Inf Syst Front (2018) 20:1173–1189
DOI 10.1007/s10796-017-9782-y

User Personality and User Satisfaction with Recommender Systems

Tien T. Nguyen[1] · F. Maxwell Harper[1] · Loren Terveen[1] · Joseph A. Konstan[1]

Published online: 3 September 2017

Abstract In this study, we show that individual users' preferences for the level of diversity, popularity, and serendipity in recommendation lists cannot be inferred from their ratings alone. We demonstrate that we can extract strong signals about individual preferences for recommendation diversity, popularity and serendipity by measuring their personality traits. We conducted an online experiment with over 1,800 users for six months on a live recommendation system. In this experiment, we asked users to evaluate a list of movie recommendations with different levels of diversity, popularity, and serendipity. Then, we assessed users' personality traits using the Ten-item Personality Inventory (TIPI). We found that ratings-based recommender systems may often fail to deliver preferred levels of diversity, popularity, and serendipity for their users (e.g. users with high-serendipity preferences). We also found that users with different personalities have different preferences for these three recommendation properties. Our work suggests that we can improve user satisfaction when we integrate users' personality traits into the process of generating recommendations.

Keywords Human factors · Personality · Recommender systems · Big-five personality traits · User preferences · Recommendation diversity · Recommendation popularity · Recommendation serendipity

✉ Tien T. Nguyen
nguy1749@umn.edu

[1] 5-240, GroupLens Research, Keller Hall, Department of Computer Science and Engineering, University of Minnesota, 200 Union St SE, Minneapolis, MN 55455, USA

Terveen's current areas of research emphasis are: peer production systems, incentive mechanisms to enhance user participation, the quantitative analysis of social media data, and geographically based online communities. He leads projects that have: revealed new information about how valuable content is created on Wikipedia, produced and deployed new interface designs to enhance participation in online communities, developed a new location-based messaging system, combined wiki and geographical information systems technologies to create novel interfaces that let people enter and access information about places in their local communities, and created the first fully functional geographical wiki. In all his work, he seeks to use knowledge gained from empirical studies to build novel systems that solve real problems.

1 Introduction

One approach to increase user satisfaction with recommendations is to generate recommendation lists with satisfactory levels of popularity, diversity, and serendipity (for example (Ziegler et al. 2005; Zhang et al. 2012; Oh et al. 2011)). However, previous research assumes that all users have the same preferences for the levels of popularity, diversity, and serendipity in their recommendation lists, and that recommendation algorithms can learn these preferences from users' ratings. We hypothesize that these assumptions are not true because users do not provide ratings for all of the movies they have watched. Therefore, we ask:

RQ1: How satisfied are users with the levels of diversity, popularity and serendipity of recommendation lists produced by rating-based recommendation algorithms?

Furthermore, we investigate factors that can be integrated into recommendation algorithms to generate recommendations with the appropriate levels of diversity, popularity, and serendipity for individual users. User personality is a stable information source about individual user (Costa and McCrae 2008; John and Srivastava 1999), and has significant connections with users' tastes and interests (Kemp 1996; Rentfrow and Gosling 2003). Therefore, we examine whether information about users' personalities can help recommendation systems deliver the appropriate levels of diversity, popularity and serendipity in recommendation lists, and increase users' satisfaction. We ask:

RQ2: Is there a correlation between users' personality traits and their preferences for diversity, popularity, and serendipity?
RQ3: Is there a correlation between users' personality traits and their satisfaction with the list of recommendations?

We conducted a study on a live movie recommendation system with more than 1800 users. We showed each user a list of 12 personalized movie recommendations. We varied the levels of diversity, popularity, and serendipity of these lists. We then asked the users to rate how satisfied they were with the levels of diversity, popularity, and serendipity of these recommendation lists, and how much they would enjoy watching the movies in the lists. Finally, we assessed users' personality using the Ten-item Personality Inventory proposed by (Gosling et al. 2003). We used these questions from Gosling et al. because these questions have been widely used to access the personality of a user in research community.

Our contributions are three-fold. First, we demonstrate that user personality, in addition to user ratings, can provide good signals for recommender systems to deliver satisfactory levels of recommendation diversity, popularity and serendipity to individual users. Our findings suggest that integrating user personality into recommendation algorithms could lead to increased user satisfaction. Second, we found that user preferences for recommendation diversity, popularity and serendipity do not correlate with user satisfaction as assumed in prior work. Last but not least, we found that recommendation algorithms based on ratings alone can not always generate recommendation lists with the desired levels of diversity, popularity, or serendipity.

In the next section, we will discuss related work in personality psychology and recommender systems, and identify the opportunities for this research project.

2 Related work

2.1 User personality and user preference

Prior work in user personality has identified five personality traits - openness to new experiences, conscientiousness, extraversion (or introversion), neuroticism and agreeableness (John and Srivastava 1999; McCrae and Costa 1987; Gosling et al. 2003). Prior work also showed that there are significant connections between these traits and people's tastes and interests, and that personality can be reliable sources to describe users' habits and behaviors. Kraaykamp and Van Eijck (2005) found that personality has effects on users' media preferences. Rentfrow and Gosling (2003) showed that personality traits "have roles to play in the formation and maintenance of music preferences". More specifically, they showed that people with openness to new experiences usually tend to have preferences for jazz, blues, and classical music. Other studies showed that people who were open to new experiences preferred familiarity to novelty (McCrae and Costa 1987), or that emotionally unstable people preferred popular media programs (Kraaykamp and Van Eijck 2005). Chausson (2010) also showed that people who are open to new experiences are likely to prefer comedy and fantasy movies while conscientious individuals are more inclined to enjoy action movies, and neurotic people tend to like romantic movies.

To the best of our knowledge, there has not been much research in the recommender system domain about the relationships between user personality and user satisfaction with the levels of recommendation diversity, popularity, and serendipity. Hu and Pu (2011) pointed out that there are few datasets with measures of user personality and their consumption (ratings). Among prior work on personality in recommender systems, Chen et al.'s work Chen et al. 2013) is closest to ours. In their work, they reported that user personality influenced user needs for movie recommendation diversity. However, they focused on the diversities of genres, or actors, or countries of origin whereas our interest is the diversity of movie content, as reported in Nguyen et al. (2014). Furthermore, Chen et al. did not examine user satisfaction, user personality and how user personality can improve user satisfaction. Therefore, in this study, we measured user personality and user evaluations of recommendation lists. We hypothesize that personality has significant connections with user preferences for the levels of diversity, popularity, and serendipity of recommendations, and how satisfied they are with those recommendations.

2.2 User preferences for recommendation diversity, popularity, and serendipity

To improve user satisfaction, prior work in recommender systems looked at three different properties of recommendation lists: diversity, popularity, and serendipity. Diversity indicates how different the recommendations are. It is measured as the average of all pairwise distances (representing the differences) of any two recommendations (Ziegler et al. 2005). Serendipity indicates how different the recommendations are from what users usually consume. One way to measure recommendation serendipity is to first identify pairs of one item in the recommendation list and one item in the list a user has consumed. Then we take the average of all pairwise distances (representing the differences) of these pairs of items (Zhang et al. 2012). Recommendation popularity indicates how common the recommendations are. One way to approximate recommendation popularity is to look at how frequent users consumed the recommended items (Celma and Cano 2008).

We note that there are distinct differences between popularity and serendipity. A less popular item is not necessarily serendipitous for a user that is not interested in the item. On the other hand, a highly popular item is not necessarily less serendipitous for a user, because he or she might not have heard about it yet due to some barriers (e.g. a cultural barrier).

Prior work on diversifying recommendations argued that users are more satisfied when they get diverse recommendations. Ali and van Wijnand (2004) coined the term "portfolio effect" to describe the problem of recommender systems generating less diverse recommendation lists with only a few genres. Each recommendation in the set is good, but the set collectively is bad since it is focused on only a few genres. The incremental utility of recommendations for users decreases when users consume these similar items over and over. This phenomenon is termed "the law of diminishing marginal returns" and is studied extensively in economics (Shephard and Fare 1974). Indeed, Ziegler et al. 2005 found that users liked diverse recommendation lists more than less diverse ones, even though less diverse lists collectively had higher predicted ratings.

Similarly, prior work also suggested that recommending popular items is less useful to users since they probably already know about these items. Herlocker et al. (2004) argued that popular recommendations were only useful when users were interested in these recommendations. In fact, recommender systems are prone to popularity bias, presenting users with popular recommendations than niche ones (Fleder and Hosanagar 2007; Adomavicius and Kwon 2009; Celma and Cano 2008). Users appreciate recommendation lists with some less popular items (Steck 2011).

Likewise, prior work also suggested that recommendation lists should be novel and interesting to users in order to be useful to them. An example of useless recommendations is recommending bananas to grocery shoppers, since they already know about bananas and may intend to buy them regardless (Herlocker et al. 2004). Users may even be annoyed if they know about bananas but do not intend to buy them. Indeed, Zhang et al. (2012) showed that increasing serendipity in recommendation lists improved user satisfaction.

2.3 The assumptions of prior work in recommender system and our research opportunities

Although researchers demonstrated in prior work that user satisfaction correlates highly with recommendation diversity, popularity and serendipity, they made two assumptions that we examine in this study.

First, prior work assumed that users' ratings contain sufficient information about their preferences for diversity, popularity, or serendipity. Oh et al. (2011) showed that recommender systems could learn about user tendency of consuming popular items via user ratings. They proposed a method called Personal Popularity Tendency Matching (PPTM) to measure these tendencies and used PPTM to generate useful recommendation lists. Zhang and Hurley (2009) proposed a user-profile-partitioning technique to capture users' ranges of tastes for novelty from their ratings. They showed that this technique could improve recommendation novelty at a small cost to overall accuracy. Vargas and Castells (2013) also applied the same technique to improve recommendation diversity. We believe that users have their individual preferences for diversity, popularity, and serendipity and that we cannot depend on user ratings or user consumption data to infer these individual preferences.

Second, they assumed that users have global preferences for diversity, popularity, and serendipity. They assumed that all users prefer diverse recommendations over unvaried ones, slightly less popular recommendations over highly popular ones, and more serendipitous recommendations over familiar ones. Thus, they focused on proposing methods to tune recommendations towards those global preferences. This assumption leads to contradicting results. For example Zhang et al. (2012) showed that showing users recommendations that the users were not familiar with improved user satisfaction, but Ekstrand et al. (2014) showed that novelty reduced user satisfaction.

Thus, we test these two assumptions in this study. First, we investigate whether user preference for recommendation diversity, popularity and serendipity can be extracted from user ratings alone. Second, we examine if users with different personalities have different preferences for recommendation diversity, popularity, and serendipity.

3 User study

Our study examines the relationship between users' personality types and their satisfaction with more or less diverse, popular and serendipitous movie recommendation lists. We ran a user study on MovieLens,[1] a well-known recommender system in the research community. We showed each user a list of movie recommendations. This list of recommendations has a low, medium or high level of diversity or popularity or serendipity. We asked users to rate how satisfied they were with the level of diversity, popularity, and serendipity of the recommendation list, and how much they would enjoy watching the movies in the list. We also asked questions to assess users' personalities.

3.1 User experiment

We ran the user experiment on MovieLens from May 12th, 2015 to October 14th, 2015.

3.1.1 Eligible users

To participate in our study, a user must a) have an account with MovieLens, and b) rate at least 15 movies prior to his or her participation. We asked for the minimum of 15 movies because MovieLens traditionally required users to rate at least 15 movies before it would give users personalized recommendations (see (Nguyen et al. 2014; Chang et al. 2015)). We showed eligible users an invitation to participate in our study when they logged in to MovieLens. In total, 1888 users participated in our study.

3.1.2 A between-subjects experiment

Our user experiment is between-subjects, with 10 conditions - 3 metrics (diversity, popularity, serendipity) × 3 different levels (high, medium, and low) and a control condition. Once a user accepted our invitation to participate in the study, we randomly assigned this user to one of the 10 conditions. Each user saw only a list of 12 recommendations selected from the top 60 recommendations personalized for him or her. Thus, all users saw different recommendation lists. We extracted the recommendation lists from the top 60 recommendations because our offline evaluation showed that the sixtieth recommendation was still a good recommendation with predicted ratings above 4 out of 5 starts for most users. We chose a between-subject design instead of a within-subject design because our offline evaluation showed that with a within-subject design, each user would receive several overlapping recommendation lists. The high level of overlap among the recommendation lists that each user would receive would negatively affect our manipulation of diversity, popularity and serendipity in three ways. First, a repeated recommendation is no longer serendipitous to users after they have seen it several times. Second, users would perceive repeated recommendations as more popular. Third, users would perceive the recommendation lists as less diverse if they contained overlapping items.

3.1.3 How we varied the levels of diversity, popularity, and serendipity of recommendation lists

Choosing a random set of 12 recommendations from the top 60 takes a non-trivial amount of time because there are C_{12}^{60} possible ways to do so. Furthermore, it is time-consuming to choose twelve recommendations such that these recommendations are the most diverse or the least diverse. However, we want our users to see the recommendations within 30 s after they accepted the study invitation. Thus, we implemented a greedy search algorithm to select twelve recommendations out of the top sixty based on our diversity metric. The algorithms for popularity and serendipity metrics are straightforward. Below are the descriptions of these algorithms.

Diversity We can estimate the diversity of a set by taking the average of all pairwise distances of all items in the set (see Ziegler et al.'s work (Ziegler et al. 2005)). We are interested in content-based distances, which we measure using the tag-genome information space, similar to (Nguyen et al. 2014). More information about the tag-genome feature space can be found in the work of Vig et al. (2012) or from GroupLens' dataset.[2] Since we want the diversity measure to be bounded by 0 and 1, we use cosine distance (i.e. 1 - cosine similarity) instead of Euclidean distance. Thus, given a set S of 12 movies, the diversity of S is defined as:

$$diversity\ score = \frac{1}{C_2^{12}} \sum_{m_i \in S} \sum_{\substack{m_j \in S \\ i \neq j}} cosine\ distance\left(m_i, m_j\right) \quad (1)$$

There are three different conditions for levels of diversity - high, medium and low. For the high-level condition, the algorithm selects twelve recommendations to form a list with a diversity score as high as possible (based on the above equation), and vice versa for the low-level condition. For the medium-level condition, the algorithm selects twelve recommendations to form a list with a diversity score in the middle.

We designed the diversity-based algorithm with two parts. In the first part, the algorithm initializes a seeding list with a pair of movies. In the second part, the algorithm iteratively adds a movie into the seeding list until the number of movies in the list is 12. To create a seeding list, the algorithm selects a pair of movies out of C_2^{60} possible pairs. This pair must satisfy the following criteria. Among all the C_2^{60} possible pairs, if the condition is high-level, the pair has the highest diversity score; if the condition is low-level, the pair has the lowest diversity score; if the condition is medium-level, the pair has the median

[1] movielens.org

[2] http://grouplens.org/datasets/movielens/tag-genome/

diversity score. The other fifty-eight movies form a remainder list. From the remainder list, the algorithm then subsequently selects a movie to add to the seeding list, so that, after adding, the new seeding list should have a) the highest possible diversity score if the condition is high-level, b) the lowest possible diversity score if the condition is low-level and c) the median diversity score if the condition is medium-level.

Popularity Popularity indicates how commonly known all the items in a set of recommendations are. We can estimate the popularity of an item by computing how many users have consumed (or rated) the item in the same system (Celma and Cano 2008). In our study, we normalize the number of ratings per movie so that the approximated popularity ranges between 0% and 100%. Thus, given a set S of 12 movies where *num rating* (m_i) is the number of ratings movie m_i receives, the popularity score is defined as:

$$popularity\ \ score = \frac{1}{12}\sum_{m_i \in S}\frac{num\ \ rating(m_i)}{total\ \ ratings}\times 100\% \qquad (2)$$

There are three different conditions for levels of popularity - high, medium and low. Thus, given the top 60 recommendations, our algorithm chooses 12 recommendations to form lists that have either high, or medium, or low level of popularity. First, for each movie m_i in the top 60 recommendations the algorithm computes the popularity score based on the above equation. The algorithm then sorts these movies from highest to lowest based on their popularity scores. The algorithm selects the top 12 movies for the high-level condition, the bottom 12 movies for the low-level condition, and the 25th to 36th movies for the medium-level condition.

Serendipity Serendipity indicates how different all the items in a set of recommendations are when compared with items the user recently consumed (or rated). We can estimate the serendipity of a recommendation list by taking the average of all pair-wise distances of between one item in the recommendation list, and one item the user has rated (Zhang et al. 2012). We are interested in tag-genome distances, similar to our calculation of diversity. Thus, we also use the tag-genome feature space to compute serendipity. Since we want the serendipity measure to be bounded by 0 and 1, we use cosine distance (i.e. 1 - cosine similarity). Furthermore, because users needed to rate at least 15 movies to participate in our study, we approximate serendipity with the 15 movies a user rated most recently. Thus, for each user, given a set S of 12 movies and a set R of the 15 most-recently-rated movies, the serendipity score is defined as:

$$serendipity\ \ score = \frac{1}{12^{*}15}\sum_{m_i \in S}\sum_{r_j \in R} cosine\ \ distance\left(m_i, r_j\right) \qquad (3)$$

There are three different levels of serendipity - high, medium and low. Thus, given the top 60 recommendations, our algorithm chooses 12 recommendations to form lists that have either high, or medium, or low level of serendipity. First, for each movie m_i in the top 60 recommendations, the algorithm computes the serendipity score based on the above equation. The algorithm then sorts these movies from highest to lowest, based on their serendipity scores, and selects the top 12 movies for the high- level condition, the bottom 12 for the low-level condition, and the 25th to 36th movies for the medium-level condition.

3.1.4 Data analyses

While 1888 users participated in our study, we only analyzed data from 1635 users. We first excluded 253 users because of the following reasons:

- 203 users were assigned to the control condition
- 47 users did not answer the questions assessing their personality
- 3 users had fewer than 15 ratings prior to their participation in our study

Because our algorithms are designed around heuristics, they could not deliver the appropriate levels of diversity, popularity, or serendipity to 416 users. Thus, we excluded these users from analysis. In our future work, we will look at how to improve these algorithms. Table 1 shows the number of users per condition in our data analyses. We discuss why we exclude these users in detail in Appendix A.

3.2 Measures

3.2.1 Satisfaction with recommendation lists

In the first set of questions, we checked if users would notice the different levels of diversity, popularity, and serendipity of the movies in the recommendation list (see Table 2). Then we evaluated their satisfaction with the recommendations (see Table 3).

3.2.2 User personality

The second set consists of 10 questions to assess users' personality traits (see Table 4). We adopted these ten questions from the work of Gosling et al. (2003) (p. 525). We only use these 10 questions since we did not want to overload our users with many questions which in turn can lead to more noise data from user's answers. Each personality trait is measured by two questions, where one is the reverse of the other. Hence, for each personality trait we first reversed the question and took the average of the answers to the two questions. The average score per personality trait tells us the magnitude of this personality trait of a user.

Table 1 The number of users who participated, and the number of users who were included in analyses per condition

Condition	Block of recommendation property	Diversity			Popularity			Serendipity		
	Level	Low	Medium	High	Low	Medium	High	Low	Medium	High
# users who participated per condition		169	176	176	205	195	182	162	171	199
# users included in data analyses per condition		150	147	171	151	135	158	114	102	92

Figure 1 shows the distributions of the five personality traits. Our users expressed that they were more introverted than extroverted. Hence, in this study, we use introversion instead of extraversion where introversion is the reverse of extraversion. Likewise, since our users expressed that they were more emotionally stable, we use emotional stability instead of neuroticism where emotional stability is the reverse of neuroticism.

In this study, for each personality trait, we divide our users into two groups - a group with a high level (score greater than 4), and a group with a low level (score less than or equal to 4) of that personality trait. Dividing our users into two groups (high vs. low) allow easy and understandable contrasts when we analyze the interaction between the personality traits and the 10 experimental conditions.

4 Results

4.1 The diversity, popularity, and serendipity of recommendation lists created by algorithms based only on user ratings did not satisfy most users

With the first research question, our goal is to estimate how satisfied users would be with the diversity, popularity, and serendipity levels of recommendations generated by default ratings-based algorithms. Our data assesses users' satisfaction with the assigned category and level (e.g. high-diversity or low-serendipity). We therefore look at users who responded that the assigned level was Just Right.

These users are those who answered Just Right to the questions assessing their preferences for diversity, popularity and serendipity (i.e. the 1st, 2nd, and 3rd questions in Table 3). We report on how often these users would receive their preferred Just Right-level recommendations from ratings-based algorithms. The reported frequency is an estimate of user satisfaction with ratings-based algorithms. We may overestimate the satisfaction because we have no evidence that users who were unsatisfied with the given level would have been happy at another level. We also may underestimate the satisfaction because users who are happy at their assigned level might be happy at multiple levels.

Diversity For diversity, 233 out of 468 users reported Just Right satisfaction (92/171, 90/147, 51/150). Of these, only 88 would have received the preferred level of diversity from ratings-based algorithms (8/92, 72/90, 8/51). This ratio suggests that ratings- based algorithms could satisfy users with medium-diversity preferences 80% of the time. However, for many users who were happy with high- or low-diversity, ratings- based algorithms could satisfy only 9%–16% of the time.

Table 2 Questions to evaluate if users perceive the differences in recommendations with different levels of diversity, popularity, or serendipity

	Targeted block of recommendation properties	Questions	5-point scale
1	Diversity	How dissimilar are the movies in the list from each other?	● ● ● ● ● Very similar to each other · Neutral · Very dissimilar from Each other
2	Popularity	How popular are the movies in the list?	● ● ● ● ● Very obscure · Neutral · Very popular
3	Serendipity	How surprised are you to see these movies being recommended to you?	● ● ● ● ● Not surprised at all · Neutral · Very surprised

After seeing the 12-item recommendation list, users answered two sets of questions

Table 3 Questions to evaluate users' satisfaction with the recommendation lists

	Targeted block of recommendation properties	Factors about user satisfaction	Questions	How our users will answer the question
1	Diversity	Diversity preference	Is the level of dissimilarity among the recommended movies right for you?	On a 5-point scale with the order as follows: Far too similar - A bit too similar - Just right for me - A bit too dissimilar - Far to dissimilar
2	Popularity	Popularity preference	Is the level of popularity of the recommended movies right for you?	On a 5-point scale with the order as follows: Far too obscure - A bit too obscure - Just right for me - A bit too popular - Far too popular
3	Serendipity	Serendipity preference	Is the level of serendipity of the recommended movies right for you?	On a 5-point scale with the order as follows: Far too unsurprising - A bit too unsurprising - Just right - A bit too surprising - Far too surprising
4	All	Enjoyment	This list contains movies I think I would enjoy watching	On a 5-point Likert scale from Strongly Disagree to Strongly Agree

Popularity We carry out the same calculations as above and found that 89 out of 238 users would have received the preferred level of popularity from ratings-based algorithms (49/76, 36/81, 4/81). This fraction suggests that ratings-based algorithms could satisfy users with high-popularity preferences 65% of the time, and users with medium-popularity preferences 44% of the time. Although about one-third of 238 users preferred low-popularity, ratings-based algorithms could satisfy their satisfaction only 5% of the time.

Serendipity With the same calculations, we found that 46 out of 124 users would have received the preferred level of serendipity from ratings-based algorithms (0/41, 1/37, 45/46). This ratio suggests that ratings-based algorithms typically satisfy users with low-serendipity preferences (98% of the time) and that they will rarely satisfy users who have high- and medium-serendipity preferences (0% of the time).

Table 5 displays the number of users who are satisfied with the levels of diversity, popularity, and serendipity of recommendation lists created by rating-based algorithms.

4.2 User personality and user preference for diversity, popularity and serendipity

The second research question asks if there is a correlation between users' personality traits and their preferences for diversity, popularity, and serendipity. To answer this question, we tested the interaction effect between each personality trait and the levels of diversity, popularity, and serendipity in the recommendation list.

Table 4 The ten questions to assess user personality adopted from (Gosling et al. 2003)

#	Personality trait	Question
		I see myself as:
1	Agreeableness	.... critical, quarrelsome. (*)
2		.... sympathetic, warm.
3	Conscientiousness	.... dependable, self-disciplined.
4		.... disorganized, careless. (*)
5	Emotional Stability	.... anxious, easily upset. (*)
6		.... calm, emotionally stable.
7	Extraversion	.... extraverted, enthusiastic.
8		.... reversed, quiet. (*)
9	Openness to experiences	.... open to new experiences, complex.
10		.... conventional, uncreative. (*)

*denotes reversed questions. These questions are on 7 Likert scale

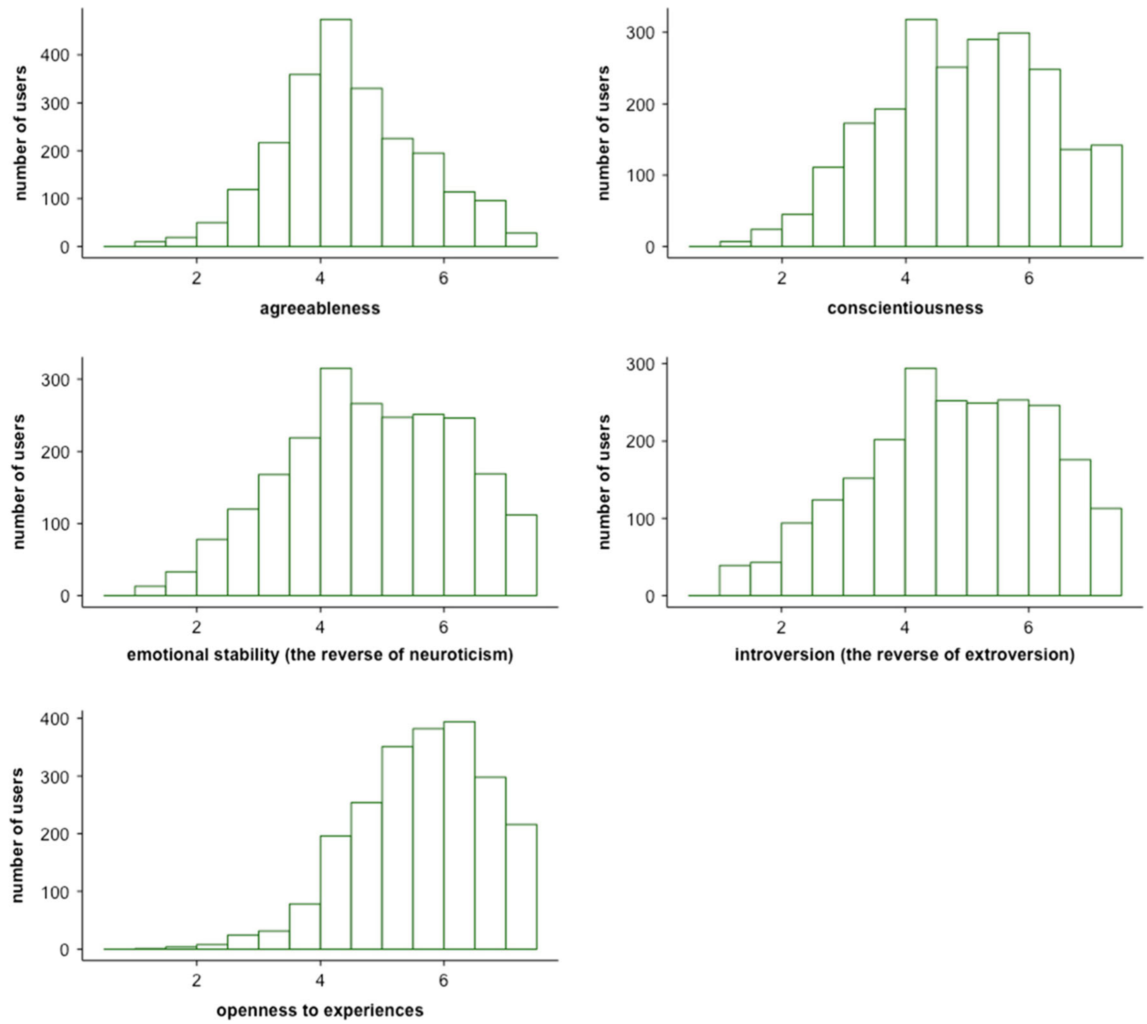

Fig. 1 The distributions of the five personality traits

Significant interaction effects suggest that users with different levels of personality trait have different preferences for how diverse, popular, and serendipitous a recommendation list should be. We used ordinal logistic regression[3] with the following model.[4]

$$foldedresponse \sim personality.level^{*}score \quad (4)$$

Ordinal logistic regression is appropriate in this case because the dependent variable is an ordered categorical variable, and we are interested in interpreting the probabilities of users choosing ordinal answers.

[3] using polr in R

[4] Although in our user experiment we have three categorical levels for diversity, popularity, and serendipity, we decide to analyze the data based on continuous variables representing the diversity, serendipity and popularity of recommendation lists. This is because the distributions for diversity, popularity and serendipity are either skewed or highly overlapped, as shown in Appendix A. Thus using continuous variables makes our analyses independent from the distributions and easily duplicated.

For all targeted blocks of recommendation properties, the dependent variable is users' answers for the 1st, 2nd, and 3rd questions assessing their preferences for diversity, popularity and serendipity respectively (as shown in Table 3). Because these questions have the most satisfactory option in the middle of the scales, we transformed the scales into new variables as follows. For diversity, we recoded *Far too similar* and *Far to dissimilar* options to *Far*, *A bit too similar* and *A bit too dissimilar* options to close, and *Just Right* to *Just Right.* For popularity, we recoded *Far too obscure* and *Far too popular* to *far*, *A bit too obscure* and *A bit too popular* to *close*, and *Just Right* to *Just Right*. For serendipity, we recoded *Far too unsurprising* and *Far too surprising* to far, *A bit too unsurprising* and *A bit too surprising* to *close*, and *Just Right* to *Just Right.*

The independent variable scores are the diversity, popularity, and serendipity scores computed based on eqs. 1, 2, and 3 respectively for diversity, popularity and serendipity.

Below, we describe the results for each recommendation property (diversity, popularity, and serendipity).

Table 5 Number of users who answered *Just Right* and number of users who would be satisfied with the levels of diversity, popularity and serendipity produced by ratings-based algorithms

	Diversity			Popularity			Serendipity		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
# Of users who answered Just Right	51	90	92	81	81	76	46	37	41
# Of the above users who would receive recommendations with their preferred levels of diversity, popularity, and serendipity created by ratings-based only algorithms	8	72	8	4	36	49	45	1	0

Diversity We found only the interaction effect between high-introversion and diversity score significant (p-value of interaction effect = 0.05[5,6]). This significant effect suggests that more high-introversion users preferred diverse recommendations than similar ones. Specifically, the effect suggests that 0.1 (10%) increases in diversity correspond to 1.12 increases in the odds of high-introversion users choosing Just Right. However, increases in diversity do not significantly change the odds of low-introversion users choosing Just Right. We visualize[7] the changes in odds of choosing Just Right for high- and low-introversion users in Fig. 2.

Popularity We found that only the interaction effect between introversion and popularity is significant (the interaction effect p-value = 0.02).[8] We also found only the main effect of conscientiousness significant (p-value = 0.02).

The interaction effect suggests that more low-introversion users preferred popular recommendations than non-popular ones, but more high-introversion users preferred non-popular recommendations than popular recommendations. A 10% increases in popularity score corresponds to a 1.01 increase in the odds of low introversion users choosing Just Right, but a 0.55 decrease in the odds of high-introversion users choosing Just Right. We visualize the changes in the odds of high- and low- introversion users choosing Just Right in Figs. 3 and 4.

The main effect of conscientiousness suggests that no matter how popular the recommendation lists, low-conscientiousness users were more likely to choose Just Right (60% of the time) than high-conscientiousness users (48% of the time).[9]

Serendipity We found three significant interaction effects. These effects are between serendipity and the conscientiousness trait (p-value of the interaction effect = 0.05), between serendipity and the introversion trait (p-value of the interaction effect = 0.00), and between serendipity and the openness trait (p-value of the interaction effect = 0.03).

We found that more high-conscientiousness users preferred low serendipity recommendations over high, and that low-conscientiousness users preferred high serendipity over low. More high-introversion users preferred high serendipity over low, and that more low-introversion users preferred low serendipity over high. More low-openness users preferred low serendipity recommendations over high. High-openness users did not have any clear preference for the level of serendipity.[10] These significant interaction effects suggest that a 0.1 (10%) increase in serendipity corresponds to:

* A 1.06 increase in the odds of low-conscientiousness users, and a 0.86 decrease in the odds of high-conscientiousness users choosing Just Right.
* A 0.82 decrease in the odds of low-introversion users, and a 1.03 increase in the odds of high-introversion users choosing Just Right.
* A 0.76 decrease in the odds of low-openness users choosing Just Right. The change in the odds of high-openness users choosing Just Right is negligible.

4.2.1 User satisfaction and user enjoyment

In summary, we found that introversion, conscientiousness, and openness personality traits contain strong signals about user preference for diversity, popularity and serendipity. However, delivering individually preferred amounts of diversity, popularity or serendipity does not necessarily lead to the increase in user enjoyment. In fact, we found correlations, though not big, between user satisfaction with recommendation diversity, popularity and serendipity. The correlation between user

5 The statistical significance test reported is of the interaction effect only; i.e., of the difference between the odds-ratios of high- and low–introversion users

6 The model is reported in the Appendix B Table 6

7 This and the subsequent visualizations are generated as follows. First, we bucket the corresponding continuous score (in this case diversity score). Then, for each bucket, we compute the percentage of users who answered *Just Right* to the corresponding questions (in this case, the question assessing user preference for diversity). Thus, a dot represents the percentage of a personality-trait group per bucket. However, the line is fitted based on the continuous score, not on the bucketed percentages.

8 The model is reported in Appendix B Table 7

9 The model is reported in Appendix B Table 11

10 These models are reported in Appendix B tables 8, 9, and 10. $response \sim personality.level \square score$ (5)

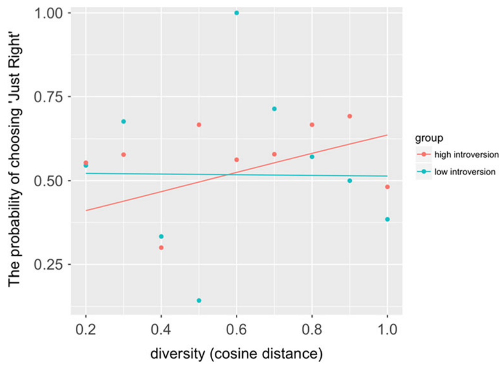

Fig. 2 The probabilities of high and low introverted users choosing *Just Right* with different diversity quantity

satisfaction with the diversity of the recommendation list and the user enjoyment of the movies in the list is 0.21. The correlation is 0.24 between users' satisfaction with the popularity of the list and their enjoyment, and 0.12 between users' satisfaction of the serendipity of the list and their enjoyment. Therefore, in the next section, we investigate the relationship between user personality and user enjoyment for recommendation lists with different levels of diversity, popularity and serendipity.

4.3 User personality and user enjoyment for recommendations with different levels of diversity, popularity and serendipity

Research question 3 asks if there is a correlation between users' personality traits and their satisfaction with the list of recommendations. For each of the experimental conditions, we test if the interaction effects of personality traits and the self-reported enjoyment (to infer user satisfaction) of the movies in the recommendation list were significant. We employed ordinal logistic regression with the following model:

The dependent variable is the self-reported enjoyment of the recommendations (users' answers to the 4th question shown in Table 3. For ease of interpretation, we recoded users' answers to this question from a 5-point Likert scale ranging from Strongly Disagree to Strongly Agree into three categories: Disagree, Neutral, and Agree.

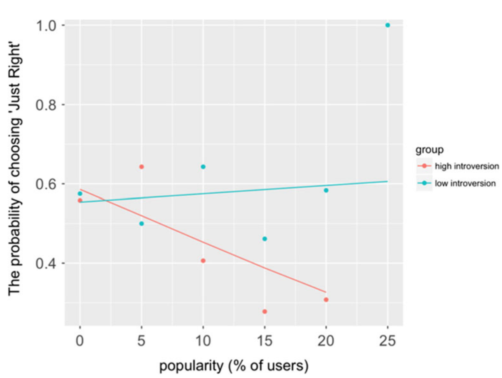

Fig. 3 The probabilities of high and low introverted users choosing *Just Right* with different popularity quantity

Diversity We found only the interaction effect between diversity and the introversion trait significant (the p-value of the interaction effect is 0.02).[11] The significant effect means low-introversion users did not enjoy the recommendations in a diverse list as much as they enjoyed the recommendations in a less diverse list. Specifically, 0.1 (10%) increases in the diversity of the list correspond to 0.86 decreases in the odds of low-introversion users expressing that they would enjoy the recommendations in the list. Meanwhile, high-introversion users would equally enjoy recommendations from a diverse list or an unvaried list. We visualize the changes in the odds of enjoying the recommendations for high- and low-introversion users in Fig. 5 below.

Popularity We found only the interaction effect between high emotional stability and enjoyment significant (the p-value of the interaction effect is 0.02).[12] The interaction effect shows that high-emotional-stability users were more likely to enjoy recommendations from a popular list than low-emotional-stability users. Specifically, it suggests that a 10% increase in the popularity of the list corresponds to a two times (2×) increase in the odds of high-emotional-stability users self-reporting that they would enjoy the recommendations in the list. However, low-emotional-stability users would equally enjoy recommendations from a popular or a less popular list. We visualize the changes in the odds of enjoying the recommendations for high- and low-emotional-stability users in Fig. 6.

Serendipity We found no significant interaction effect between personality traits and serendipity scores (all p-values >0.1).

5 Conclusion

In summary, we found that users have different preferences for diversity, popularity, and serendipity. We also found that ratings-based algorithms do not consistently deliver the diversity, popularity, and serendipity levels preferred by individual users. On the other hand, personality traits provide important signals to predict users' individual preferences and overall recommendation enjoyment.

[11] The model is reported in Appendix B Table 7.
[12] The model is reported in Appendix B Table 8

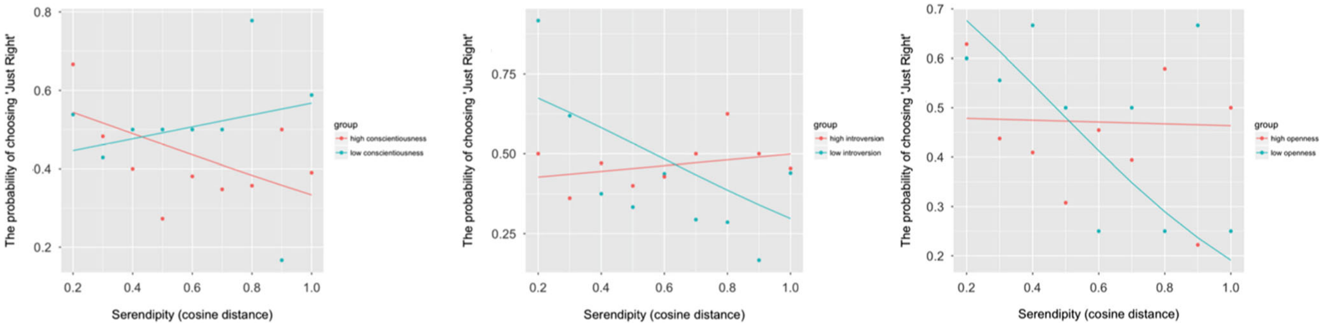

Fig. 4 The probabilities of users choosing *Just Right* with different amounts of serendipity. Left: conscientiousness; Middle: introversion; Right: openness

Our findings demonstrate that systems can recommend content with more appropriate levels of diversity, popularity, and serendipity by integrating personality information. Prior work concentrated on incorporating the information to improve prediction accuracy. For example Hu and Pu (2011) blended the personality information into collaborative filtering frameworks and saw improvements in MAE and ROC sensitivity metrics. Tkalcic et al. (2009) showed that using personality-based user-similarity in a collaborative filtering algorithm does not diminish the prediction accuracy when compared with ratings-based algorithms. Other researchers focused on different aspects of user experience such as acceptance issues of personality-based recommender systems (Hu and Pu 2009). Our work contributes to the understanding of improving the overall user experience from the angle of user preferences for diversity, popularity, and serendipity.

While it is not common today for a recommender system to know personality traits, we are not advocating giving a personality test before signing up. There is an increasing amount of user data available through federated sites, including sites that claim to have personality information. It is reasonable to believe that future commercial sites (e.g., Amazon, Facebook or Pinterest) will have access to personality assessment data for at least some of their users. Moreover, it's not necessary to always assess user personality of all users via questionnaires. Prior work showed that we could reliably predict users' personality based on their online footprint. Quercia et al. (2011) and Bachrach et al. (2012) demonstrated that we could effectively predict users' personality based on their Twitter or Facebook profiles. Youyou et al. (2015) went further to prove that well-trained algorithms could do a better job than humans at predicting user personality. These implicitly inferred personality signals could be used to improve recommendations (Wu and Chen 2015).

Our work also confirms that satisfaction with recommendations is a property of the entire set of recommendations and not only of the individual items recommended (e.g. (McLaughlin and Herlocker 2004; McNee et al. 2006)). This satisfaction, in turn, has strong connections with the other properties of the recommendation sets, such as diversity and serendipity. Although popularity is a property of an individual recommendation, our study proves that when considered as a property of a set, popularity also has a strong connection with user satisfaction. We can infer the directions of these connections via users' personality traits.

We think it is important to continue exploring the ways in which personality data can be used to improve user experience with recommender systems. One direction for future work is to take the n-dimensional matrix factorization approach proposed by Karatzoglou et al. (Karatzoglou et al. 2010) and treat the effects of personality traits that we found in our study as

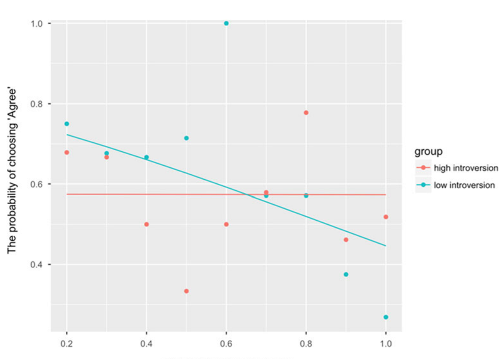

Fig. 5 The probabilities of high and low introverted users reported that they would enjoy the recommendations with different diversity quantity

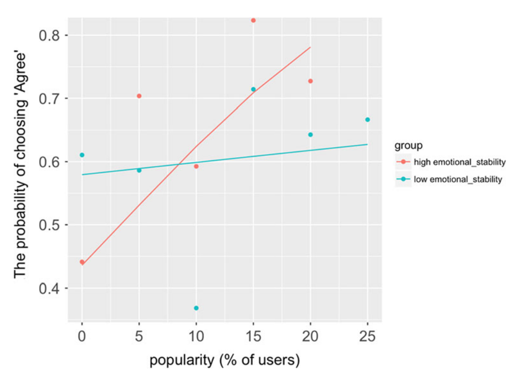

Fig. 6 The probabilities of high and emotional stability users reported that they would enjoy the recommendations with different popularity quantity

new dimensions. An alternative is to consider Bayesian personalized ranking (e.g. Rendle et al.'s work (Rendle et al. 2009)), treating the effects of personality as a prior belief. We also believe that future work should investigate the effects of all five-personality traits together, as each trait plays an equal role in influencing users' decisions.

Our study is not without limitation. The choice of using a between-subjects design for our experiment, even after careful evaluations, poses several challenges for our analyses. First, we have limited data per user, and our data does not have contrast in users' preferences for different levels of diversity, popularity and serendipity. We also cannot calibrate our users' desired levels of diversity, popularity, and serendipity. Moreover, we had to exclude users for whom randomly assigned levels of diversity, popularity, and serendipity could not be achieved. Furthermore, the statistical models that we used do not perfectly fit our observed data. This may be because we used only ten questions to approximate user personality and that the distributions of user personality questions are skewed.

We also only examine user satisfaction and user personality in the movie domain with a specific recommender system - MovieLens. In the future, we would like to conduct similar research in different domains - such as books or news. Duplicating this study in different domains will enhance our knowledge of user satisfaction with recommendations.

Appendix A Reason to exclude some users from the Analyses

The score ranges from 0 to 1. Red indicates high level, blue medium level and green low level. There are three curves indicating the three peaks of high, medium and low levels of popularity.

Before explaining why we exclude some users from our main analyses, we discuss the results of our manipulation-check analyses, which reveal reasons for the exclusion.

Our manipulation-check analyses show that:

- for the diversity metric, users perceived the differences in recommendations of high and low levels (p-value = 0.000), and of medium and low levels (p-value = 0.000). Users did not perceived the differences in recommendations of medium and high levels (p-value = 0.255).
- for the popularity metric, users perceived the differences in recommendations of high and low levels (p-value = 0.000), in recommendations of medium and high levels (p-value = 0.000). Users also perceived the differences in recommendations of medium and low levels (with marginally p-value = 0.057).
- for the serendipity metric, users perceived the differences in recommendations of high and low levels (p-value = 0.035), in recommendations of medium and low levels (p-value = 0.021). Users did not perceive the differences in recommendations with medium and high levels (p-value = 0.956).

We plot out the distributions of diversity, popularity, and serendipity to investigate why in some level comparisons (e.g. high vs. medium diversity level) users did not perceive the differences. Figures 7 and 8 show these distributions. In each distribution, we observe overlapping regions of high, medium and low levels. These overlapping make users not perceive differences in some recommendations with different levels.

Both scores range from 0 to 1. Red indicates high levels, blue medium levels and green low levels. In each figure, there are three curves indicating the three peaks of high, medium, and low levels of diversity (left figure) or of serendipity (right figure)

In this study, we want to examine the recommendation experience of users who perceived the differences in the diversity, popularity, or serendipity quantities per level. Thus, we remove from our analyses users to whom the recommendations cannot deliver to the appropriate quantities of diversity, popularity, and serendipity.

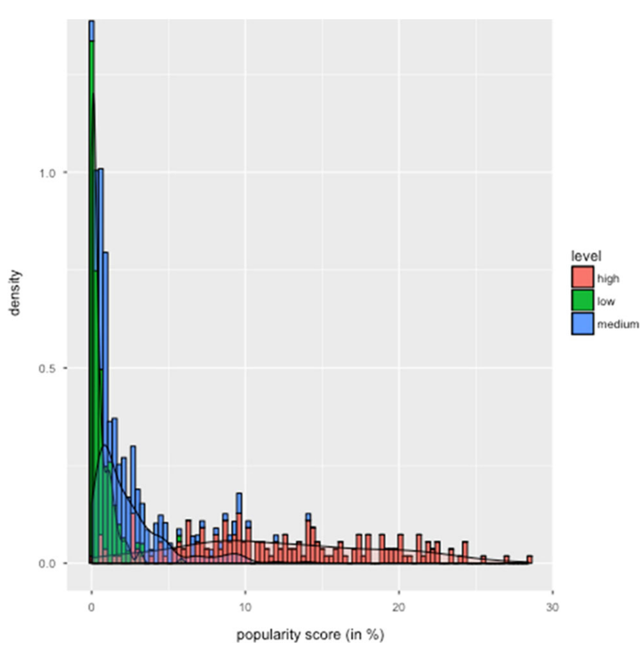

Fig. 7 The distributions of popularity scores

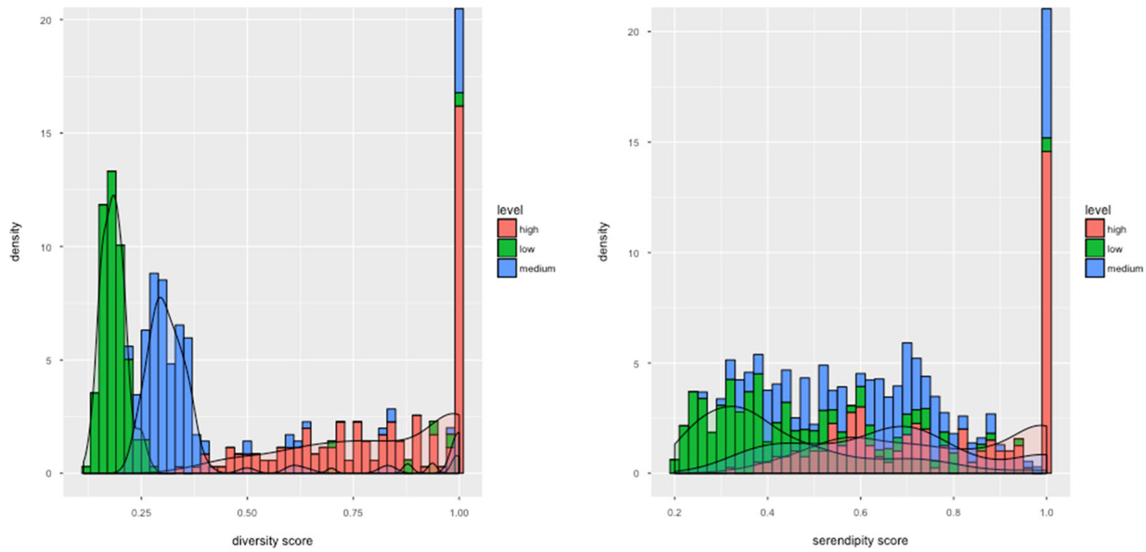

Fig. 8 The distributions of diversity scores (on the left) and serendipity scores (on the right)

Our process to remove users from the analyses is as follow.

Diversity From Fig. 8 (left), there are clear cut-off points for low, medium and high levels of diversity at 0.24 and 0.41. Thus, we analyze 150/169 users who were assigned to the low-level with diversity scores less than 0.24, 147/176 users assigned to the medium-level with diversity scores from 0.24 to 0.41, and 171/176 users assigned to the high level with scores greater than 0.41.

Popularity From Fig. 7, there are clear cut-off points for low, medium and high levels at 0.6% and 5.0%. Thus, we analyze 151/205 users assigned to the low-level with popularity scores from 0% to less than 0.6%, 135/195 users assigned to the medium-level with popularity scores from 0.6% to 5.0%, and 158/182 users assigned to the high-level with popularity scores greater than 5.0%.

Serendipity From Fig. 8 (right), there are clear cut-off points for low, medium and high levels of serendipity at 0.46 and 0.80. Thus, we analyze 114/162 users who were assigned to the low-level with serendipity scores less than 0.46, 102/171 users assigned to the medium-level with serendipity scores from 0.46 to 0.80, and 92/199 users assigned to the high-level with serendipity score greater than 0.80.

Appendix B Tables

Table 6 Ordinal regression results for diversity preferences of introversion personality trait

Variables	Logistic coefficient	Standard error	95% CI	*p* value	Increase in odds ratio
Personality level					
High	−0.69	0.33	[−1.34, −0.03]	0.04	0.50
Low	Ref				
Diversity quantity					
Score (cosine distance)	−0.04	0.46	[−0.95, 0.87]	0.90	0.96
Interaction effect					
high*score	1.19	0.62	[−0.02, 2.41]	0.05	3.29
low*core	Ref				

Goodness of fit: Deviance chi-square = 790, df = 463 (p = 0.00)

Table 7 Ordinal regression results for popularity preferences of introversion personality trait

Variables	Logistic coefficient	Standard error	95% CI	p value	Increase in odds ratio
Personality level					
High	0.13	0.25	[−0.35, 0.61]	0.50	1.14
Low	Ref				
Popularity quantity					
Score (% of users rated)	0.001	0.02	[−0.03, 0.05]	0.60	1.0
Interaction effect					
high*score	−0.06	0.03	[−0.12, −0.01]	0.02	0.93
low*score	Ref				

Goodness of fit: Deviance chi-square = 742, df = 439 (p = 0.00)

Table 8 Ordinal regression results for serendipity preferences of conscientiousness personality trait

Variables	Logistic coefficient	Standard error	95% CI	p value	Increase in odds ratio
Personality level					
High	0.73	0.58	[−0.41, 1.87]	0.20	2.01
Low	Ref				
Serendipity quantity					
Score (cosine distance)	0.60	0.69	[−0.75, 1.97]	0.30	1.84
Interaction effect					
high*score	−1.70	0.87	[−3.41, 0.00]	0.05	0.18
low*score	Ref				

Goodness of fit: Deviance chi-square = 536, df = 303 (p = 0.00)

Table 9 Ordinal regression results for serendipity preferences of introversion personality trait

Variables	Logistic coefficient	Standard error	95% CI	p value	Increase in odds ratio
Personality level					
High	−1.49	0.61	[−2.69, −0.29]	0.02	0.23
Low	Ref				
Serendipity quantity					
Score (cosine distance)	−1.99	0.71	[−3.42, −0.61]	0.00	0.14
Interaction effect					
high*score	2.34	0.88	[0.62, 4.11]	0.00	10.44
low*score	Ref				

Goodness of fit: Deviance chi-square = 534, df = 303 (p = 0.00)

Table 10 Ordinal regression results for serendipity preferences of openness personality trait

Variables	Logistic coefficient	Standard error	95% CI	*p* value	Increase in odds ratio
Personality level					
High	1.35	0.80	[−2.92, 0.21]	0.08	0.25
Low	Ref				
Serendipity quantity					
Score (cosine distance)	−2.73	1.08	[−4.90, −0.62]	0.01	0.07
Interaction effect					
high*score	2.66	1.17	[0.35, 4.95]	0.02	14.17
low*score	Ref				

Goodness of fit: Deviance chi-square = 535, df = 303 ($p = 0.00$)

Table 11 Ordinal regression results for popularity preferences of conscientiousness personality trait (main effects)

Variables	Logistic coefficient	Standard error	95% CI	*p* value	Increase in odds ratio
Personality level					
High	−0.46	0.19	[−0.84, 0.08]	0.02	0.63
Low	Ref				

Table 12 Ordinal regression results for users' self-reported enjoyment with different diversity quantities for introversion personality trait

Variables	Logistic Coe cient	Standard error	95% CI	*p* value	Increase in Odds Ratio
Personality level					
High	-0.95	0.35	[-1.65, -0.27]	0.00	0.39
Low	Ref				
Diversity quantity					
Score (cosine distance)	-1.47	0.46	[-2.37, -0.58]	0.00	0.23
Interaction e ect					
High*score	1.46	0.60	[0.28, 2.64]	0.02	4.32
Low*score	Ref				

Goodness of t: Deviance chi-square = 877, df = 463 ($p = 0.00$)

Table 13 Ordinal regression results for users' self-reported enjoyment with different popularity quantities for emotional stability personality trait

Variables	Logistic Coe cient	Standard error	95% CI	*p* value	Increase in Odds Ratio
Personality level					
High	-0.57	0.24	[-1.06, -0.11]	0.02	0.56
Low	Ref				
Diversity quantity					
Score (cosine distance)	0.00	0.02	[-0.03, 0.05]	0.60	1.01
Interaction e ect					
High*score	0.07	0.03	[0.01, 0.13]	0.02	1.01
Low*score	Ref				

Goodness of t: Deviance chi-square = 866, df = 439 ($p = 0.00$)

References

Adomavicius, G., & Kwon, Y. (2009). Toward more diverse recommendations: Item re-ranking methods for recommender systems. In *Workshop on Information Technologies and Systems*.

Ali, K., & van Wijnand, S. (2004). TiVo: making show recommendations using a distributed collaborative filtering architecture. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 394–401.

Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). Personality and patterns of Facebook usage. In *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 24–32.

Celma, O., & Cano, P. (2008). From hits to niches?: Or how popular artists can bias music recommendation and discovery. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*. ACM, 5.

Chang, S., Maxwell Harper, F., & Terveen, L. (2015). Using groups of items for preference elicitation in recommender systems. In *Proceedings of the 18th ACM conference on computer supported cooperative work; social computing (CSCW '15)* (pp. 1258–1269). New York, NY, USA: ACM.

Chausson, O. (2010). Who watches what?: Assessing the impact of gender and personality on film preferences. Paper published online on the MyPersonality project website. http://mypersonality.org/wiki/doku.

Chen, L., Wu, W., & He, L. (2013). How personality influences users' needs for recommendation diversity? In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 829–834.

Costa, P. T., & McCrae, R. R. (2008). The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment, 2*, 179–198.

Ekstrand, M. D., Maxwell Harper, F., Willemsen, M. C., & Konstan, J. A. (2014). User perception of differences in recommender algorithms. In *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 161–168.

Fleder, D. M., & Hosanagar, K. (2007). Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce*. ACM, 192–199.

Gosling, S. D., Rentfrow, P. J., & Swann Jr., W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality, 37*(6), 504–528.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5–53.

Hu, R., & Pu, P. (2009). Acceptance issues of personality-based recommender systems. In *Proceedings of the third ACM conference on Recommender systems*. ACM, 221–224.

Hu, R., & Pu, P. (2011). Enhancing collaborative filtering systems with personality information. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 197–204.

John, O. P., & Srivastava, S. (1999). The big five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research, 2*, 102–138.

Karatzoglou, A., Amatriain, X., Baltrunas, L., & Oliver, N. (2010). Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In proceedings of the fourth ACM conference on recommender systems. ACM, 79–86.

Kemp, A. E. (1996). *The musical temperament: Psychology and personality of musicians*. Oxford: Oxford University Press.

Kraaykamp, G., & van Eijck, K. (2005). Personality, media preferences, and cultural participation. *Personality and Individual Differences, 38*(7), 1675–1688.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81.

McLaughlin, M. R., & Herlocker, J. L. (2004). A collaborative filtering algorithm and evaluation metric that accurately model the user experience. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 329–336.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006). Being accurate is not enough: How accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*. ACM, 1097–1101.

Nguyen, T. T., Hui, P.-M., Maxwell Harper, F., Terveen, L., & Konstan, J. A. (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*. ACM, 677–686.

Oh, J., Park, S., Yu, H., Song, M., & Park, S.-T. (2011). Novel recommendation based on personal popularity tendency. In *In Data Mining (ICDM), 2011 I.E. 11th International Conference on. IEEE* (pp. 507–516).

Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011). Our twitter profiles, our selves: Predicting personality with twitter. In *In Privacy, Security, Risk and Trust (PASSAT) and 2011 I.E. Third International Conference on Social Computing (SocialCom), 2011 I.E. Third International Conference on. IEEE* (pp. 180–185).

Rendle, S., Freudenthaler, C., Gantner, Z., & SchmidtThieme, L. (2009). BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. AUAI press, 452–461.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84*(6), 1236.

Shephard, R. W., & Fare, R. (1974). *The law of diminishing returns*. Berlin: Springer.

Steck, H. (2011). Item popularity and recommendation accuracy. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 125–132.

Tkalcic, M., Kunaver, M., Tasic, J., & Kosir, A. (2009). Personality based user similarity measure for a collaborative recommender system. In *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real world challenges*. 30–37.

Vargas, S., & Castells, P. (2013). Exploiting the diversity of user preferences for recommendation. In *Proceedings of the 10th conference on open research areas in information retrieval* (pp. 129–136). Le centre de hautes etudes internationals d'informatique documentaires.

Jesse Vig, Shilad Sen, and John Riedl. (2012). The tag genome: Encoding community knowledge to support novel interaction. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 3, 13.

Wu, W., & Chen, L. (2015). Implicit acquisition of user personality for augmenting movie recommendations. In *In International Conference on User Modeling, Adaptation, and Personalization. Springer* (pp. 302–314).

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*. 112(4), 1036–1040.

Zhang, M., & Hurley, N. (2009). Novel item recommendation by user profile partitioning. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*. IEEE computer society, 508–515.

Zhang, Y. C., Seaghdha, D. O., Quercia, D., & Jambor, T. (2012). Auralist: Introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 13–22.

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. ACM, 22–32.

Tien T. Nguyen is now a software engineer focusing on data analytics at Pinterest. He obtained his Ph.D. from the University of Minnesota, Twin-Cities in Computer Science. His research interests are recommender systems, applied machine learning, data mining and human-computer interaction. He has served in program committee for top-tier conferences in recommender system such as RecSys 2017, WWW 2017. He is also a reviewer for top-tier conferences in human-computer interaction such as CHI, CSCW, and IUI.

F. Maxwell Harper is a research scientist with the GroupLens Center for Social and Human-Centered Computing at the University of Minnesota. He is an experienced software engineer, with nine years of professional development experience, as well as an experienced researcher, with over 30 peer-reviewed publications in the areas of human-computer interaction, social computing, and recommender systems. He is currently the lead developer for MovieLens, a research platform with thousands of active monthly users.

Loren Terveen is a Professor of Computer Science at the University of Minnesota. Before joining the University of Minnesota, he received his PhD in Computer Sciences from the University of Texas at Austin, and then spent 11 years at AT&T Labs / Bell Labs. Terveen is an expert in the areas of human-computer interaction and social computing. He has published over 100 scientific papers, holds 9 patents, has advised several startup companies, consulted on intellectual property cases, and has held many leadership positions in his profession, including serving as President of ACM's Special Interest Group on Computer-Human Interaction and on the CM Council.

Joseph A. Konstan is Distinguished McKnight Professor and Distinguished University Teaching Professor in the Department of Computer Science and Engineering at the University of Minnesota, where he also holds the title Adjunct Professor of Public Health. His research addresses a variety of human-computer interaction issues, including personalization (particularly through recommender systems), eliciting online participation, designing computer systems to improve public health, and ethical issues in research online. He is probably best known for his work in collaborative filtering recommenders (the GroupLens project, which recently won the ACM Software Systems Award). Dr. Konstan received his Ph.D. from the University of California, Berkeley in 1993. He is a Fellow of the ACM, AAAS, and IEEE, a member of the CHI Academy, Past-President of ACM SIGCHI, the 4500-member Special Interest Group on Human-Computer Interaction, and a three-time member of the ACM Council. He chaired the first ACM Conference on Recommender Systems in 2007, and has also chaired ACM's UIST 2003 and CHI 2012 conferences. He is currently serving as co-Chair of ACM's Publications Board.