations where users behave, with various intensities, in a way that could influence the evolution in diversity/novelty. Collaborative filtering algorithms are based on the definition of neighborhoods, either of items or of users, to recommend items to users.

The research question experimentally addressed by this work could therefore be formulated as: *Do collaborative filtering algorithms, by themselves (i.e., by their internal functioning and without any human interaction), lead to the narrowing of the diversity/novelty of recommended lists?*

Section 2 presents the experimental settings and methodology. Results are shown and discussed in Section 3, while Section 4 concludes the paper.

2 EXPERIMENTAL SETTINGS

All the developments were made by using the Python 3 programming language.

2.1 Dataset

The data used in this paper are a MovieLens data set containing ratings, from 1 to 5, assigned by users to movies. The full data set includes 20,000,263 ratings assigned by 138,493 users to 27,278 movies and is provided by GroupLens [4] (see “MovieLens 20M Dataset”).

Each tuple contains the id of the user who gave the rating, the id of the rated movie, the rating and the timestamp, which refers to the time the rating was given.

2.2 Algorithms

This work focuses on Collaborative Filtering (CF) algorithms, and therefore compares 4 of such algorithms, described in this section: an item-based CF algorithm (ICF), a user-based CF algorithm (UCF), a CF algorithm incorporating more diversity (DivCF), and a CF algorithm incorporating more novelty (NovCF). Two baseline algorithms, an algorithm offering random recommendations (which does not require further details), and an algorithm recommending popular items, are also reported.

To recommend n items to users, CF algorithms explore and compare the known preferences of each user (i.e., their purchase history), on the basis that similarities of the past should still be effective in the future [8]. In this work, these preferences are saved in a so-called adjacency matrix, $\mathbf{A}$, where each row corresponds to a user u and each column to an item i . An element a_{ui} of the matrix takes a value of 1 if user u watched movie i (no matter the rating), and 0 otherwise.

2.2.1 Item-based CF algorithm (ICF). An item-based CF algorithm is based on the existing similarities between the items. These similarities could be computed through different measures: Pearson correlation, cosine similarity, etc. [7, 8]. In this work, the cosine similarity between each pair of items i, j is used [1], computed by

$$\text{sim}(i, j) = \text{Cos}(i, j) = \frac{\sum_{u \in U_{ij}} a_{ui} a_{uj}}{\sqrt{\sum_{u \in U_i} a_{ui}^2} \sqrt{\sum_{u \in U_j} a_{uj}^2}}, \quad (1)$$

where U_{ij} is the set of users who consumed both items i and j , U_i (U_j) is the set of users who consumed item i (j).

These similarities are then used to identify the k -nearest neighbors of each item i , which correspond to the k items with the highest similarity scores with item i .

The list of n items recommended to user u is chosen among the items not yet consumed by user u and contains the n items closest to the set of items already consumed by user u .

2.2.2 User-based CF algorithm (UCF). A user-based CF algorithm is based on the existing similarities between the users. The similarity between each pair of users u, v is computed here by

$$\text{sim}(u, v) = \text{Cos}(u, v) = \frac{\sum_{i \in I_{uv}} a_{ui} a_{vi}}{\sqrt{\sum_{i \in I_u} a_{ui}^2} \sqrt{\sum_{i \in I_v} a_{vi}^2}}, \quad (2)$$

where I_{uv} is the set of items consumed by both users u and v , I_u (I_v) is the set of items consumed by user u (v).

Based on the similarities between each pair of users, the algorithm identifies the k -nearest neighbors of each user, and the list of recommendations is further built for each user u by adding the items not yet consumed by user u and mostly consumed by the k -nearest neighbors of user u , each nearest neighbor contribution being weighted by his similarity with user u .

2.2.3 CF algorithm incorporating more diversity (DivCF). This third algorithm refers to the CF algorithm developed by [2], designed to incorporate more diversity into the recommendations offered to users.

Firstly, the algorithm creates a temporary list of $5 * n$ items that could be recommended to the user, based on the classical UCF presented above. The first element of this temporary list is added to the list of recommendations for the user. Then, items are added one by one to the list of recommendations by choosing the item offering the largest diversity with the items already included in the list of recommendations. The diversity between m items is measured as follows, where $\text{sim}(i_j, i_k)$ corresponds to the cosine similarity between items j and k ,

$$\text{Div}(i_1, \dots, i_m) = \frac{\sum_{j=1..m-1} \sum_{k=j+1..m} (1 - \text{sim}(i_j, i_k))}{\frac{m}{2}(m-1)} \quad (3)$$

The algorithm stops when the number of items in the list of recommendations reaches n .

2.2.4 CF algorithm incorporating more novelty (NovCF). In the same way as the previous algorithm (focused on diversity), the fourth one aims at incorporating more novelty, instead of diversity. Such as the previous algorithm, it creates a temporary list of $5 * n$ items by using the UCF. Then, n items are selected from this temporary list to constitute the list of recommendations by choosing the most novel items. The novelty of an item j , $\text{Nov}(j)$, ranges between 0 and 1, and is computed by Equation (4), where $\text{pop}(i_j)$ is the degree of popularity of item i_j . In this work, $\text{pop}(i_j)$ corresponds to the number of users who consumed item i_j divided by the total number of users in the database [6].

$$\text{Nov}(j) = 1 - \text{pop}(i_j) \quad (4)$$

2.2.5 Algorithm recommending popular items. This algorithm recommends popular items to users. Again, the popularity relates to the number of views of an item: the more an item is viewed, the more it is considered as popular. Firstly, the algorithm counts the number of views of each item. Then, the items are ordered by growing popularity, the most popular items being recommended to all the users.

2.3 Metrics

In order to evaluate and compare the recommendations established by the recommender algorithms described in Section 2.2, we use two metrics closely related to the potential enclosure of users in their dominant preferences: a diversity and a novelty score. Both scores are computed within the list of recommendations offered by an algorithm for a user, and are further averaged over the users to obtain an aggregated novelty/diversity score.

Diversity. Equation (3) is used to measure the diversity between all items included in a list of recommendations.

Novelty. The novelty of the n items included in a list of recommendations is computed by Equation (5), where, again, $pop(i_j)$ is set to the degree of popularity of item j , which corresponds to the number of users who consumed item j during the period studied divided by the number of users within the database.

$$Novelty(i_1, \dots, i_n) = \frac{\sum_{j=1..n} (1 - pop(i_j))}{n} \quad (5)$$

2.4 Methodology and scenarios

2.4.1 Data preprocessing. To experimentally analyze the potential gradual creation of filter bubbles, the data set is split into subsets, based on timestamps. More precisely, the data set is first ordered on a chronological timestamp basis (see Figure 1, Step 1), and then split into x subsets (different values of x , the number of time periods, were tested – see later) (see Figure 1, Step 2). Moreover, to analyze the evolution of the recommendations offered to each user, the data related to users who are not present in all subsets is removed (see Figure 1, Step 3).

Finally, an adjacency matrix (where columns represent the items and rows the users) is built for the first subset so that algorithms can be applied to recommend items to users (see Figure 1, Step 4).

This adjacency matrix evolves over the following time periods, depending on the users' behavior in the acceptance of recommended items (see Subsection 2.4.3 for details about the various explored scenarios). Indeed, after each recommendation period, each item that a user will watch will be added as "seen/watched" in his purchase history considered as basis of the next period, leading to changing the corresponding element of the adjacency matrix from 0 to 1. Then, at the next recommendation period, new recommendations are suggested based on this new adjacency matrix, and so on.

Choice of the number of time periods. First remember that each user must be present in all subsets (i.e., all time periods) to be able to analyze the evolution of the suggested recommendations over time periods. Keeping in mind that we need to have enough users for a robust analysis, we decided to divide our data set into 5 subsets / time periods. Indeed, the values of 5, 10 and 20 were tested but

the number of users present on each period with 10 or 20 time periods was insufficient for the relevance of the analysis. With 5 time periods, the same set of 335 different users is present in each subset / time period.

Choice of the number of recommendations. We tested the values of 10, 20 and 100. As expected, the diversity and the novelty within the recommendations are known to be higher (or stable for the algorithm suggesting random recommendations) with a higher number of recommendations. It is however important to consider that users do not consult necessarily all the suggested recommendations, as shown in the literature, where the number of recommendations often do not exceed 20 (see, e.g., [2, 9]). We therefore fix the number of recommendations at 20 items, based on the literature.

2.4.2 Human and algorithmic components. As the items recommended by any system depend on both the purchase history of users and the paradigm of the algorithm, it is important to distinguish the *human component* from the *algorithmic component* in those processes.

The *human component* is related to the own choices of the user, independently of potential suggested recommendations. In our data set, each subset contains items that users have consumed in a time period. The adjacency matrix therefore evolves due to this human component, by periodically (i.e., for each subset) completing the matrix with all the data associated to the next time period (therefore completing the purchase history of each user with the *newly* (i.e., regarding the previous period) watched movies).

On the other hand, the *algorithmic component* is related to the behavior of users when facing suggested recommendations, as users can watch all or some of the movies recommended by the algorithm.

Various behaviors of users were analyzed in this work, as detailed in the next subsection.

2.4.3 Scenarios. To compare different potential users' behaviors, 5 scenarios were analyzed:

- The first scenario (S1) incorporates both the human and the algorithmic components with an acceptance of 100% of the recommendations regarding the algorithmic component. This percentage may be considered as "too optimistic" compared to reality, but it will be used as a maximum threshold of the influence of algorithms.
- The second scenario (S2) also deals with both the human and the algorithmic components but, each recommendation has a probability p to be considered as "accepted/watched" by the user. Several values of p could be tested, but it is not the focus of our study. Therefore, we put the value of p to 25%, as this value represents better the reality of the number of movies users really watch among the recommendations [17]. It means that each recommendation is accepted according to a probability of 25%.
- The third scenario (S3) only includes the human component, therefore allowing us to analyze the effect of human choices on the recommendations without any influence of any algorithm.
- The final two scenarios only incorporate the algorithmic component, by considering that users only watch recommended movies, where the probability of acceptance of each

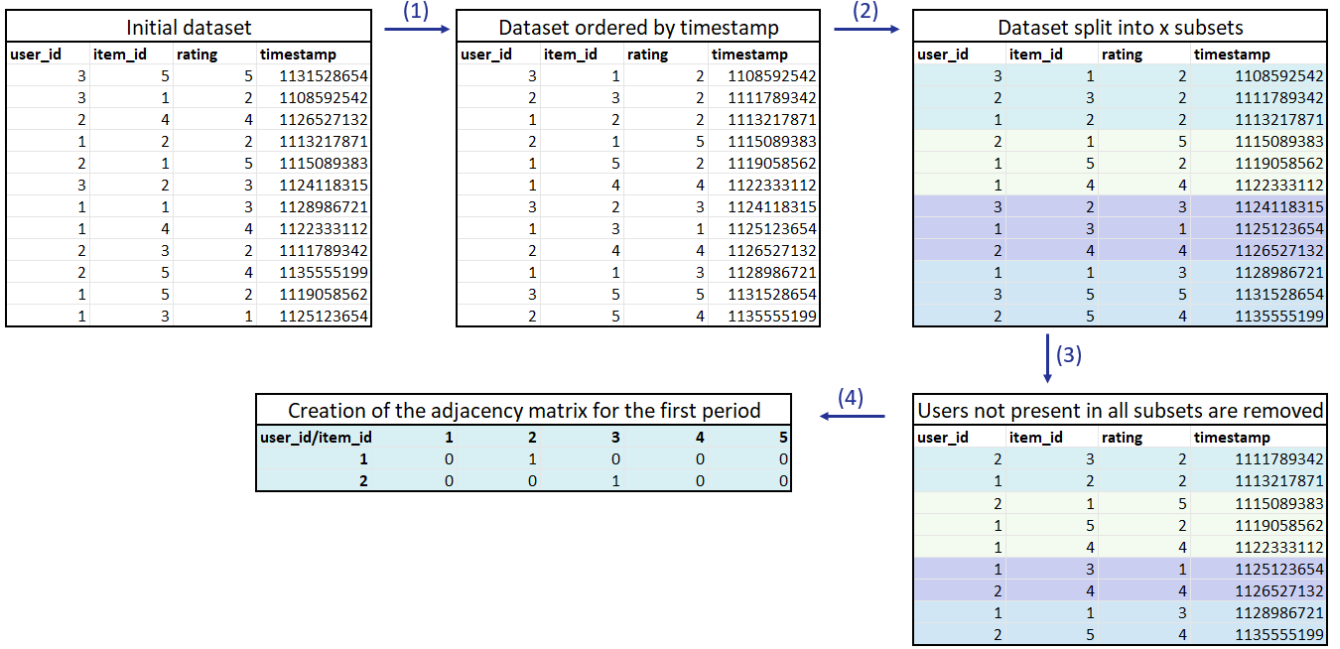


Figure 1: Data preprocessing.

recommendation is 100% for the fourth scenario (S4), and 25% for the fifth one (S5).

3 RESULTS AND DISCUSSION

This section first analyzes the global evolution of the diversity and the novelty within the recommendations through time for each considered algorithm, and then analyzes the impact of the human and algorithmic components on the recommender systems.

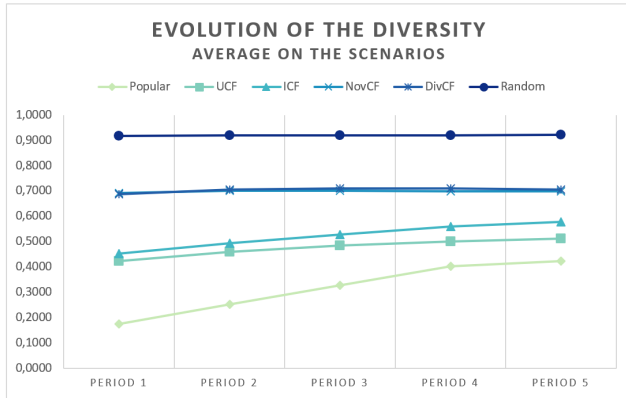


Figure 2: Diversity scores (averaged on the scenarios) for each algorithm, over the time periods.

3.1 Temporal evolution of the metrics

Three algorithms (i.e., random, DivCF and NovCF) have a significantly higher level of diversity already in the first time period,

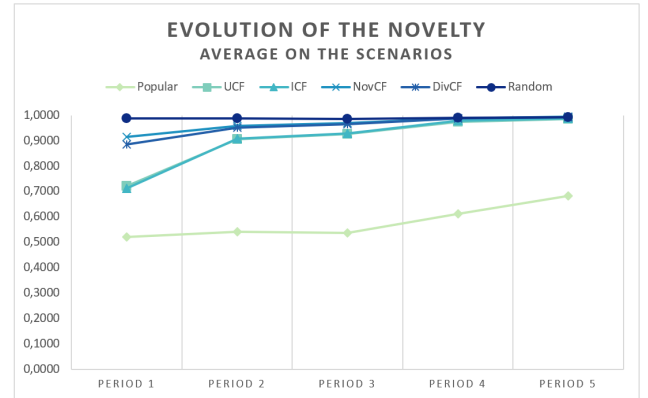


Figure 3: Novelty scores (averaged on the scenarios) for each algorithm, over the time periods.

compared to the three others (i.e., ICF, UCF and popular) (see Figure 2). As expected, the algorithm showing the highest diversity (but the lowest personalization), no matter the time period (1 to 5) or the scenario (S1 to S5), is the one recommending random items. The two following (i.e., in terms of high diversity) algorithms are DivCF and NovCF. For these three recommender systems presenting a higher diversity since the first time period, the level of diversity is rather stable through time.

However, the other three algorithms (i.e., ICF, UCF, popular recommendations) have a smaller level of diversity at period 1, but this level increases through time (see Figure 2), no matter the scenario applied. Despite this increase, it is important to note that

the level of diversity within the recommendations is still smaller for those recommender systems than the one offered by the other three algorithms, even at the last period (see Figure 2). Moreover, the degree of increase is smaller through time for the algorithms where we observe an increase through time.

Regarding the novelty, except that the start level of this metric (i.e., in the first period) is higher than the start level of diversity, similar conclusions apply (see Figure 3): the level of novelty increases through time for UCF and ICF and is rather stable for DivCF, NovCF and random recommendations. Note that the level of novelty starts to increase only after the third period for the algorithm recommending popular items.

3.2 Analysis of the human and algorithmic components

We now focus on the analysis of the responsibility of the users and algorithms on the level of diversity/novelty within the suggested recommendations, by analyzing the different scenarios of users' behaviors. Note that the results obtained with novelty are very similar to the ones obtained by diversity, and are therefore further highlighted in the sequel only if different.

3.2.1 The algorithmic component may increase the diversity. Two algorithms present a quite similar behavior faced to the users' behaviors: ICF and UCF. For those algorithms, the diversity grows more significantly when S1 or S4 are applied compared to the other scenarios of users' behavior (see Figure 4 for ICF and Figure 5 for UCF).

As a reminder, S1 integrates the algorithmic and the human components while S4 integrates only the algorithmic component. But, the common point of S1 and S4 is that we consider, through the algorithmic component, that users watch all the movies recommended. It highlights the importance of the algorithmic component for those algorithms because a higher diversity is provided within the recommendations when users watch all the recommended movies (i.e., in S1 and S4) compared to when each recommendation is watched with only a probability of 25% (i.e., in S2 and S5). To support the importance of the algorithmic component on the increase of diversity for UCF and ICF, we may notice that the worst scenario is when users refuse all of the recommendations and thus, watch only movies of their choices (i.e., in S3). So, the higher is the number of recommendations consumed by the users, the higher will be the diversity in the future.

Even if S1 and S4 present a higher level of diversity (for UCF and ICF) compared to other scenarios, we can distinguish them. Indeed, Figures 4 and 5 show that S1 offers the best overall level of diversity. This result may induce that the human component, only paired with the algorithmic one, may generate a higher diversity.

To sum up, if the algorithm is based on the purchase histories (as is the case of UCF and ICF), *the users contribute more than the recommender systems to enclose them in their prevailing preferences*. Indeed, we can see that following all the recommendations suggested by those algorithms seems to increase the diversity within the list of recommendations. However, we may nuance this statement by adding that users who accept all the recommendations may still increase the diversity within the recommendations by consuming other movies in addition to the recommended ones.

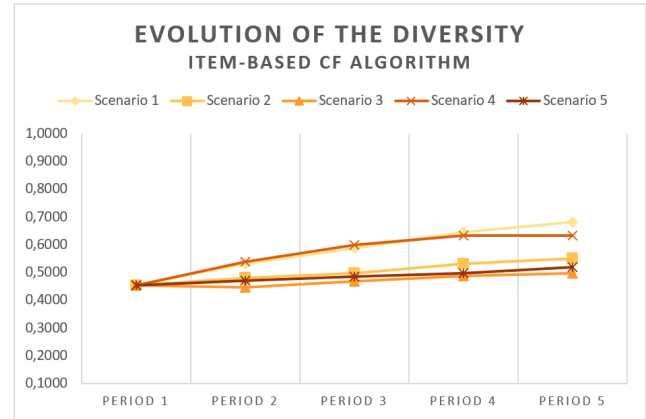


Figure 4: Diversity within the recommendations of ICF for each scenario, over the time periods.

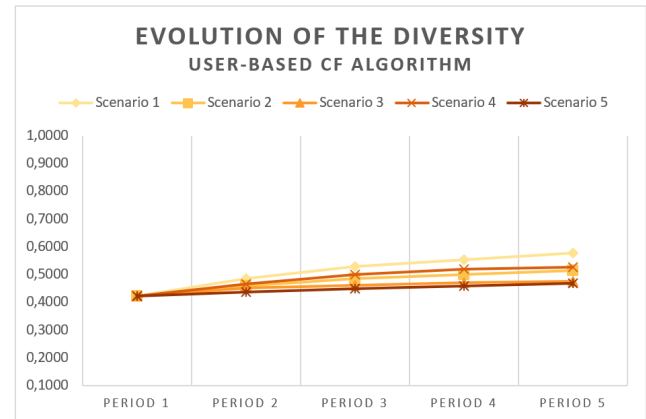


Figure 5: Diversity within the recommendations of UCF for each scenario, over the time periods.

3.2.2 The human component may increase the diversity. The other studied algorithms lead however to different conclusions.

Popular recommendations. For the algorithm recommending popular items, we can clearly identify a distinction between the scenarios with and without the human component (see Table 1, where the level of diversity within the suggested recommendations is represented for each scenario and time period). As mentioned earlier, we see that the diversity grows through time no matter the scenario of users' behavior applied. However, there is a greater increase for the first three scenarios, which include the human component, compared to the last two scenarios. So, for this algorithm, *users play a role in the establishment of more diverse lists of recommendations* by watching movies not recommended to them in addition to or instead of the recommended ones.

DivCF and NovCF. Regarding now DivCF and NovCF, by looking closer to each scenario, we can identify a slight increase of the diversity through time for some scenarios or a slight decrease for others (see Table 2 for DivCF and Table 3 for NovCF). Such as

Table 1: Diversity within the recommendations of the algorithm recommending popular items.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Period 1	0.1739	0.1739	0.1739	0.1739	0.1739
Period 2	0.2813	0.2813	0.2812	0.2100	0.2100
Period 3	0.3577	0.3575	0.3571	0.2821	0.2821
Period 4	0.4186	0.4186	0.4186	0.3792	0.3792
Period 5	0.4404	0.4403	0.4403	0.3992	0.3992

the algorithm recommending popular items, those two algorithms provide better results, in terms of evolution of the diversity, when the human component is present (i.e., S1, S2 and S3). So, it highlights the importance of the human component. In other words, diversity is lower when algorithms hinder users from accessing to items by themselves (i.e., S4, S5). This statement is further enhanced by the fact that the scenario including only the human component (i.e., S3) is always better than the one including only a pure algorithmic component (i.e., S4).

Moreover, we can explore deeper the role of the algorithmic component for those algorithms by comparing the probability of acceptance of the recommendations (i.e., S1 versus S2, and S4 versus S5). For those two algorithms (i.e., DivCF and NovCF), we can see in Tables 2 and 3 that the diversity is higher when recommendations are consumed with a probability of 25% compared to a total acceptance of the recommendations (i.e., S2 is better than S1, and S5 is better than S4). It demonstrates that it is important not to give too much weight to the algorithmic component.

Finally, we may notice that the best scenario for DivCF and NovCF, in terms of diversity, is S2 (see Tables 2 and 3), which means that the human component is essential but, the presence of the algorithmic component sparingly also increases the diversity. So, we can conclude that algorithms incorporating naturally a higher diversity or novelty (i.e., DivCF and NovCF) provide more diverse recommendations when users *watch movies not recommended to them in addition to a part of the recommended ones*.

Table 2: Diversity within the recommendations of the DivCF.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Period 1	0.6880	0.6880	0.6880	0.6880	0.6880
Period 2	0.7194	0.7162	0.7099	0.6835	0.6940
Period 3	0.7257	0.7337	0.7222	0.6768	0.6940
Period 4	0.7227	0.7391	0.7244	0.6663	0.6960
Period 5	0.7138	0.7382	0.7251	0.6629	0.6898

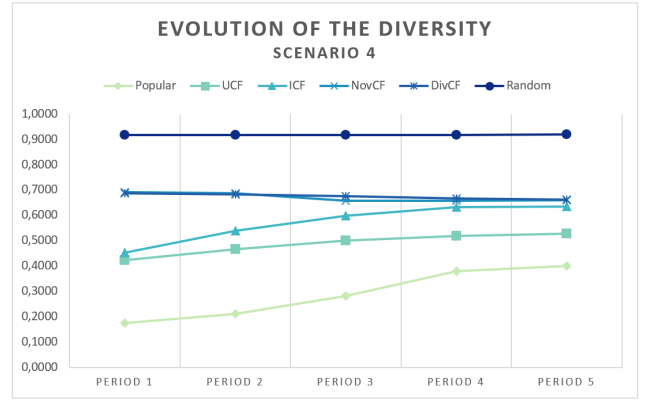
3.3 Discussion

It is important to note that, no matter the scenario applied, DivCF and NovCF always offer a higher diversity (in absolute terms) compared to UCF, ICF and the algorithm recommending popular items. More precisely, our results show that S4 is one of the best scenarios, in terms of evolution of the diversity, for ICF and UCF while it is the worst one for DivCF and NovCF. However, Figure 6 shows that the level of diversity is still higher for DivCF and NovCF than for

Table 3: Diversity within the recommendations of the NovCF.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Period 1	0.6911	0.6911	0.6911	0.6911	0.6911
Period 2	0.7145	0.7086	0.6999	0.6862	0.6970
Period 3	0.7166	0.7251	0.7083	0.6582	0.6979
Period 4	0.7045	0.7265	0.7071	0.6578	0.7001
Period 5	0.7056	0.7247	0.7063	0.6592	0.6928

UCF and ICF, even if the level of diversity in period 5 for ICF is close to the level of diversity of the other two algorithms. So, it proves that, even in the worst scenario, DivCF and NovCF offer more diverse lists of recommendations to the users.

**Figure 6: Diversity in scenario 4 for each recommender system.**

Another finding concerns the combination of the algorithmic and the human components. For all algorithms, we identified that the best overall scenario is one including the human component and the algorithmic component (with an acceptance probability of 25% or 100% depending on the algorithm). So, we could conclude that both of the components are necessary to increase the diversity.

Table 4 sums up the main points of our analysis by highlighting, for each algorithm, the main factor promoting diversity within the recommendations. It also presents the additional factor (if there is one) that, coupled to the first one, keep on increasing the diversity.

4 CONCLUSIONS AND FUTURE RESEARCH

Our answer to our research question (*Do collaborative filtering algorithms, by themselves (i.e., by their internal functioning and without any human interaction), lead to the narrowing of the diversity/novelty of recommended lists?*) is clearly "No".

Indeed, results show that, without any external interaction, none of the studied CF algorithms progressively leads to the narrowing of diversity/novelty in the recommended list, and therefore to the creation of any filter bubble, even if such CF algorithms are designed to offer recommendations that increasingly stick to the profile of users. It therefore seems that a collaborative filtering world where algorithms decide by themselves to lock us in a filter

Table 4: Main points from the analysis of the results.

Algorithm	Evolution of diversity through time	Main factor promoting diversity	Additional factor promoting diversity
<i>ICF and UCF</i>	Positive	Acceptation of all the recommendations	The users watch movies not recommended in addition to the recommended ones
<i>DivCF and NovCF</i>	Stable	The users watch movies not recommended to them	Acceptation of a part of the recommendations
<i>Popular recommendations</i>	Positive	The users watch movies not recommended to them in addition to or instead of the recommended ones	/
<i>Random recommendations</i>	Stable	/	/

bubble is unlikely. More precisely, in the case of CF algorithms a priori designed to be biased through more diversity/novelty (such as DivCF or NovCF), the human component is important to still improve diversity/novelty over time. Moreover, if the CF algorithm is a *classical* one (such as UCF or ICF), the algorithmic component still helps to increase the diversity/novelty over time (i.e., users should accept part of the recommendations, with a probability of at least 25% to improve diversity/novelty). In other words, in this case, the algorithmic component is more *open-minded* than users themselves. Finally, regardless of the type of CF algorithm, best results (in terms of an increase in diversity/novelty) are always obtained by a hybrid strategy (i.e., including both human and algorithmic components) rather than a pure human strategy or a pure algorithmic strategy.

To confirm these preliminary insights, future work will include a similar study on a data set with more users, which could be divided in more than 5 time periods, and testing other types of algorithms than CF ones. The apparent stabilization of diversity/novelty through time could also be analyzed more deeply. It could also be interesting to apply our study on another data set and compare the results.

Moreover, future work could focus on the analysis of another way to compute the diversity. Indeed, in our analysis, the computation of the diversity is based on similarities, which are calculated through the users' ratings. This measure of diversity, most often used in the literature, could then introduce a bias in our analysis since the algorithms base also their recommendations on the past users' ratings. Therefore, it would be interesting to compare our results with another measure of diversity, for example, based on the categories of the movies or on the tags associated with the movies.

REFERENCES

- [1] John S. Breese, David Heckerman, and Carl Kadie. 2013. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. arXiv:1301.7363 [cs.LG]
- [2] Sylvain Castagnos, Armelle Brun, and Anne Boyer. 2014. La diversité: entre besoin et méfiance dans les systèmes de recommandation. *Revue d'Intelligence Interaction Intelligence* (2014).
- [3] Daniel Geschke, Jan Lorenz, and Peter Holtz. 2019. The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology* 58, 1 (2019), 129–149.
- [4] GroupLens. 2021. *MovieLens*. <https://grouplens.org/datasets/movielens/>
- [5] Kartik Hosanagar, Daniel Fleder, Dokyun Lee, and Andreas Buja. 2014. Will the Global Village Fracture Into Tribes? Recommender Systems and Their Effects on Consumer Fragmentation. *Management Science* 60, 4 (2014), 805–823. <http://www.jstor.org/stable/42919571>

- [6] 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.
- [7] Yehuda Koren. 2010. Collaborative Filtering with Temporal Dynamics. *Commun. ACM* 53 (04 2010), 89–97. <https://doi.org/10.1145/1721654.1721677>
- [8] N.K. Lathia. 2010. *Evaluating collaborative filtering over time*. Ph.D. Dissertation. UCL (University College London).
- [9] 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 (WWW '14). Association for Computing Machinery, New York, NY, USA, 677–686. <https://doi.org/10.1145/2566486.2568012>
- [10] Dimitar Nikolov, Diego F. M. Oliveira, Alessandro Flammini, and Filippo Menczer. 2015. Measuring online social bubbles. *PeerJ. Computer science* 1 (2015), e38.
- [11] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin UK, New York, NY.
- [12] Robert D. Putnam. 2000. *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster.
- [13] KE Stanovich, RF West, and ME Toplak. 2013. Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science* 22, 4 (2013), 259–264. <https://doi.org/10.1177/0963721413480174>
- [14] David Sumpter. 2018. *Outnumbered: from Facebook and Google to fake news and filter-bubbles - the algorithms that control our lives*. Bloomsbury Sigma, London.
- [15] CR Sunstein. 2002. The law of group polarization. *Journal of Political Philosophy* 10, 2 (2002), 175–195. <https://doi.org/10.1111/1467-9760.00148>
- [16] Cass R. Sunstein. 2007. *Republic.com 2.0*. Princeton university press, Princeton.
- [17] Zoltán Szilávik, Wojtek Kowalczyk, and Martijn Schut. 2011. Diversity measurement of recommender systems under different user choice models. In *Fifth International AAAI Conference on Weblogs and Social Media*.