Simulating the Impact of Recommender Systems on the Evolution of Collective Users' Choices

Conference Paper · July 2020

DOI: 10.1145/3372923.3404812

3 authors:

Naieme Hazrati
Free University of Bozen-Bolzano

Mehdi Elahi
University of Bergen

Francesco Ricci
Free University of Bozen-Bolzano

Simulating the Impact of Recommender Systems on the Evolution of Collective Users' Choices

Naieme Hazrati
Free University of Bozen-Bolzano
nhazrati@unibz.it

Mehdi Elahi
University of Bergen
mehdi.elahi@uib.no

Francesco Ricci
Free University of Bozen-Bolzano
fricci@unibz.it

ABSTRACT

The major focus of recommender systems (RSs) research is on improving the goodness of the generated recommendations. Less attention has been dedicated to understand the effect of an RS on the actual users' choices. Hence, in this paper, we propose a novel simulation model of users' choices under the influence of an RS. The model leverages real rating/choice data observed up to a point in time in order to simulate next, month-by-month, choices of the users. We have analysed choice diversity, popularity and utility and found that: RSs have different effects on the users' choices; the behaviour of new users is particularly important to understand collective choices; and the users' previous knowledge, i.e., their "awareness" of the item catalogue greatly affects choice diversity.

CCS CONCEPTS

• **Information systems → Information systems applications**;

KEYWORDS

recommender systems, choice behaviour, simulation

ACM Reference Format:

Naieme Hazrati, Mehdi Elahi, and Francesco Ricci. 2020. Simulating the Impact of Recommender Systems on the Evolution of Collective Users' Choices. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media (HT '20), July 13–15, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3372923.3404812

1 INTRODUCTION

Recommender Systems (RSs) can support users in their choice-making process and are typically evaluated by measuring the precision and quality of their recommendations [7, 18]. However, it is also important to assess how users are actually using RSs [12, 13, 21]. In that respect, a few previous works focused on understanding if and how RSs can influence the actual users' choices [5, 6, 9, 16, 20]. These studies have analysed aggregated measures of the impact of RSs on the distribution of the choices made by their users. Two approaches have been used: *online experiments* and *simulations*.

In online experiments, a web application allows online users to choose items to consume while browsing recommendations [15, 16]. While these studies have shown how users react to recommendations under certain conditions, they could not explore the large number of possible alternative settings (e.g., alternative RSs, data sets). Hence, *simulations* procedures of users' choices have been proposed [5, 9, 20]. In these simulations it is assumed that users adopt an algorithmic choice model. For example, recommended items are more likely to be chosen than not recommended ones. Repeated simulated choices are produced and the collective choice behaviour of a population of users is measured.

Even though previous simulation studies have obtained interesting results, some significant limitations need to be addressed. First of all, synthetic data (items and users' profiles) has mostly been used. This has produced results of dubious validity and hard to observe in more real operational settings. In order to cope with that issue, we use two well-known data sets of users' ratings, for initialising the users' preference model (utility), which is necessary to simulate realistic choices. Moreover, in previous studies the impact of the users' knowledge of the item catalogue, what is here called their *awareness* set, was not fully analysed. However, it is evident that many items are chosen independently from the recommendations, because users do have knowledge of the existence of these items from other sources. Hence, also the impact on choice behaviour of the size of the awareness set is worth to be analysed.

Hence, with the aim of better assessing RSs' effect on users' choices, we have designed a novel choice making simulation procedure where alternative RSs are employed and users are simulated to make repeated choices in a sequence of time intervals (months). During each interval, users choose some items according to a multinomial-logit choice model which is based on the estimated utility they obtain from the items. Furthermore, the simulation makes the assumption that users are not aware of the entire items' catalogue; recommendations change this knowledge and are more likely to be chosen than other items with similar utility. The simulation is initialised by using information about real user choices, taken from two data sets: *MovieLens* 100*K* and Amazon *Kindle*. We consider five different RSs, personalized and non-personalized. We analyse the effect of RSs on the Gini index and the popularity of the chosen items, which are measures of the choice diversity, and on the utility of the chosen items, which gives an indication of the quality of the users' choices. Finally, in order to determine the role of the users knowledge of the items' catalogue on their choices we have conducted simulations with alternative awareness set sizes.

While some of our results confirm hypothesis already proved in previous studies, s.a., that the choice diversity produced by non-personalised RSs is smaller than that produced by personalised RSs, we have also obtained new and unexpected results that reveal subtle connections between RSs, users, previous choices data, and the properties of the users' choices (distribution and quality).

HT '20, July 13–15, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7098-1/20/07…$15.00
https://doi.org/10.1145/3372923.3404812

Firstly, among the non-personalised RSs, it is not the RS suggesting the most popular items (*POP*) but that suggesting items with the largest average rating (*AR*) that leads users to make less diverse choices.

Secondly, Among the personalised RSs, *LPCF* (Low Popularity-based Collaborative Filtering RS), which is a collaborative RS that tends to avoid the recommendation of popular items, can produce high diversity and even optimize the utility of the choices made under its influence.

Thirdly, the awareness set size modulates the recommendations' acceptance: when users have smaller awareness sets, the RS will affect more the users' choices compared to when the awareness set is larger.

In conclusion, the obtained results are important as they show that, depending on the goals of system designer, a specific RS can be applied in a target context and its effect predicted. Moreover, the awareness set and the RS can be simultaneously manipulated based on system's goal.

2 RELATED WORK

Recommender systems have drawn a lot of attention in the research community for their ability to support online users of almost any web portal in their decision making process. However, the effect, which RSs can have on the choice behaviour of a single user and globally on a community of users, has not received adequate attention. In particular, only a few simulations studies, similar to the one that is here proposed, that focused on the effect of RSs on users' choice behaviour have been conducted [2, 5, 9, 20].

Fleder et. al. [5] introduced a simulation in which the users iteratively select items among a small set of candidate fictitious products based on a probabilistic multinomial-logit choice model [3]. Their model is based on a *randomly* generated utility function of the users. The higher is the utility, the more likely the item is going to be chosen. They also assumed that the users can only select items in a proper subset of the catalogue, which is called "awareness set". Moreover, if an item is recommended to a user, then the probability that the user select the recommended item is increased. Finally, they observed RSs' effect on the users' choices in terms of diversity. This study led to interesting results and inspired our analysis. However, their simulation has been conducted under rather simplifying modelling assumptions (e.g., they used a random utility function), as well as, they considered a small number of users and items, generated synthetically from a random distribution. Consequently, their findings are limited in providing a proper picture of choice behaviour in realistic settings.

Also Szlávik et. al. [20] designed a simulation procedure and observed the impact of RSs on the users' choice diversity. They discovered that different choice models lead to different dynamics in the choice diversity. They created alternative choice models, considering different acceptance probability of the recommendations, i.e., forcing the simulated users to select the recommended item or giving them the chance to select among other items. They discovered that when the users are forced to make the same number of choices, the diversity is not necessarily increased, and the given ratings of the users also decreases.

Finally, in a short paper [9], we have reported some initial and preliminary results of our analysis of the effect of a few RSs on users' choice behaviour. We have analysed their impact only on the Gini index, as an indicator of choice diversity, by using only one reference data set, and when the recommender suggests a single item. We found that Gini index does vary under the influence of alternative RSs. However, no analysis of the impact of the awareness set and the presence of new users was done.

3 SIMULATION OF USERS' CHOICES

We simulate the iterative process of choice making of users for items in a system. Users select items in monthly time intervals. We use the observed choices in a data set, up to a certain time point t_0, as the starting point of the simulation. We use this initial data set to train a RS, and then we simulate the choices in successive months. At the end of each month, the RS is re-trained by also considering the simulated choices of that month as real signal of their true preferences. Additionally, in order to correctly simulate users' choices we predict the user's preferences over the items (ratings) by using a Factorization Machine [17] trained on the full set of ratings available in the data set.

The basic schema of the proposed simulation is shown in Figure 1. We assume that in each month interval, the users select items one after another. When a user is simulated to make a choice, first her "awareness" set is built. The awareness set limits the options a user can choose. Then, a RS suggests a set of items that are added to the user's awareness set. Finally, the user makes a choice based on a choice model, where the "utility" of a choice is assumed to be its predicted rating and influences the probability to make a choice. New users enter the simulation on successive months, when they actually started to make choices according to the reference data set. However, in these cases, personalized RSs cannot generate recommendations for these new users, because of the lack of information of their previous choices, which are needed to create a recommendation model for them.

3.1 Awareness Set

We assume that a generic user u is not aware of the entire items' catalogue and can only choose items from her *Awareness Sets* A_u, which are all assumed to be of size A. Every time u is simulated to make a choice, her awareness set is built by including the top A items in the list D_u. D_u is a ranked list which aggregates two ranked lists: B_u and H_u:

- B_u contains the items, which have not been chosen by u before, sorted with respect to their popularity. Popularity of an item is equal to the number of times the item was chosen by the users in the previous time intervals.
- H_u contains the items, which have not been chosen by u before, sorted with respect to their utility for u (predicted rating).

The two ranked lists are then aggregated by using the Borda count method [19], and the top A items are included in the awareness set. Hence, in practice, we assume that users are aware of popular and high utility items. Then, the RS is supposed to modify the awareness set of a user. In fact, if an item j is recommended to a user u, then this item enters in u's awareness set (if it is not already in). Moreover, when the user u selects an item j from her awareness set, j is removed from her awareness set; this because we assume that a user chooses an item only once (e.g., a book is bought only once).

In the performed simulations we varied the awareness set sizes in order to investigate its effect on the users' choices distribution. In the first used data set, $Kindle$, we consider larger values, varying from 500 to 50000, while in the second, $MovieLens$, we consider, for lack of space, only one value, 200 (data set are fully described in Section 4.2).

3.2 Choice Model

When a user is simulated to make a choice (for an item) is is supposed to use a multinomial-logit choice model. We adopt this model because it is a simple but effective approach, which has been previously validated. This also makes our results comparable with earlier simulations [5]. The utility of the item j for the user u is assumed to be known by the user and equal to the best estimation of the rating of the user u for the item j (using the full knowledge of the reference data set): $v_{uj} = \hat{r}_{uj}$, where $\hat{r}_{uj}$ is the predicted rating of the item j for the user u. u is supposed to choose an item j among her awareness set's items, with the following probability:

$$p(u\ chooses\ j) = \frac{e^{v_{uj}}}{\sum_{k \in A_u} e^{v_{uk}}} \tag{1}$$

In practice, items that have larger predicted rating are more likely to be chosen, but the user does not necessarily select the items with the largest predicted ratings. This assumption tries to take into account the potential human errors introduced by utility estimation and the fact that no decision maker is perfectly rational.

3.3 Recommendations

The following five RSs are considered in the simulation and they recommend 10 items each time a user is simulated to make a choice.

- *PCF* - Popularity-based CF: is a neighborhood-based collaborative filtering that identifies the nearest neighbors of a target user u (by using the cosine similarity between the users' 0/1 choices' vectors). The most popular items among the choices of the nearest neighbor users are recommended to the target user.
- *LPCF* - Low Popularity-based CF: is similar to *PCF*, but it penalizes the score of popular items, computed by *PCF*, by multiplying it with the inverse of their popularity. The highest scored items are recommended.
- *FM* - Factor Model: is a RS which generates recommendations following the approach proposed in [11].
- *POP* - Popularity-based: the most popular items in terms of the number of times that they were selected before are recommended.
- *AR* - Average Rating: The items are scored with a variation of the average rating. This methods is used by IMDB.com. A weighted average is calculated for each item as: $WR = (\frac{v}{v+m} \times R) + (\frac{m}{v+m} \times C)$, Where R is the average rating for the item, v is the number of times that this item is rated, m is the minimum number of ratings required to be considered by the RS, and C is the average of all of the ratings in the data set. The highest scored items are recommended.

It is here important to note that if an item j is recommended to the user u, by a RS, then it is added to the awareness set but also

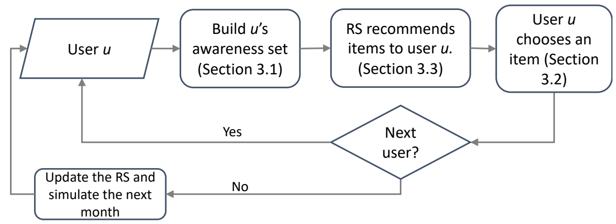


Figure 1: Simulation procedure of one month's choices

its utility v_{uj} is boosted by a multiplicative factor δ:

$$v_{uj} = \delta * v_{uj}$$

Hence, recommended items are more likely to be chosen by the user, compared with items having the same (estimated) utility, but not recommended. This simulates the effect of recommendations on user choice behaviour. Moreover, in order to simulate that recommended items are however not always chosen, then δ was set to 2. We experimentally checked that with this value users choose one of the 10 recommended items with 60-70% probability.

4 EXPERIMENTAL STRATEGY

4.1 Evaluation Metrics

By running the above described simulation procedure we are interested to measure the effect of RSs on the distribution of the users choices. Particularly, we are interested in their diversity and quality. Hence, we introduce here three metrics that capture these properties. The simulation results shown later are obtained by averaging the measured metrics over 5 repetitions of the simulation.

Moreover, for each month in the simulation, we show metric results computed over the "accumulated" choices up to the simulated month, i.e., from the first simulated month to a target simulated month. We will then analyse the variation of the metric at successive time intervals.

In order to measure choice diversity we use the **Gini index** that has also been considered in related studies [1, 5, 14, 16, 20]. Gini index definition is based on the "Lorenz curve" $L(x)$, which is the fraction of the choices generated by the lowest 100*x% chosen items, $x \in [0, 1]$ [4]. The Gini index is then: $G = \frac{A}{A+B}$, where $A = \int_0^1 (x - L(x))dx$ and $B = \frac{1}{2} - A$. The Gini index measures inequality distribution with a single value $G \in [0, 1]$. G is 0 when a perfectly uniform distribution of choices across items is observed, while it is close to 1 when choices are including only a small part of the items' catalogue.

We also measure the **average utility** of the users' choices. It is the mean of the (predicted) rating of the users chosen items. For each user, we compute the average rating of the chosen items, then we average the users' mean values. We are interested in this metric to understand if a RS helps the users to find valuable items. The predicted rating of a user for an item is in fact the only measure that we have at our disposal to assess the quality of their choices.

Finally, we measure the **popularity** of the chosen items. In every simulation month, the average popularity of an item is equal to the number of times that this item has been chosen by the users up to that time point, divided by the overall number of choices in the previous months.

4.2 Datasets

We searched for time-stamped choice data sets for our study and we selected two of them that have different characteristics and exemplify others: Amazon $Kindle$[1] [10] and *Movielens* 100K [8].

MovieLens contains 100,000 ratings given by 943 users to 1682 movies in less than 8 months. The first four months' choices are considered as the starting point of the simulation. Then we simulate the successive four months' choices. *MovieLens* 100K is used here since it is well known and has been analysed in several previous papers. However, in this data set, the full set of rated items could not be considered as equivalent to the user choices. In fact, a user can rate a movie in the catalogue without having chosen it before. Therefore, we decided to consider as users' choices the movies with ratings equal or larger than 4, which is a common approach to separate relevant from not relevant recommendations.

However, having a data set with users' actual choices should help to better simulate their choices. Hence, we decided to use the $Kindle$ data set which actually contains ratings mostly for real book purchases. It contains 3,205,467 ratings given by 1,406,890 users to 430,530 items over roughly 16 years. This data set is very sparse, especially in the first 10 years of data. So we decided to consider a higher number of months' choices as starting point of the simulation. We have therefore initially trained the RSs with the first 144 months data (12 years) and then simulated the users' choices (and iteratively retrained the RSs) in the next 10 months. From the starting date to month 154 in the $Kindle$ data set, 29,059 ratings are present, expressed by 18,764 users for 12,335 items.

In the $Kindle$ dataset there are more users and items, compared to *Movielens* 100K, and the average number of choices per user is much smaller: 1.5 v.s. 57. Hence, $Kindle$ data is much sparser than *Movielens*. This data set is therefore useful to understand the effect of high sparsity on the Gini index, which is not clear yet.

5 EXPERIMENTAL RESULTS

Choice diversity. We analyse first the effect of RSs on the diversity of the simulated choices by considering the evolution of the Gini index, at the end of each simulation month. In Figure 2 the x axis is the simulation month number and the y axis shows the Gini index calculated over the choices made from the beginning of the simulation until the end of month x. The label "Observed" refers to the Gini measured on the actual choices of the users observed in the data set. Here the awareness set size is 200 for MovieLens and 2000 for Kindle.

PCF, *LPCF* and *FM* (personalised RSs) produce lower Gini than *AR* and *POP* (non-personalised RSs). We discovered that Gini, for the personalised RSs is also influenced by the presence of many new users. For these users personalized RSs do not generate recommendations. Hence, in the absence of recommendations, only the item utility determines the choices (see Eq. 1), which will be more uniformly distributed among the candidate choices, as no item has an increased probability to be chosen. Conversely, when the non-personalised RSs are used, all the users, including the new ones, will receive the same recommendations. Hence, especially the new users, in comparison with personalised RSs, are more likely to choose from a narrower range of items. Hence, Gini, for non-personalised RSs is larger. Among personalised RSs, we can observe that *LPCF* produces the lowest Gini (highest diversity). In fact, we also measured that *LPCF* recommends the highest number of distinct items compared to the other RSs.

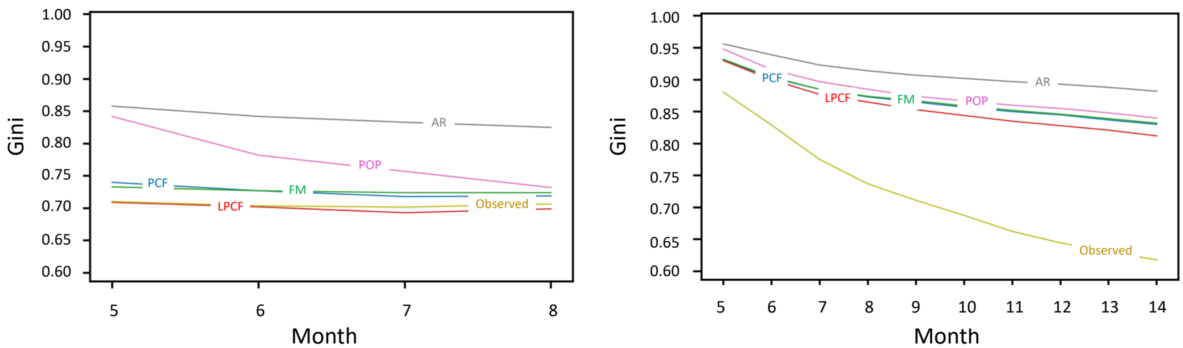
Among the non-personalized RSs, quite surprisingly, *AR* produces a much lower choice diversity, compared to *POP*. In fact, *AR* recommends the items with the largest average rating/utility. So the items recommended by *AR* have typically already a higher utility, compared to those recommended by *POP* (as we have measured). Hence, the difference in utility between the recommended and not recommended items is larger for *AR*, and, since the choice probability grows with item utility (see Eq. 1), there is a higher probability that users will select the items recommended by *AR* compared to those recommended by *POP*. Consequently, the choices influenced by *AR* will be less diverse.

Popularity and utility of the choices. We have also measured the popularity of the chosen items, not shown here for lack of space. In both data sets, non-personalised RSs produced choices for items with higher popularity, compared to the personalized ones. In fact, *POP* and *AR* push users to select a narrower range of items, the recommended items are chosen many times, and their popularity increases. However, quite surprisingly, *POP* causes a lower popularity of the chosen items compared to *AR*. This is due to the fact that the recommended items in *POP* have a relatively lower utility and thus the simulated users are less likely to choose them. Among the personalized RSs, instead, the choices influenced by *LPCF* are for the least popular items (as expected by its definition).

Moreover, we also calculated the average utility/rating of the users' choices. We discovered that in both data sets, personalized RSs lead to choices with higher average utility. Additionally, when *LPCF* is used, on average, users choose items that have higher (predicted) utility. We recall that the utility of an item for a user is estimated by using a factorization machine algorithm that considers all the available rating data (hence it is independent from the *LPCF* predicted score of the items). Hence, the choices made under the influence of *LPCF* seems to be more "accurate". This is unexpected, as we thought that *FM* should excel in that respect. This can be explained by the fact that by decreasing the score of popular items, *LPCF* can find and suggest items which are more specific to the users, hence with larger utility. It is instead easier to explain that under the influence of *POP* the utility of the chosen items is the lowest, since the RS maximises the popularity of the recommendations, which may not always yield the best items for each user.

Impact on choice diversity of the awareness set. Figure 3 shows the evolution of the Gini index when the choices are influenced by the *PCF* RS, for different awareness set sizes ($Kindle$ data set). Clearly, by increasing the awareness set size a higher diversity is observed, getting closer to the Gini index of the observed choices. We decided to further study this effect by considering the joint effect of the RS and the awareness set size A. In the $Kindle$ data set, we considered A =1000, 2000 and 5000.

[1] http://jmcauley.ucsd.edu/data/amazon/



(a) MovieLens 100K dataset

(b) Kindle dataset

Figure 2: Evolution of the Gini index of the observed and simulated choices under five different RSs.

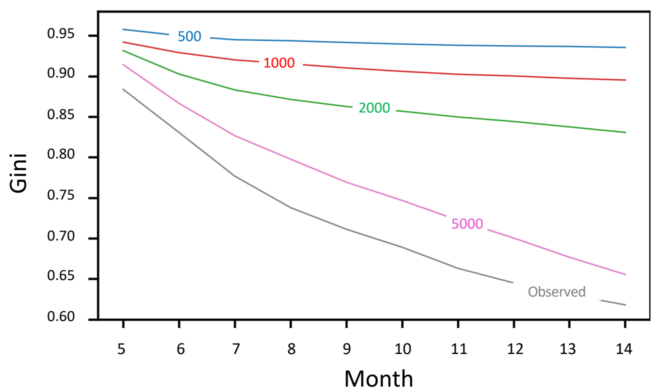


Figure 3: The evolution of Gini for different awareness set sizes 500, 1000, 2000 and 5000, (PCF RS and Kindle dataset)

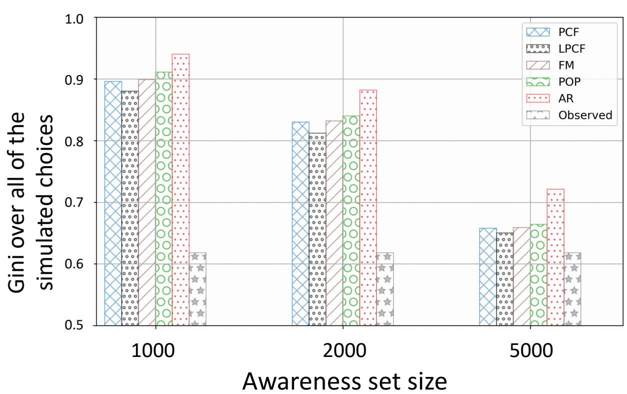


Figure 4: Gini index at the end of the simulation, month 14 - Kindle data set

Figure 4 shows that the awareness set size has an important effect on Gini, even larger than the effect produced by the RSs. In the figure it is shown this metric for the last month of the simulation process (all the accumulated choices are counted). Firstly, by increasing the awareness set size, the Gini index decreases the for all the RSs. This is because higher awareness set sizes allow the users to select items among a larger collection. Hence, the choices become more diverse. RSs' relative impact on the diversity of the choices does not seem to change with the awareness size. Additionally, larger awareness sizes, produces choices less focused on the recommended items, and this explains why the Gini index of the RSs are more similar to each other, when the larger awareness set is used.

6 DISCUSSION AND FUTURE WORK

In this paper, we have proposed a simulation procedure modelling month-by-month users' choice making under the effect of alternative RSs. The procedure was designed to be as close as possible to a real evolution of users' choices. We discovered that both the RS algorithm and the user awareness of the catalogue affect choices' distribution, popularity and utility. We discovered that the user awareness size is even more important than the RS to capture the real dynamics of the user choices. We also discovered, some novel properties of RSs, such that a non-personalised RSs suggesting items with average high rating can produce the lowest diversity of the choices, and for popular items, even more than the simple recommendation of popular items. We also discovered that a personalised RS that penalises popular items (*LPCF*) can improve choice diversity without penalising their utility.

Hence, we have shown how informative can be our simulation approach. We believe that it can help other researchers to investigate RSs' effect on users' behaviour. It can be used in operational systems to investigate the impact of changing different system parameters, such as the recommendation approach or the number of recommended items. In the future, one can easily experiment variations of the proposed simulation. For instance, it is possible to simulate users repeatedly choosing the same item, as it is common in the music domain. Or, one could simulate the effect of having users with significant differences in their awareness set sizes.

Moreover, we still need to better understand and correctly model other factors that influence users' choices, s.a., the relationships between users, and the users' knowledge of the items. Furthermore, we need to be able to predict the number of choices made by each user, in order to build a more flexible and predictive model.

REFERENCES

[1] Panagiotis Adamopoulos, Alexander Tuzhilin, and Peter Mountanos. 2015. Measuring the concentration reinforcement bias of recommender systems. *rN (i)* 1 (2015), 2.

[2] Dimitrios Bountouridis, Jaron Harambam, Mykola Makhortykh, Mónica Marrero, Nava Tintarev, and Claudia Hauff. 2019. SIREN: A Simulation Framework for Understanding the Effects of Recommender Systems in Online News Environments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 150–159.

[3] William A Brock and Steven N Durlauf. 2002. A multinomial-choice model of neighborhood effects. *American Economic Review* 92, 2 (2002), 298–303.

[4] Robert Dorfman. 1979. A formula for the Gini coefficient. *The review of economics and statistics* (1979), 146–149.

[5] Daniel Fleder and Kartik Hosanagar. 2009. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management science* 55, 5 (2009), 697–712.

[6] Daniel M Fleder and Kartik Hosanagar. 2007. Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce*. ACM, 192–199.

[7] Asela Gunawardana and Guy Shani. 2015. Evaluating Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer, 265–308. https://doi.org/10.1007/978-1-4899-7637-6_8

[8] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015), 1–19.

[9] Naieme Hazrati, Mehdi Elahi, and Francesco Ricci. 2019. Analysing Recommender Systems Impact on Users' Choices. *In 1st International Workshop on the Impact of Recommender Systems at RecSys* (2019).

[10] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*. 507–517.

[11] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*. Ieee, 263–272.

[12] Farid Huseynov, Sema Yildiz Huseynov, and Sevgi Özkan. 2016. The influence of knowledge-based e-commerce product recommender agents on online consumer decision-making. *Information Development* 32, 1 (2016), 81–90.

[13] Dietmar Jannach and Michael Jugovac. 2019. Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems (TMIS)* 10, 4 (2019), 1–23.

[14] Dokyun Lee and Kartik Hosanagar. 2019. How Do Recommender Systems Affect Sales Diversity? A Cross-Category Investigation via Randomized Field Experiment. *Information Systems Research* 30, 1 (2019), 239–259.

[15] Christian Matt, Alexander Benlian, Thomas Hess, and Christian Weiß. 2014. Escaping from the filter bubble? The effects of novelty and serendipity on users' evaluations of online recommendations. (2014).

[16] Christian Matt, Thomas Hess, and Christian Weiß. 2013. The differences between recommender technologies in their impact on sales diversity. *ICIS* (2013).

[17] Steffen Rendle. 2012. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (TIST)* 3, 3 (2012), 57.

[18] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. In *Recommender systems handbook*. Springer, 1–34.

[19] Donald G Saari. 1985. The Optimal Ranking Method in the Borda Count. *pure.iiasa.ac.at* (1985).

[20] Zoltán Szlávik, 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*.

[21] Zhang Ying, Chen Caixia, Gu Wen, and Liu Xiaogang. 2018. Impact of recommender systems on unplanned purchase behaviours in e-commerce. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*. IEEE, 21–30.