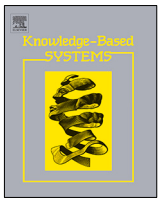
Knowledge-Based Systems 275 (2023) 110699

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys



On the problem of recommendation for sensitive users and influential items: Simultaneously maintaining interest and diversity

Alvise De Biasio [a,b,*], Merylin Monaro [c], Luca Oneto [d], Lamberto Ballan [a], Nicolò Navarin [a]

[a] Department of Mathematics, University of Padova, Via Trieste 63, Padova, 35131 PD, Italy
[b] estilos srl, Via Ca' Marcello 67/D, Venezia, 30172 VE, Italy
[c] Department of General Psychology, University of Padova, Via Venezia 8, Padova, 35131 PD, Italy
[d] DIBRIS, University of Genova, Via Opera Pia 11/A, Genova, 16145 GE, Italy

ARTICLE INFO

Article history:
Received 23 May 2022
Received in revised form 22 April 2023
Accepted 5 June 2023
Available online 9 June 2023

Keywords:
Recommender systems
Machine learning
Personality traits
Diversity
Social networks

ABSTRACT

Recommender systems, in real-world circumstances, tend to limit user exposure to certain topics and to overexpose them to others to maximize performance. However, repeated exposure to biased content could lead to the so-called echo chamber phenomenon: especially in social network environments, people encounter only information that reflects their previous beliefs and opinions, reinforcing them. This phenomenon could have worrying consequences for society, including the spread of aggressive, unhealthy, or risky behaviors. Some persons can be more affected than others by echo-chambers. We define as *sensitive* the users whose behavior could be influenced by the over- or under-exposure to certain items due to the echo-chamber effect, and as *influential* the items that could influence the behavior of such users. In this paper, we address the problem of recommending influential items to sensitive users. We formalize the problem and propose three techniques that can be used to diversify the distributions of influential items in order to positively affect sensitive users' behavior. Recommendations that meet this *diversity* criterion could potentially avoid dangerous societal consequences and simultaneously promote healthier lifestyles. We tested the proposed techniques in a real-world dataset by considering two different case studies that involved potentially aggressive and potentially depressed users. All techniques have been proven to be effective and allow high performance to be maintained while diversifying recommendations.

1. Introduction

Recommender systems are a fast-growing technology that aims to provide item suggestions to users [1,2]. We encounter these systems every time we look for a product to buy on an online site, a news story to read, or a social page to follow. Recommender systems help us make our choices by suggesting subsets of items that are more likely to be of interest to us. This high level of personalization helps us to focus on the most important items when we have to choose from a large number of alternatives [3].

However, in real-world circumstances, the tendency of these systems to encourage selective user exposure to a subset of content in order to maximize performance could result in controversial effects. In the social networks field, the repeated exposure to certain types of information would lead to the occurrence of echo chambers [4], i.e., environments in which users reinforce their position on certain topics due to repeated exposure to similar contents. According to the theory of group polarization [5], this repeated exposure would lead users in a group to develop extreme positions without evaluating alternative positions. Therefore, while in some cases the effects of polarization are limited to extremization of public opinion or to the reinforcement of beliefs and bias, in other cases, they may be associated with an increased risk of violent and aggressive behaviors [6] or with the spread of unhealthy and risky behaviors (e.g., drug use, self-injury).

In our paper, we address the problem of optimizing the performance of a recommender system that diversifies the distribution of certain items to positively affect the behavior of some users who may be more sensitive than others to specific topics. We define as *sensitive*, the users whose behavior can be influenced by certain items (e.g., depressed and aggressive users). Similarly we refer to as *influential*, the items that can influence the behavior of sensitive users. Based on the effect of influential items on the behavior of sensitive users, we distinguish two subgroups of items: *controversial* and *favorable* items. We define as *controversial* the items that could have negative consequences on the behavior of sensitive users (e.g., violence, delinquency, and weapons content),

* Corresponding author at: Department of Mathematics, University of Padova, Via Trieste 63, Padova, 35131 PD, Italy.
E-mail address: alvise.debiasio@phd.unipd.it (A. De Biasio).

https://doi.org/10.1016/j.knosys.2023.110699
0950-7051/

and as *favorable* the items that could have a positive impact (e.g., sports, and hobbies content).

Currently, to the best of the authors knowledge, there are no specialized papers in the literature that focus on the problem we address in this paper. However, there are related research fields that specifically address diversity and fairness, albeit with different research goals than ours. The main goal of studies in the field of diversity [7] is to diversify recommendations to increase user satisfaction, reflecting the full spectrum of user interests or including unexpected items that might be of interest to the user. Instead, the goal of major studies in the field of fairness [8,9] is to remove bias in model predictions toward certain user groups who are considered protected based on certain attributes. The proposed strategies in both research fields are quite general and cannot be optimally applied to our problem. Diversity methods traditionally make no distinction between sensitive and non-sensitive users. Fairness methods, on the other hand, do not consider how certain influential items, if recommended, may impact the behavior of sensitive users.

In this paper we propose two approaches that take inspiration from both the research fields of diversity and fairness. The first technique we propose is a redesign of a well-known [10] calibration algorithm. The original algorithm allows the topic proportions of a user's recommendations to be calibrated based on the topic proportions of the ground truth. To suit the method to our context, we modified the objective function in order to calibrate the proportions of influential items for sensitive users based on the distribution of the same items for non-sensitive users. The second technique we propose takes inspiration from an existing algorithm [11] that aims to maximize the value of a ranking under a set of constraints. In this case, we redesigned the algorithm to maximize the expected value of recommendations to sensitive users according to constraints based on target percentages of influential items.

Moreover, we also propose a joint approach that can be used to combine the outputs of the two techniques and any additional ranker together to further improve the results taking inspiration from the literature of hybrid recommenders [12] and rank fusion [13].

Compared to the current literature [7–9] that traditionally proposes very general methods that only consider item or user characteristics separately, the methods we propose consider them simultaneously. This allows for appropriate diversification of recommendations, avoiding potentially negative consequences on sensitive users behavior while simultaneously leading to potentially positive implications for society, without significantly affecting the overall performance of the recommender system.

We evaluated our approaches by exploiting a subset of a real-world dataset containing the social data of 92,255 users who completed a self-reported psychological questionnaire to determine their personality profile according to the Big Five personality theory [14]. We used correlations between personality traits and respectively depressive disorders [15] and violent behaviors [16] to conduct a comprehensive case analysis on two subsets of *sensitive* users that can be negatively affected by the echo chamber effect, i.e. *potentially depressed* and *potentially aggressive* users. We compared the results obtained with a strong baseline algorithm [17] in the diversity literature, which was the algorithm best suited in the literature to address the problem proposed in this paper. All proposed techniques proved successful in diversifying the distribution of influential items in sensitive user recommendations while maintaining high overall performance.

The main contributions of this article can be summarized as follows:

- we formalize the problem of maximizing the performance of a recommender system that diversifies the recommendations of influential items for sensitive users;
- we introduce two techniques that could be used to address the proposed problem taking inspiration from diversity and fairness studies and a joint approach that can be used to combine the output of any technique together to achieve better results;
- we conducted a full case analysis about potentially depressed and aggressive users based on a real-world dataset to test the proposed techniques.

The remainder of this paper is organized as follows. In Section 2 we introduce background concepts and discuss related work. In Section 3 we formalize the problem. In Section 4 we describe the techniques used to address the problem. In Section 5 we present the experiments and results obtained on a real-world dataset. In Section 6 we discuss the strengths and drawbacks of the proposed techniques and the limitations of the study. Section 7 concludes the article and highlights future research directions.

2. Background and related work

In this section, we introduce some background concepts. In Section 2.1 we introduce Recommender Systems and discuss some studies in the field of diversity and bias. In Section 2.2 we introduce the Big Five personality model and present the personality traits related to depressive disorders and aggressive behaviors.

2.1. Recommender systems, diversity, and bias

A *Recommender system* (*RS*) is a technology that aims to make suggestion of items to users [1,2]. In the RS field, an item is a general term used to denote what the system suggests to users. The suggestions, also known as recommendations, can concern: products to buy, news to read, social pages to follow, and others depending on the domain in which the system is used. These domains include many different industry sectors, such as e-commerce [18], online streaming [19], social networks [20], and others [21]. A recommender system is designed primarily to support a user evaluating a large number of alternatives [3] by proposing subsets of items that could be of greatest interest. In addition to this core functionality, the factors for which a service provider would want to use this technology are many [22], including increasing sales and profitability and improving user satisfaction and retention [23].

Recommender systems are mainly designed to determine the degree of user interest in a particular item (*Prediction Problem*) or to identify a set of k items of interest to a certain user (*Top-k Recommendation Problem*) [24]. Regardless of the problem they are designed to solve, recommendations for a user are offered primarily in the form of a ranking of items. Within this ranking, the top items are the products or services that best suit the user's preferences. To generate the ranking, the RS relies on user information collected either explicitly (e.g., product review ratings, item features, and others) or implicitly, based on the user's interactions with the platform hosting the service.

According to one of the best-known taxonomies [25], recommender systems can be divided into: *Collaborative Filtering* (*CF*), *Content Based Filtering* (*CB*) and *Hybrid Systems* (*HS*). CF [26,27] is a class of algorithms based on a very intuitive principle: people's decisions are often influenced by the advice and behavior of other people. Therefore, CF recommends to the user the items of interest from users similar to him/her according to some explicit (e.g., ratings, likes) or implicit (e.g., clicks, impressions) feedback measure. CB [28,29] was developed based on a different principle: people tend to choose products with characteristics similar to those of other products they have chosen in the past.

On the other hand, HS [12] considers both the above principles to further improve system performance. Within this taxonomy, further distinctions can be made. CF systems can be divided into neighborhood and model-based methods [30]. Neighborhood methods exploit similarity criteria on items or users to build new recommendations. Model-based methods, instead, build a predictive model from stored ratings to make recommendations.

Three of the main approaches used in the state-of-the-art belong to the model-based class. *Matrix Factorization* (*MF*) [31,32] is the first implemented and has shown very promising performance in large-scale recommendation tasks. The basic idea behind the model is to factorize user preferences and item characteristics into a common latent space. Subsequently, it generates recommendations based on the dot product of user and item factors. Another widely used model is *SLIM* [33]. The algorithm is based on an efficient elastic-net sparse linear model that generates high-quality top-*k* recommendations. The approach demonstrated better performance than MF, generating faster recommendations. Finally, a very promising algorithm is *Mult-VAE* [34]. This approach exploits multinomial function-based variational autoencoders for collaborative filtering. The algorithm outperformed several baselines in real-world datasets.

2.1.1. Diversity in recommender systems

The literature on RSs has traditionally focused on the design of systems to optimize the accuracy of the recommendation [35]. However, accuracy is not the only quality indicator of algorithm predictions [36]. Beyond accuracy, other studies in the field of recommender systems have proposed evaluating diversity factors [7]. The right amount of diversity in user recommendations could address the well-known over-fitting problem [37,38]. Furthermore, more diverse recommendations are associated with higher levels of user satisfaction [39,40].

Diversity in the RS literature has been defined in various ways. Depending on the problem considered, it may be convenient to exploit one definition of diversity instead of another. The main definition of diversity [41] is based on the concept of dissimilarity between pairs of items in the result set. According to this definition, a rank that minimizes the similarity of items recommended to a user achieves a higher level of diversity. Alternative definitions [42,43], integrate diversity into the well-known *Normalized Discounted Cumulative Gain* (*NDCG*[1]) evaluation framework, or formalize it [44] as a combination of item topics (e.g., film genres), coverage, and non-redundancy factors.

From the diversity definitions mentioned above, in the literature [7], various models have been proposed. Some research has proposed methodologies based on item features. For example, a well-known study [17] proposed a post-processing approach to diversify recommendation topics. Another research [45] proposed a method that considers the impact of each individual item on total diversity. Another paper [46] instead proposed to avoid showing many documents of the same category by exploiting an algorithm based on matroid constraints. A specialized study in the music domain [47] proposed an approach to characterize the diversity of musical user interests by extracting features from audio files, while a study focused on the web [48] instead studied an algorithm that relies on the correlation of web content to provide the user with different search results while browsing. Finally, another article [49] studied how a recommender changes over time and what the impact of this change on diversity is.

Other works proposed instead methodologies based on user interaction data. For example, a research [50] proposed to exploit re-ranking methodologies to increase diversity, while another well-known study [51] proposed to reduce popularity bias, thus promoting long-tail items through evolutionary algorithms. Another paper [52] proposed to combine user-based and item-based techniques, while another research [53] proposed an approach using Hamming Distance to increase item diversity by exploiting collaborative data only. A specialized study on long-tail items [54], proposed an approach to improve diversity considering recommendations as resources to be allocated to items. Another work [55] proposed a probabilistic modeling approach to select subsets of items based on determinantal point processes to jointly encode the diversity and relevance of the item set. Instead, another study [56] proposed a methodology that relies on user ratings to create higher diversity recommendations through items' clusters, while another research [57] proposed a graph theory-based algorithm that exploits entropy concepts. Finally, recent trends are focusing on: balancing accuracy and diversity, while also including personalized explanations extracted from knowledge graphs [58]; identifying the correct level of diversity for each individual user [59]; increasing the efficiency of algorithms without compromising the quality of recommendations [60]; providing diverse recommendations without the use of tuning parameters [61].

Overall, the diversity literature aims to solve a completely different problem from the one presented in this paper in Section 3. The available strategies are quite generic. These do not consider that there are items that might influence the behavior of certain users who are more sensitive to certain topics than others. Therefore, although there are methodological similarities with the approaches we propose in Section 4, as we will see in the experimental Section 5.4.4, the techniques from the diversity literature cannot be used to effectively address our problem.

2.1.2. Addressing bias in recommender systems

Addressing the problem of bias in artificial intelligence systems is very important nowadays. These systems are used in many areas of our lives when we need to make critical decisions [8] (e.g., bank loans, legal processes, job selection). Therefore, it is essential that these decisions do not reflect discriminatory behavior that could be harmful to people. One of the most notorious biases that has traditionally been studied in the literature of recommender systems is demographic bias [62]. This particular type of bias occurs when the recommender system discriminates against users from a particular group (e.g., women/men, young/elderly). To mitigate the effect of bias, machine learning systems introduce algorithmic fairness [63]. The objective of fairness is to eliminate discrimination in model predictions.

There are multiple definitions of fairness in the literature [8]. Depending on the case considered, one definition might be more appropriate than another [64,65]. One of the most known definitions is *Demographic Parity* [66]. Consider a generic predictor that is required to assign a class to an individual. A predictor is said to achieve demographic parity if a specific outcome is equally likely to be assigned to individuals from different groups.

Some studies have focused on the introduction of algorithmic fairness in recommender systems [9,67,68]. Three main scopes have been identified depending on whether the RS should not discriminate: users/consumers (*C-fairness*), items/providers (*P-fairness*) or both (*CP-fairness*) [69,70]. Techniques are distinguished into pre-, in-, or post-processing according to when fairness is introduced into the learning process [71,72]. For example, some studies have proposed in-processing methodologies to introduce algorithmic fairness in matrix factorization [73] or in SLIM [69] by changing the objective function of the algorithms. Instead, another study [10] proposed a re-ranking algorithm based on the KL-divergence distance function to calibrate item topics in recommendations based on ground truth user preferences. In particular, the post-processing methodologies have been

[1] An introduction to NDCG can be found in Section 5.3.

Table 1
Adjectives describing the two polarities (high level vs. low level) of each Big Five personality trait [92].

Personality trait	High level	Low level
Openness (O)	wide interests, imaginative, intelligent, original, insightful, curious, sophisticated	commonplace, narrow interests, simple, shallow, unintelligent
Conscientiousness (C)	organized, thorough, planful, efficient, responsible, reliable, dependable	careless, disorderly, frivolous, irresponsible, slipshot, undependable, forgetful
Extraversion (E)	talkative, assertive, active, energetic, outgoing, outspoken, dominant	quiet, reserved, shy, silent, withdrawn, retiring
Agreeableness (A)	sympathetic, kind, appreciative, affectionate, soft-hearted, warm, generous	fault-finding, cold, unfriendly, quarrelsome, hard-hearted, unkind, cruel
Neuroticism (N)	tense, anxious, nervous, moody, worrying, touchy, fearful	stable, calm, contented, unemotional

applied to various recommenders by re-ranking the predicted scores [74], introducing fairness constraints [11] or considering temporal aspects to amortize fairness on series of multiple rankings [75,76]. Other studies [77–80] have instead proposed methodologies for improving certain fairness parameters in recommendations addressed to user groups [81] (i.e., sets of subjects who collectively receive a recommendation), such as friends who are planning a vacation [77]. Recent trends are investigating instead how to prevent the target user from being affected by sensitive features of neighboring users by using graph neural networks [82] or how to integrate fairness into medical applications [83].

As with diversity, the fairness literature aims to solve a problem that, although related, is completely different from the one proposed in this article in Section 3. The available techniques are quite generic because they do not consider how the recommendations of certain types of items may impact the behavior of certain kinds of user. Therefore, it is not possible to use them as is. However, as we will see in Section 4 where we describe algorithmic approaches, it was possible to redesign some of them [10,11] to address the problem of this paper.

2.1.3. Personality-aware recommender systems

One of the main problems with recommender systems is the cold-start [2,84,85], i.e., recommenders cannot draw inferences about users or items about which they have not yet collected sufficient information. The most widely adopted strategy to deal with cold-start is to introduce contextual features related to users or items [86,87]. In fact, since users and items tend to exhibit similar behaviors based on these characteristics, the latter can be exploited in the learning phase as additional data to guide inference in the space of the most likely solutions [88,89]. In the field of contextual feature research, a particular class of recommender systems known as *Personality-Aware Recommender Systems* (*PARS*) [90] has been introduced to address the cold-start [91]. These algorithms are based on particular types of contextual features known as personality traits,[2] which have been shown to be particularly predictive of user behavior.

Because of this special relationship that links personality traits with human behavior, several studies have been proposed in the PARS literature [91]. One of the first studies [93] showed that the use of similarity measures based on personality information was statistically equivalent to the use of ratings. Similarly, another study [94] demonstrated that combining rating and personality information yields better results than using ratings alone. These considerations have also been extended to matrix factorization approaches [95]. Another traditional recommendation issue in which PARS has proven useful is the generation of diverse recommendations [91]. A study [96] investigated the relationship between personality traits and diversity, showing that high Openness to experience and low Conscientiousness correlate with greater diversity in individual preferences. Based on these findings, a study [97] proposed using personality as a moderating factor to adjust the degree of diversity in recommendations, obtaining promising results. These considerations were further extended [38], showing that users with low Openness to experience tend to prefer thematic diversity to categorical variation. Personality traits have also been used in application fields to recommend computer games to players [98]. Recent trends have proposed exploiting personality traits to make cross-domain recommendations [99] since this information can be used to make transfer learning from one domain to another.

In our experiments we used Big Five personality data to define subsets of sensitive users (Section 5). However, as can easily be inferred from the context, while partially related to the PARS literature, the objectives of our study are completely different (Section 3). In fact, we do not exploit personality traits to achieve better accuracy in recommendations. Instead, we propose to use this information for a full case analysis involving subsets of potentially depressive and aggressive sensitive users identified according to the main correlations in the literature of depressive disorders and violent behavior with personality traits.

2.2. Big five personality traits and social issues

The *Big Five* theory is a taxonomy of personality traits originally hypothesized by Tupes and Christal [14] in the 1960s and subsequently developed in the 1980s and 1990s by different authors [92,100,101]. It defines five basic dimensions through which personality can be described. Personality is defined as a set of cognitive and behavioral patterns that account for individual differences. The *Big Five* theory is based on the hypothesis that the relevant individual differences are expressed in language. Therefore, researchers have identified a set of adjectives that can be grouped into five clusters that are capable of describing permanent traits of human behavior: Openness to experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism (OCEAN model). Each trait is a continuous dimension between two polarities (high level vs. low level) on which the person is located (Table 1). A personality trait leads to a specific behavioral response that is repeated with a certain constancy of time, regardless of the stimulus that causes it.

Self-reported questionnaires are currently the gold-standard for assessing personality in psychology. Different questionnaires have been proposed in literature for the measurement of the five main personality traits and their sub-dimensions according with the Big Five theory. Two of the most popular are the NEO Personality Inventory (NEO-PI) [100] and the International Personality Item Pool (IPIP) [102].

In addition to questionnaires, more recently other methods have been proposed to determine personality, such as using machine learning models to predict personality traits from various sources of information [103]. Pioneristic studies investigated the possibility of extracting personality information, according to the

[2] We will introduce the psychological theories of personality traits in Section 2.2.

Big Five personality model, from user activity on social networks [104–110]. These studies revealed that personality information is related to different aspects of social networking, such as the general activity defined as online time, the number of friends, posts, and likes [110]. Furthermore, personality appears to be related to the semantic context of texts and images (e.g., topics that subjects like and post) and the emotional connotation of multimedia content [109]. The social network that has been used primarily for research purposes is Facebook. Data are incredibly interesting: Garcia and Sikström [107] demonstrated that the semantic content of Facebook updates can accurately predict Psychopathy, [107] as people with psychopathic personality have a general tendency to post content with a negative valence. Sumner et al. [111] proposed a method based on the bag-of-words approach to predict the dark triad (i.e., the narcissism, Machiavellianism and psychopathy traits) from text data [111]. Some authors applied Artificial Intelligence methods to Facebook likes and successfully predicted personality [112], while others analyzed the link between the level of participation of Facebook users in photos and activities related to photos and their Big Five personality traits [113]. For example, low levels of Consciousness have been found to be correlated with an intensive and addictive use of social networks and a high number of pictures uploaded on Facebook; people with low levels of Agreeableness have fewer friends and are generally less tagged in photos [106]. Recent trends in personality detection methods focus instead on: providing interpretable approaches to identify personality traits from language by extracting semantic features from microblogs [114]; identifying personality traits in images through fuzzy and genetic algorithms [115]; utilizing multi-modal methods based on different types of data to increase the accuracy of predictions [116]; exploiting personality characteristics to improve the performance of user sentiment [117] or interest [118] mining methods in social networks.

2.2.1. Personality traits, depressive disorders and aggressive behavior

One of the most important social problems today involves mental disorders. A significant part of the population suffers from clinical conditions such as anxiety, depression, and substance abuse. In particular, depression has a high prevalence, affecting around 7% of the population [119] with major socio-economic impacts (e.g., increased mortality, direct costs of medications and hospitalization, generation of indirect costs for absence from work, turnover and disability compensation) [120]. Another relevant social problem concerns aggressive behaviors, such as bullying, racial violence, physical and verbal abuse, minor and major crimes, and anti-social behaviors. Although in recent years there has been a reduction in major crimes in many countries,[3] aggressive behaviors persist and are also evolving in the world of social networks (e.g., cyberbullying, hate speech) [121]. Different studies revealed that being the victim, as well as the perpetrator, of cyber-aggressions is related to lower well-being and mental health [122,123]. In the following, a brief discussion on the relationships between personality traits and these social issues is reported.

2.2.1.1. Personality traits and depressive disorders. The correlation between personality traits and mental disorders is known in the literature. A meta-analysis [124] examined the relationship between the Big Five traits and personality disorders defined in *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*) [119]. The study reports that each personality disorder is associated with a particular pattern of personality traits. For example, personality disorders characterized by emotional distress (e.g., paranoid, schizotypal, borderline, avoidant, and dependent disorders) show a positive correlation with Neuroticism; histrionic and narcissistic disorders show a positive correlation with Extraversion while schizoid, schizotypal, and avoidant disorders show a negative correlation with this trait. Disorders characterized by difficulties in relationships (e.g., paranoid, schizotypal, antisocial, borderline, and narcissistic) show negative correlations with Agreeableness. Furthermore, obsessive-compulsive disorder seems to be positively correlated with Conscientiousness, while antisocial and borderline disorders show a negative correlation with this trait. Other studies have investigated specific mental disorders in more depth. A meta-analysis of 175 studies published from 1980 to 2007 [15] found that depressive, anxiety, and substance abuse disorders in adults are predominantly correlated with traits of high Neuroticism and low Conscientiousness. People affected by dysthymic disorder and social phobia show low levels of Extraversion. Similar results on depression were also reported by another study [125]. A meta-analysis of 10 cohort studies suggests that depressive symptoms and personality traits are prospectively related. Personality traits are associated with the development of depressive symptoms, while depressive symptoms, are associated with temporary or persistent personality changes.

2.2.1.2. Personality traits and aggressive behavior. There are strong scientific evidence supporting that violent acts are highly correlated with specific personality profiles [126]. The personalities that are more strongly associated with a higher risk of committing criminal behavior are antisocial personality disorder and psychopathy. The literature also demonstrated the link between specific personality profiles and the risk of committing mild violence, such as bullying. For example, a meta-analysis revealed that studies that assessed personality through the Big Five personality model consistently reported an association between lower levels of Agreeableness and Conscientiousness and higher levels of Neuroticism and Extraversion with both bullying and victimization. On the contrary, cognitive and affective empathy was negatively associated only with bullying behavior [127]. Consistently, other studies [128,129] found that Agreeableness is negatively correlated with delinquency and aggressive behavior. Sharpe and Desai [130] found also that Neuroticism tends to be positively correlated with aggressive behavior while Conscientiousness tends to be negatively correlated. Another research [16] distinguished between physical aggression and violent behavior, showing that the former is directly and indirectly related to Openness, Agreeableness, and Neuroticism, while the latter is indirectly related to Openness, and Agreeableness. Recent studies have also confirmed the relationship between the Big Five traits and other forms of aggression, such as relational aggression [131] and sexual aggression [132].

3. Problem statement

In real-world circumstances, recommender systems tend to recommend items that belong to topics of interest to the user to maximize performance [62]. However, this reduces user exposure to a narrow subset of content leading to the well-known echo chamber phenomenon [4]. Echo chambers are environments in which users reinforce their opinions on certain topics due to repeated exposure to content of similar positions. Consequently, according to the theory of polarization [5], this will lead users belonging to the same group to increasingly reinforce their beliefs toward extreme positions without valuing different opinions. According to various studies, these extreme positions can trigger

[3] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Crime_statistics

a number of different issues, including an increased risk of local or international violent conflict [6]. It is important that platforms, especially social ones, address this problem and promote depolarization interventions with the aim of transforming conflicts into more constructive forms [133].

In this paper, we consider the consequences of polarization on the behavior of different users. Among all users, there are some who are more sensitive than others to certain topics. We define *sensitive*, as those users whose behavior can be influenced by certain types of items. Some examples of sensitive users may include depressed or aggressive users. The behavior of these users can be influenced by some particular content. For these properties, we define *influential*, as those items that could influence the behavior of sensitive users. In turn, influential items can be further divided into controversial or favorable items, depending on the type of behavioral influence. We define *controversial*, as those items that could have a dangerous influence on the behavior of sensitive users. Controversial items may include items associated with violence, alcohol and weapons. Over-recommending controversial items to aggressive users, could result in a potential spread of verbal and physical aggression, self-injury acts, depressive symptoms, fears, anxiety, and delinquency. On the other hand, we define *favorable*, as those items that might have a positive influence on sensitive users behavior. Favorable items may include items associated with sports, support groups and hobbies. Recommending enough favorable items could positively affect depressed users by supporting people in difficulty, promoting emotional balance, or encouraging healthy lifestyles.

It is indeed fundamental to correctly balance controversial and favorable items in sensitive user recommendations. Some strategies have been introduced to degenerate the feedback loop of recommendations by promoting the diversity of items [134]. However, these strategies are quite generic, do not consider user sensitivity, and could significantly decrease system performance. Therefore, the aim of this paper is to address the problem of maximizing the performance of a recommender system while diversifying the items suggested for sensitive users. Recommendations that meet the latter criterion could avoid the potential negative consequences that come from item distributions too skewed toward controversial item sets. Moreover, these could simultaneously carry positive consequences by promoting distributions that are more skewed toward favorable item sets.

In the following Section, we formalize the problem and we define a probability criterion that a RS should satisfy to preserve influential items' diversity in order to positively affect sensitive users' behavior.

3.1. Recommending influential items to sensitive users

Consider the task of recommending top-k items to a user. Let $u \in \mathcal{U}$ be a user belonging to a set $\mathcal{U} = \{u_1, \ldots, u_m\}$ of users and $s_u \in \{0, 1\}$ be a variable indicating whether user u belongs to a set $\mathcal{S} \subseteq \mathcal{U}$ of sensitive users. Let $i \in \mathcal{I}$ be an item that belongs to a set $\mathcal{I} = \{i_1, \ldots, i_n\}$ of items and $l_i \in \mathcal{Z} = \{0, 1\}$ be a binary variable defined over $\mathcal{Z}$ categories indicating whether the item i belongs to a set $\mathcal{L} \subseteq \mathcal{I}$ of influential items. We define the set of influential items as a set of items that can potentially affect the behavior of sensitive users. These include a combination of controversial and favorable items depending on the type of influence on sensitive user behavior. Let $f_i \in \{0, 1\}$ and $c_i \in \{0, 1\}$ be two variables indicating whether the item i belongs to a set $\mathcal{F} \subseteq \mathcal{L}$ of favorable items or $\mathcal{C} \subseteq \mathcal{L}$ of controversial items, respectively. For the case studies analyzed in our paper and to simplify the problem, the influential item set will contain only favorable ($\mathcal{F} = \mathcal{L} \wedge \mathcal{C} = \emptyset$) or controversial ($\mathcal{C} = \mathcal{L} \wedge \mathcal{F} = \emptyset$) items. Then, consider a recommender system that learns a function $\mathbf{R} \rightarrow \hat{\mathbf{R}}$ to predict the matrix of the predicted scores $\hat{\mathbf{R}} \in \mathbb{R}^{m \times n}$ from the user–item interaction matrix $\mathbf{R} \in \mathbb{R}^{m \times n}$. Let $r_{u,i}$ and $\hat{r}_{u,i}$ be the rating and the predicted score of the user u for the item i, respectively. Finally, let $\mathcal{Y}_{u,k}$ be the set of k recommended items to the user and $y_{u,i} \in \{0, 1\}$ be a variable indicating whether the item i belongs to the set $\mathcal{Y}_{u,k}$. For sake of notation, in the following we omit the dependency from the user u if clear from the context.

According to our objectives, a recommender satisfies a first diversity criterion with respect to the sensitive user group $\mathcal{S}$ and the influential item set $\mathcal{L}$ if:

$$\mathbb{P}(y_i = 1 | s_u = 1, l_i = 1) = \mathbb{P}(y_i = 1 | s_u = 0, l_i = 1) \tag{1}$$

The criterion requires an equality between the probability that an influential item is selected in the top-k recommendations for a sensitive user and the probability that the same item is selected in the recommendations for a non-sensitive user.

To give an example of how this affects a real-world scenario, let us examine two different cases. First, consider a set of potentially aggressive sensitive users and a set of influential items composed of social pages about controversial topics (e.g. weapons, alcohol). Then, assume to have a recommender system that tends to over-recommend these types of item to sensitive users compared to the non-sensitive group. According to echo chamber theory, in these circumstances, the system could influence the users, potentially making them even more aggressive. On the contrary, a recommendation system that meets the diversity criterion in Eq. (1) would balance the distributions of controversial items of potentially aggressive users while reducing the echo chamber effect. Second, consider a set of potentially depressed sensitive users and assume to have a recommender system that tends to under-recommend sets of favorable social pages (e.g., sports, support groups) to this group of users. In this case, the system under normal circumstances could keep users in their depressive state without providing any kind of help. In contrast, if it meets the diversity criterion, it could positively impact these people's mental state. Indeed, the exposition to positive stimuli (e.g., sports, social activities), could affect their emotional state and encourage them to engage in activities (e.g., practicing sports, seeking social or clinical support) that promote their well-being, leading them a step forward to get out of their depression condition. Note that while the diversity criterion in Eq. (1) may be useful in some cases, it may not always be sufficient. Promoting an imbalanced item distribution to highly sensitive users could bring greater benefits on their behavior. For highly depressed users, it might be desirable to promote more favorable items compared to non-sensitive users. Conversely, for highly aggressive users, it might be desirable to recommend even less controversial items. Therefore, in these cases, a more general formulation of the previous criterion might be useful:

$$\begin{cases} \mathbb{P}(y_i = 1 | s_u = 1, f_i = 1) \geq \alpha \\ \mathbb{P}(y_i = 1 | s_u = 1, c_i = 1) \leq \beta \end{cases} \tag{2}$$

with $\alpha > \mathbb{P}(y_i = 1 | s_u = 0, f_i = 1)$ as a lower bound on the probability of recommending favorable items to sensitive users and $\beta < \mathbb{P}(y_i = 1 | s_u = 0, c_i = 1)$ as an upper bound on the probability of recommending controversial items. The recommender system that meets the diversity criterion in Eq. (2) is able to recommend even fewer controversial items and even more favorable items to sensitive users compared to the non-sensitive group. Note that α and β are defined in the range $[0, 1]$. Thus, by setting $\alpha = 0$ or $\beta = 1$ it is possible to enforce just one of the two constraints defined in the above equation.

The algorithmic implementations we propose in the next section will rely on the criteria in Eqs. (1) and (2).

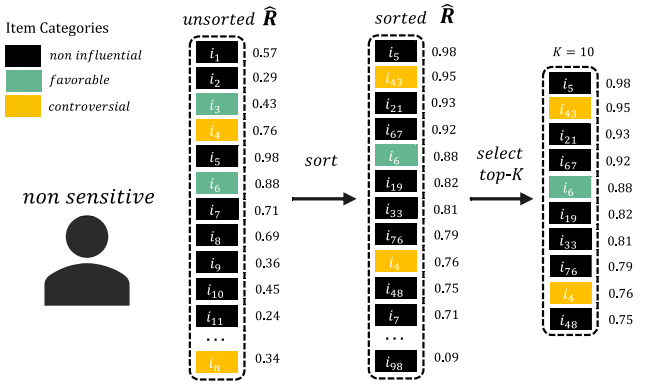


Fig. 1. Behavior of Calibration and Re-ranking for non-sensitive users: the algorithms sort the predicted scores $\hat{\mathbf{R}}$ in descending order and return the top-k items.

4. Algorithmic approach

In the following, we introduce the techniques designed to solve the problem introduced in Section 3. We took inspiration from some existing methods used in the field of fairness and redesigned them to suit our context.

4.1. Calibrating influential item distribution of sensitive users

We developed a first possible solution by modifying a well-known calibration approach [10]. The original algorithm was designed to solve a class imbalance problem to reflect the full spectrum of ground truth interests of the users in the recommendations. We redesigned the algorithm to calibrate the recommendations for each sensitive user based on a target distribution of items from all non-sensitive users. A brief description of the proposed approach follows based on the notation introduced in Section 3.1.

Given the categorical probability distribution $p(\mathcal{Z}|i)$ of categories $\mathcal{Z}$ for each item i, we define the distribution $p(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S})$ over categories $\mathcal{Z}$ of the set of items recommended over all non-sensitive users as:

$$p(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S}) = \frac{\sum_{u \in \mathcal{U} \setminus \mathcal{S}} \sum_{i \in \mathcal{Y}_{u,k}} a_i \cdot p(\mathcal{Z}|i)}{\sum_{u \mathcal{U} \setminus \mathcal{S}} \sum_{i \in \mathcal{Y}_{u,k}} a_i} \tag{3}$$

where a_i is a weight associated with item i that can be used to weight the distribution. Some possible choices to define a_i can be the predicted score $\hat{r}_{u,i}$, the position of item i in the ranking $\mathcal{Y}_{u,k}$ or others. For our experiments, we weighed all items uniformly i.e., $a_i = 1$. Then, we define the probability distribution $q(\mathcal{Z}|u \in \mathcal{S})$ over categories $\mathcal{Z}$ of the set of items recommended to a single sensitive user as:

$$q(\mathcal{Z}|u \in \mathcal{S}) = \frac{\sum_{i \in \mathcal{Y}_{u,k}} a_i \cdot p(\mathcal{Z}|i)}{\sum_{i \in \mathcal{Y}_{u,k}} a_i} \tag{4}$$

For the sake of notation, in the following we omit the category and user dependence in the distributions p and q if clear from the context.

We can now define an utility function (inspired by the one proposed by Steck [10]), to find the optimal set $\mathcal{Y}^*_{u,k}$ of k items to recommend to the sensitive user ($s_u = 1$) as:

$$\underset{\mathcal{Y}_{u,k}}{\text{argmax}} \quad (1-\lambda) \sum_{i \in \mathcal{Y}_{u,k}} s_u \hat{r}_{u,i} - \lambda \cdot KL(p \parallel q) \tag{5}$$

Algorithm 1 Calibration

```
1:  Input:
2:  u: user identifier;
3:  S: set of sensitive users;
4:  r̂_u: scores predicted by the backbone model for user u;
5:  p(Z|U \ S): categorical distribution of recommended items
    over all non-sensitive users;
6:  γ: objective function regularization parameter;
7:  k: number of items to be recommended;
8:  Output:
9:  Y*_{u,k}: top-k items to recommend to user u;
10: Procedure:
11: if u ∉ S then
12:     return argsort(r̂_u, order = descending)[0 : k]
13: else
14:     Y*_{u,k} ← ∅
15:     while |Y*_{u,k}| < k do
16:         obj^best ← −inf
17:         j^best ← −1
18:         for j ∈ r̂_u.items do
19:             if j ∉ Y*_{u,k} then
20:                 Y'_u ← Y*_{u,k} ∪ j
21:                 KL_{Y'_u} ← KL(p || q_{Y'_u})
22:                 obj ← (1 − λ) Σ_{i∈Y'_u} s_u r̂_{u,i} − λ · KL_{Y'_u}
23:                 if obj > obj^best then
24:                     obj^best ← obj
25:                     j^best ← j
26:                 end if
27:             end if
28:         end for
29:         Y*_{u,k} ← Y*_{u,k} ∪ j^best
30:     end while
31:     return Y*_{u,k}
32: end if
```

where $\lambda \in [0, 1]$ is a regularization parameter. The algorithm optimizes the predicted interest while calibrating the distribution (4) of each individual sensitive user to make it as close as possible to the target distribution (3) defined over all non-sensitive users. The term $KL(p \parallel q)$ is the Kullback–Leibler divergence, defined to quantify the distance between the two distributions as:

$$KL(p \parallel \tilde{q}) = \sum_{\mathcal{Z}} p(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S}) \log \frac{p(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S})}{\tilde{q}(\mathcal{Z}|u \in \mathcal{S})} \tag{6}$$

with $\tilde{q}(\mathcal{Z}|u \in \mathcal{S}) = (1-\eta) \cdot q(\mathcal{Z}|u \in \mathcal{S}) + \eta \cdot p(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S})$ and $\eta = 0.01$ as an approximation of the distribution $q(\mathcal{Z}|u \in \mathcal{S})$ to avoid that the KL function diverges when $q(\mathcal{Z}|u \in \mathcal{S}) = 0$ and $p(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S}) > 0$.

The problem is solved with a greedy approach that operates as follows (Algorithm 1). For each user: if the user is non-sensitive, the algorithm sorts the predicted scores in descending order and returns the top-k items (Fig. 1); if the user is sensitive, the algorithm returns a set of k items by exploiting an iterative procedure, depicted in Fig. 2 and described in the following.

For any sensitive user, the algorithm starts from an empty set $\mathcal{Y}^*_{u,k} \leftarrow \emptyset$ (line 14) and iteratively adds items until a set of the required size is obtained (line 15). At each iteration, for each item $j : j \in \mathcal{I}, j \notin \mathcal{Y}^*_{u,k}$ the algorithm computes the KL divergence in Eq. (6) between the distribution $q_{\mathcal{Y}'_u}(\mathcal{Z}|u \in \mathcal{S})$ of the set $\mathcal{Y}'_u = \mathcal{Y}^*_{u,k} \cup j$ and the distribution $p(\mathcal{Z}|\mathcal{U} \setminus \mathcal{S})$ of the top-k items of all non-sensitive users (line 21). The algorithm then adds the item j^{best} to $\mathcal{Y}^*_{u,k}$ (line 29) that maximizes the objective function in Eq. (5) (line 22).

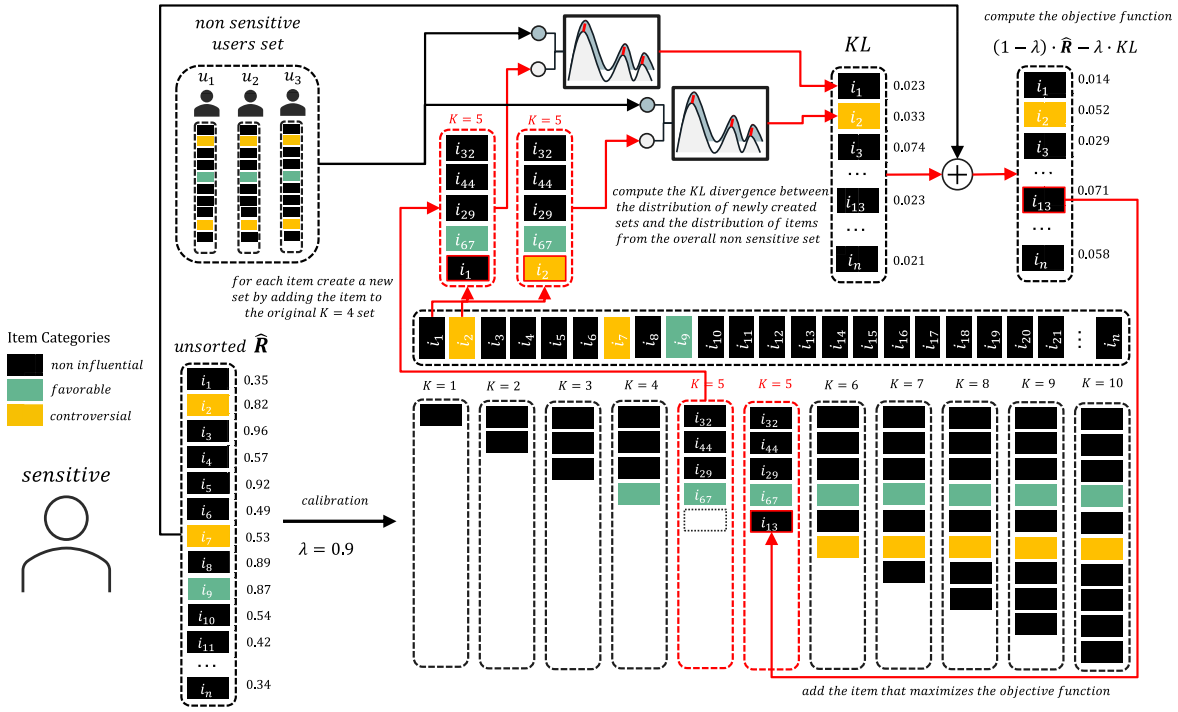


Fig. 2. Execution of Calibration (Algorithm 1) on a sensitive user: at each iteration (e.g., 5-th iteration), it is evaluated which item to be added based on the utility function in Eq. (5) that considers the predicted score $\hat{r}_{u,i}$ and the KL divergence in Eq. (6).

In Fig. 2 we provide an illustrative example of the execution of the Calibration algorithm on a sensitive user to obtain a list of 10 items. We focus on the 5-th iteration. For each item $j : j \in \mathcal{I}, j \notin \mathcal{Y}^*_{u,k}$ the algorithm defines the set $\mathcal{Y}'_u = \{i_{32}, i_{44}, i_{29}, i_{67}, j\}$. Thus, the algorithm first calculates the KL divergence between the set $\mathcal{Y}'_u$ and the top-k items of the set of non-sensitive users and then the objective function. Item i_{13} is added to $\mathcal{Y}^*_{u,k}$ because it is the one that maximizes the value of the objective function.

4.2. Recommending influential items to sensitive users under constraints

Another simple yet effective solution was designed taking inspiration from a well-known methodology in the literature on fairness [11,135]. The original algorithm aimed to eliminate discrimination in rankings by achieving demographic parity through fairness constraints. The algorithm we propose allows to determine the list of top-k items that maximize the expected value of sensitive users recommendations under constraints on the percentage of controversial and favorable items. Below is a brief description of the proposed methodology based on the notation defined in Section 3.1.

The list of top-k items $\mathcal{Y}_{u,k}$ recommended to the sensitive user ($s_u = 1$) satisfies the diversity constraint in Eq. (2) if:

$$\begin{cases} \frac{\sum_{i \in \mathcal{Y}_{u,k}} s_u f_i}{k} \geq \alpha \\ \frac{\sum_{i \in \mathcal{Y}_{u,k}} s_u c_i}{k} \leq \beta \end{cases} \tag{7}$$

where α and β in the range [0, 1] are, respectively, a lower and an upper bound on the percentage of favorable and controversial items allowed in the top-k list.

The optimization problem we propose aims to find the optimal set $\mathcal{Y}^*_{u,k}$ of k items that maximizes the predicted value for the sensitive user subject to the constraints in Eq. (7):

$$\underset{\mathcal{Y}_{u,k}}{\text{argmax}} \sum_{i \in \mathcal{Y}_{u,k}} s_u \hat{r}_{u,i} \tag{8}$$

Algorithm 2 Re-ranking

Input:
u: user identifier;
$\mathcal{S}$: set of sensitive users;
$\hat{\mathbf{r}}_u$: scores predicted by the backbone model for user u;
α: minimum percentage of favorable items allowed;
β: maximum percentage of controversial items allowed;
k: number of items to be recommended;
Output:
$\mathcal{Y}^*_{u,k}$: top-k items to recommend to user u;
Procedure:
if $u \notin \mathcal{S}$ **then**
  return argsort($\hat{\mathbf{r}}_u$, order = *descending*)[0 : k]
else
  $\mathcal{Y}^*_{u,k} \leftarrow \emptyset$
  $\hat{\mathbf{r}}^{sorted}_u \leftarrow$ sort($\hat{\mathbf{r}}_u$, order = *descending*)
  for $j \in \hat{\mathbf{r}}^{sorted}_u$.*items* **do**
    if $|\mathcal{Y}^*_{u,k}| < k$ **then**
      $con_\alpha \leftarrow ((f_j + \sum_{i \in \mathcal{Y}^*_{u,k}} s_u f_i)/k \geq \alpha)$
      $con_\beta \leftarrow ((c_j + \sum_{i \in \mathcal{Y}^*_{u,k}} s_u c_i)/k \leq \beta)$
      if $con_\alpha \wedge con_\beta$ **then**
        $\mathcal{Y}^*_{u,k} \leftarrow \mathcal{Y}^*_{u,k} \cup j$
      end if
    else
      return $\mathcal{Y}^*_{u,k}$
    end if
  end for
end if

The problem is solved with an efficient greedy algorithm that works as follows (Algorithm 2). As for the previous approach: if the user is non-sensitive, the algorithm returns the k items with

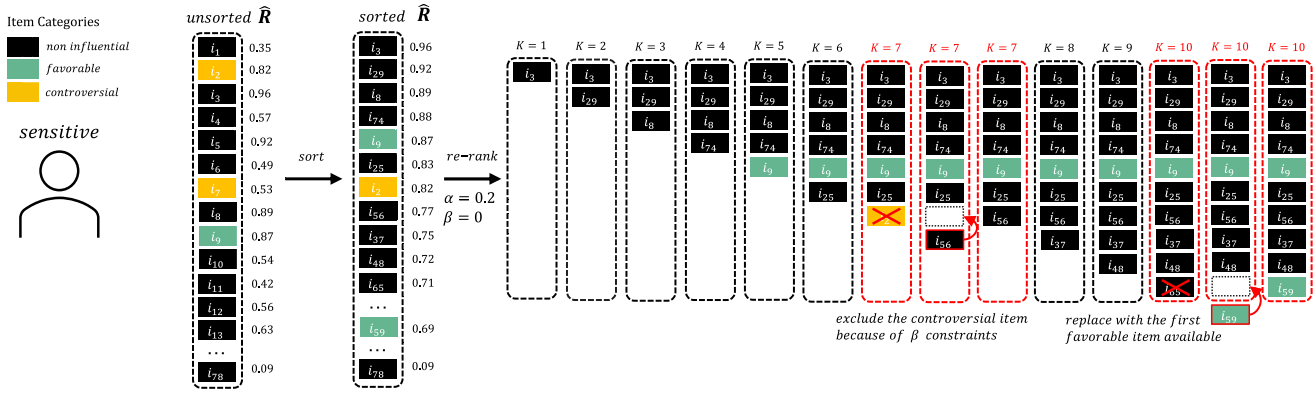


Fig. 3. Execution of Re-ranking (Algorithm 2) on a sensitive user: after sorting the items according to the predicted scores $\hat{\mathbf{R}}$ in descending order, at each iteration the algorithm adds the item that matches the constraints in Eq. (7) until a list of k items is obtained.

the highest predicted scores (Fig. 1); if the user is sensitive, the algorithm exploits an iterative procedure to determine the set of top-k items (Fig. 3).

For each sensitive user, the algorithm starts from an empty set $\mathcal{Y}^*_{u,k} \leftarrow \emptyset$ (line 14) and sorts the items according to the predicted scores $\hat{\mathbf{r}}_u$ in descending order (line 15). The algorithm then cycles through the ordered list $\hat{\mathbf{r}}_u^{sorted}$ (line 16). At each iteration the item j that matches the constraints in Eq. (7) (lines 18–19) is added to the set $\mathcal{Y}^*_{u,k}$ (line 21) until a list of the required size is obtained.

In Fig. 3 we provide an illustrative example of the execution of the Re-ranking algorithm on a sensitive user to obtain a list of 10 items. In the first 6 iterations, the algorithm adds the items in order of predicted scores because they all match the constraints in Eq. (7). In the 7-th iteration, the algorithm discards the controversial item i_2 because it does not match the constraint $\beta = 0$ and replaces it with the next item i_{56}. Finally, at the 10-th iteration, the algorithm discards the non-influential item i_{65} because it does not match the constraint $\alpha \geq 0.2$ and replaces it with the next available favorable item i_{59}.

4.3. Combining different approaches for recommending influential items to sensitive users

In the previous section, we presented two algorithms for recommending influential items for sensitive users. As it will be detailed in Section 5, it is difficult to decide a-priori which algorithm will perform best on a given problem setting. A possible approach to relieve a user from the need to select a single algorithm is to develop a technique that can automatically combine multiple strategies.

A solution for combining different approaches to recommend influential items to sensitive users together was designed by taking inspiration from the literatures of hybrid recommenders [12] and rank fusion [13]. The former is a branch of recommender systems research [1,2] that deals with combining the outputs of different recommenders to achieve greater performance. The latter, on the other hand, is a branch of information retrieval research [136] related to that of recommenders concerned with mixing the ranks generated by different IR systems. The algorithm we propose is inspired by the classes of algorithms implemented in these literatures to determine the best rank generated from a set of different rankers. Below is a brief description of the proposed methodology based on the notation defined in Section 3.1.

Let $\mathcal{G}$ be a set of rankers that can be applied to recommend $\mathcal{Y}'_{u,k}$ lists of items to sensitive users (e.g., exploiting Algorithm 1 or Algorithm 2). Let *AIR*[4] be a metric indicating the percentage of influential items in any $\mathcal{Y}_k$ list of k recommended items. The optimization problem we propose aims to find the optimal set $\mathcal{Y}^*_{u,k}$ of k items that minimize the deviation in *AIR* between the recommendations for the single sensitive user $u \in \mathcal{S}$ and those for all non-sensitive users belonging to the set $\mathcal{U} \setminus \mathcal{S}$:

$$\underset{\mathcal{Y}'_{u,k}}{\text{argmin}} \quad |AIR(\mathcal{Y}'_{u,k}) - AIR(\mathcal{Y}_{\mathcal{U}\setminus\mathcal{S},k})| \tag{9}$$

The problem can be solved by the following iterative greedy algorithm (Algorithm 3). First, the recommendations for non-sensitive users $\mathcal{Y}^*_{\mathcal{U}\setminus\mathcal{S},k}$ are determined by sorting the predicted scores $\hat{\mathbf{R}}_{\mathcal{U}\setminus\mathcal{S}}$ in descending order and selecting the top-k items for each user (line 10).

Algorithm 3 Combination

1: **Input:**
2: $\mathcal{U}$: set of all users;
3: $\mathcal{S}$: set of sensitive users;
4: $\hat{\mathbf{R}}$: scores predicted by the backbone model for all users;
5: $\mathcal{G}$: set of rankers to be combined;
6: k: number of items to be recommended;
7: **Output:**
8: $\mathcal{Y}^*_k$: top-k items to recommend to all users;
9: **Procedure:**
10: $\mathcal{Y}^*_{\mathcal{U}\setminus\mathcal{S},k} \leftarrow \text{argsort}(\hat{\mathbf{R}}_{\mathcal{U}\setminus\mathcal{S}}, \text{order} = \textit{descending})[0:k]$
11: $\mathcal{Y}^*_{\mathcal{S},k} \leftarrow \emptyset$
12: **for** $u \in \mathcal{S}$ **do**
13: $\quad |\Delta_{AIR}|^{best} \leftarrow +\text{inf}$
14: $\quad \mathcal{Y}^{best}_{u,k} \leftarrow \emptyset$
15: $\quad$ **for** ranker $\in \mathcal{G}$ **do**
16: $\quad\quad \mathcal{Y}'_{u,k} \leftarrow \text{ranker}(\hat{\mathbf{r}}_u)$
17: $\quad\quad |\Delta_{AIR}|' \leftarrow |AIR(\mathcal{Y}^*_{\mathcal{S},k} \cup \mathcal{Y}'_{u,k}) - AIR(\mathcal{Y}^*_{\mathcal{U}\setminus\mathcal{S},k})|$
18: $\quad\quad$ **if** $|\Delta_{AIR}|' < |\Delta_{AIR}|^{best}$ **then**
19: $\quad\quad\quad |\Delta_{AIR}|^{best} \leftarrow |\Delta_{AIR}|'$
20: $\quad\quad\quad \mathcal{Y}^{best}_{u,k} \leftarrow \mathcal{Y}'_{u,k}$
21: $\quad\quad$ **end if**
22: $\quad$ **end for**
23: $\quad \mathcal{Y}^*_{\mathcal{S},k} \leftarrow \mathcal{Y}^*_{\mathcal{S},k} \cup \mathcal{Y}^{best}_{u,k}$
24: **end for**
25: **return** $\mathcal{Y}^*_{\mathcal{U}\setminus\mathcal{S},k} \cup \mathcal{Y}^*_{\mathcal{S},k}$

[4] We will define this metric later in the paper in Section 5.3 devoted to experimental evaluation.

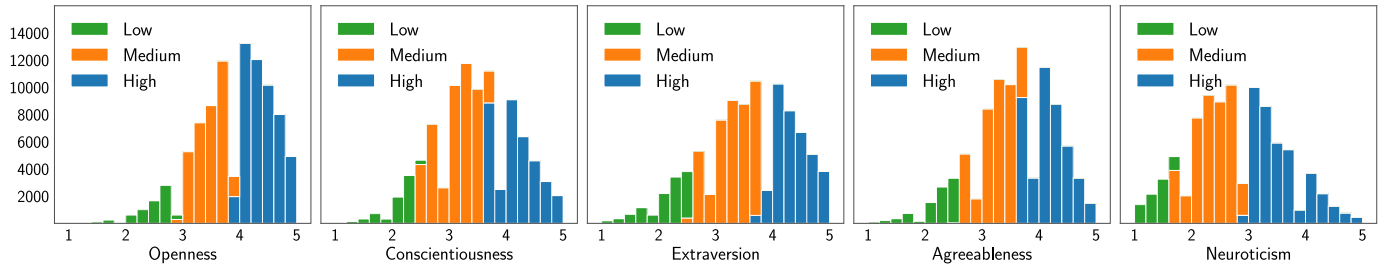


Fig. 4. The *myPersonality* dataset [104,137,138] distribution of user Big Five personality traits.

Then, starting with an empty matrix $\mathcal{Y}^*_{\mathcal{S},k} \leftarrow \emptyset$ the algorithm proceeds by iterating the following steps for each sensitive user (line 12). Different candidate lists $\mathcal{Y}'_{u,k}$ are generated (line 16) from a set $\mathcal{G}$ of rankers (line 15). For each of the candidate lists, the corresponding $|\Delta_{AIR}|'$ (line 17) is calculated based on: the *AIR* that would be obtained if the list $\mathcal{Y}'_{u,k}$ is added to the recommendations already selected for sensitive users at the current iteration; the *AIR* of recommendations for non-sensitive users. Then, the candidate list $\mathcal{Y}^{best}_{u,k}$ that minimizes the $|\Delta_{AIR}|'$ (line 20) is selected and added to $\mathcal{Y}^*_{\mathcal{S},k}$ (line 23). Finally, the algorithm returns $\mathcal{Y}^*_{\mathcal{U}\setminus\mathcal{S},k} \cup \mathcal{Y}^*_{\mathcal{S},k}$ joining the recommendations for sensitive and non-sensitive users (line 25).

Note that by design, since the algorithm is incremental, for the first few iterations the term $AIR(\mathcal{Y}^*_{\mathcal{S},k} \cup \mathcal{Y}'_{u,k})$ used to compute $|\Delta_{AIR}|'$ (line 17), tends to vary because it is influenced more by the rank of the individual user $\mathcal{Y}'_{u,k}$ than the rank of all sensitive users $\mathcal{Y}^*_{\mathcal{S},k}$. However, as the number of sensitive users processed increases, $\mathcal{Y}^*_{\mathcal{S},k}$ has more weight in the calculation and consequently $|\Delta_{AIR}|^{best}$ tends to stabilize and decrease.

5. Experiments

In this section, we present the results of applying the algorithmic approaches presented in Section 4 for the problem described in Section 3 on two real-world case studies. In Section 5.1 we present the dataset used. In Section 5.2 we introduce the backbone recommender systems exploited. In Section 5.3 we define the metrics used for performance evaluation. In Section 5.4 we present the experiments. Finally, in Section 6 we discuss the results and limitations of the study.

5.1. Dataset description

One of the most famous datasets that has been collected with the aim of studying the relationship between personality and activity in social networks is *myPersonality*.[5] [104,137,138] Many studies on user personality information in recommender systems are based on this dataset (Section 2.1.3). The dataset contains data from about 4.3 millions of Facebook users who contributed to psychological research between 2007 and 2012 by filling out a personality questionnaire through a social application. In our experiments, we used a subset of the original dataset.

The dataset contains personal information of anonymized users (e.g., gender, age, etc.), information about their activities on Facebook, personality traits, and other psychological information. Personality data were collected through a 20-item mini-IPIP questionnaire [139] that allows the determination of the Big Five personality traits (Section 2.2). According to the main standards, the personality traits of the users in the dataset are represented as a value defined in a range $[1, 5]$ for each trait. Information is also present in discrete form. Each variable can be associated with a value in the set $\{Low, Medium, High\}$ according to threshold criteria indicating the influence of each trait. In addition, the dataset contains information about which Facebook pages the user likes. These data were collected in the form of a topical decomposition of the users resulting from a 600-component Latent Dirichlet Allocation where each user was treated as a document containing the words from its own dictionary of likes. The Latent Dirichlet Allocation model (LDA) [140] is a well-known probabilistic topic modeling technique in natural language processing that allows to represent a document from a set of underlying topics. The model is based on a two-step Bayesian generative process where each document is considered as a set of words that combined together compose one or more subsets of latent topics, each of which is characterized by a particular word distribution. In the dataset, each LDA topic is represented by a set of 5 distinct pages and the user's preference for the topic is expressed through a value in the range [0, 1]. The same page can be found on multiple topics. Each user can thus be represented by a weighted combination of topics, the interpretation of which may indicate a particular taste in films and music groups, sexual and religious orientation, or a political view. The dataset contains a total of 4,282,857 users and 6,171,599 pages. For the experiments we present in the following sections, we considered a subset containing all the users of the dataset that have associated the Big Five personality traits and the information of the pages the user likes in the LDA format. Moreover, since the same pages could be found multiple times in different LDA topics, we decomposed the topics into individual pages through an averaging operation. As a result, we obtained a subset of 92,255 users and 1,836 pages.

The distribution of the Big Five personality traits of the users is reported in Fig. 4. As we can see, for each personality trait, there are three subsets indicating the influence of the trait (low vs. medium vs. high). The criteria defining the membership of the subset are variable according to the trait. For example, when comparing Openness and Neuroticism we observe that the range indicating a low trait influence is wider for Openness while the range indicating a high trait influence is wider for Neuroticism. Generally, it would also appear that for each trait, the subset indicating a low trait influence receives fewer users than the subset indicating a medium or high influence.

5.2. Backbone recommender systems

In this Section we introduce the two state-of-the-art top-k recommender systems used as backbones to generate recommendations considering the notation in Section 3.1.

5.2.1. Sparse linear method

SLIM [33] is an efficient sparse linear model able to compute top-k recommendations from the purchase/rating user profiles by solving a regularized optimization problem under non-negativity constraints.

[5] http://mypersonality.org/

Let $\mathbf{R} \in \{0,1\}^{m \times n}$ be the user–item sparse interaction matrix and $\mathbf{B} \in \mathbb{R}_{+}^{n \times n}$ the SLIM coefficients. The algorithm computes the recommendation score $\hat{r}_{u,i}$ for the user u and the un-rated item i as $\hat{r}_{u,i} = \mathbf{r}_u^T \mathbf{b}_i$ with $\mathbf{r}_u^T$ as the row vector of items interactions of user u and $\mathbf{b}_i$ as the column vector of the model coefficients for item i. The latters are learnt by solving the following optimization problem:

$$\begin{aligned} \underset{\mathbf{b}_i}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{r}_i - \mathbf{R}\mathbf{b}_i\|_2^2 + \frac{\gamma}{2}\|\mathbf{b}_i\|_2^2 + \delta\|\mathbf{b}_i\|_1 \\ \text{subject to} \quad & \mathbf{b}_i \geq 0, \quad b_{i,i} = 0 \end{aligned} \tag{10}$$

with γ and δ as regularization parameters to obtain the best trade-off between complexity and performance.

After the learning phase, the algorithm is able to recommend the top-k items for the user u by sorting the $\hat{\mathbf{r}}_u^T$ un-rated recommendation scores in descending order.

5.2.2. Variational autoencoders with multinomial likelihood

Mult-VAE [34] is a recent non-linear probabilistic variational autoencoder algorithm that exploits multinomial likelihood for collaborative filtering recommendations. Mult-VAE is a probabilistic latent-variable model based on the following generative process. The model samples a multidimensional latent representation $\mathbf{z}_u$ from a standard Gaussian prior for the user u. Then, $\mathbf{z}_u$ is transformed into a probability distribution $x(\mathbf{z}_u)$ over n items through a non-linear multilayer perceptron with parameters $\boldsymbol{\Theta}$ and softmax activation. The process assumes that the vector $\mathbf{r}_u$ of user interests is drawn from a multinomial distribution $Mult(t_u, x(\mathbf{z}_u))$ with t_u sum of the number of ratings for the user. To determine $\boldsymbol{\Theta}$, the model has to estimate the posterior distribution $g(\mathbf{z}_u|\mathbf{r}_u)$. However, $g(\mathbf{z}_u|\mathbf{r}_u)$ is intractable and therefore is approximated by a diagonal Gaussian distribution $h(\mathbf{z}_u) = \mathcal{N}(\boldsymbol{\mu}_u, diag(\boldsymbol{\sigma}_u^2))$ with $\{\boldsymbol{\mu}_u, \boldsymbol{\sigma}_u^2\}$ free variational parameters such that the Kullback–Leibler divergence between the two distributions $KL(h(\mathbf{z}_u) \parallel g(\mathbf{z}_u|\mathbf{r}_u))$ is minimized. However, the number of parameters $\{\boldsymbol{\mu}_u, \boldsymbol{\sigma}_u^2\}$ to optimize grows with the number of users and items in the dataset and can become a bottleneck in real-world applications. Thus, the variational autoencoder replaces the variational parameters of the function $h(\mathbf{z}_u)$ by turning it into a function $h_{\Phi}(\mathbf{z}_u|\mathbf{r}_u) = \mathcal{N}(\mu_{\Phi}(\mathbf{r}_u), diag(\sigma_{\Phi}^2(\mathbf{r}_u)))$ parameterized by Φ that, if optimized, approximates the intractable posterior $g(\mathbf{z}_u|\mathbf{r}_u)$. The model attempts to minimize the distance between the distributions by optimizing an evidence lower bound that is interpreted as composed of a first reconstruction error and a second regularization term:

$$\mathbb{E}_{h_{\Phi}(\mathbf{z}_u|\mathbf{r}_u)}[\log g_{\boldsymbol{\Theta}}(\mathbf{r}_u|\mathbf{z}_u)] - \psi \cdot KL(h_{\Phi}(\mathbf{z}_u|\mathbf{r}_u) \parallel g(\mathbf{z}_u)) \tag{11}$$

where $\psi \in [0,1]$ is a regularization variable.

After the learning phase, the algorithm is able to recommend the top-k items for user u by sorting in descending order the un-rated predicted scores.

5.3. Evaluation metrics

We evaluated the top-k recommendation algorithms using *Normalized Discounted Cumulative Gain* (*NDCG@k*) [141], *Average Influential Ratio* (*AIR@k*) and *Sensitive ImBalance* (*SIB@k*) The results of the experiments are reported by aggregating the metrics by user group and item set. Taking into account the notation introduced in Section 3.1, we define each metric below.

We used *NDCG@k* as a performance measure for recommendations. Let $rel_{u,j}$ be a relevance metric that indicates whether the item i recommended at position j is relevant or not for user u. The relevance value of item $i \in \mathcal{Y}_{u,k}$ for user u corresponds to the value of the ground truth rating $r_{u,i}$. Thus, $NDCG_u@k$ for user u can be calculated as:

$$NDCG_u@k = \frac{DCG_u@k}{IDCG_u@k} \tag{12}$$

where $DCG_u@k = \sum_{j=1}^{k} \frac{rel_{u,j}}{log_2(j+1)}$ is the discounted cumulative gain resulting from placing each item in a given position in the ranking $\mathcal{Y}_{u,k}$ and $IDCG_u@k$ is the ideal cumulative gain obtained by sorting all the items relevant to the user in descending order. Thus, *NDCG@k* is given by the average of $NDCG_u@k$ of users of the test set.

We exploited *AIR@k* as a measure of the number of influential items in user recommendations. The $AIR_u@k$ for the user u can be defined as:

$$AIR_u@k = \frac{1}{k} \sum_{i \in \mathcal{Y}_{u,k}} l_i \tag{13}$$

and the overall *AIR@k* is given by the average of $AIR_u@k$ on all users of the test set.

Finally, we used $SIB_i@k$ as an item-level measure that indicates how frequently certain items are recommended on average to sensitive users compared to non-sensitive ones. The metric was developed by repurposing a widely adopted metric in the field of fairness to suit our context, that is, *Non-Parity Unfairness* (*NP*) [73]. Let $\frac{1}{|S|}\sum_{u \in S} f(\hat{r}_{u,i})$ be the average predicted score of item i from the sensitive user group and $\frac{1}{|U \setminus S|}\sum_{u \in U \setminus S} f(\hat{r}_{u,i})$ the average predicted score from the non-sensitive one with $f(\hat{r}_{u,i}) = 1$ if $i \in \mathcal{Y}_{u,k}$ as a binarization function. Binarization has been used to normalize the predicted scores of different recommender systems. Thus, $SIB_i@k$ for item i can be calculated as:

$$SIB_i@k = \frac{1}{|S|}\sum_{u \in S} f(\hat{r}_{u,i}) - \frac{1}{|U \setminus S|}\sum_{u \in U \setminus S} f(\hat{r}_{u,i}) \tag{14}$$

We use this metric exclusively to provide an interpretation of the pages recommended the most and least frequently.

5.4. Experimental results

This Section is dedicated to the description of the experimental methodology and the obtained results.

- In the first experiment (Section 5.4.1) we first checked if the Big Five personality information was present in the LDA-format pages associated with the users and then we gave a preliminary interpretation by studying the most correlated pages with the various personality traits. Although the literature seemed to confirm the initial hypothesis [142], we still preferred to verify it. This was done because our dataset, unlike those used in previous studies, was based on LDA-format pages and had a lower number of available data points. In addition, the most correlated pages were used to select the influential item sets used in the next experiment.
- In the second experiment (Section 5.4.2) we studied the performance of recommender systems based on two different case studies: recommend favorable items to potentially depressed users and recommend controversial items to potentially aggressive users. The subsets of users were selected based on the Big Five personality traits most correlated with the depressive disorders and aggressive behaviors described in Section 2.2.1. Item sets have been selected manually from the most correlated pages identified in the previous experiment according to the topic of the page (e.g. sports, hobbies, weapons, alcohol, and others).
- In the third experiment (Section 5.4.3), given the tendency of SLIM and Mult-VAE to over-recommend controversial items and under-recommend favorable items, we exploited

Table 2
The RMSE and the corresponding number of non-zero W_i coefficients of Lasso evaluated on the test set divided by Big Five personality trait.

Metric	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
RMSE	0.6096	0.6979	0.7703	0.6836	0.7828
$\|\mathbf{w}\|_0$	1,100	954	964	966	952

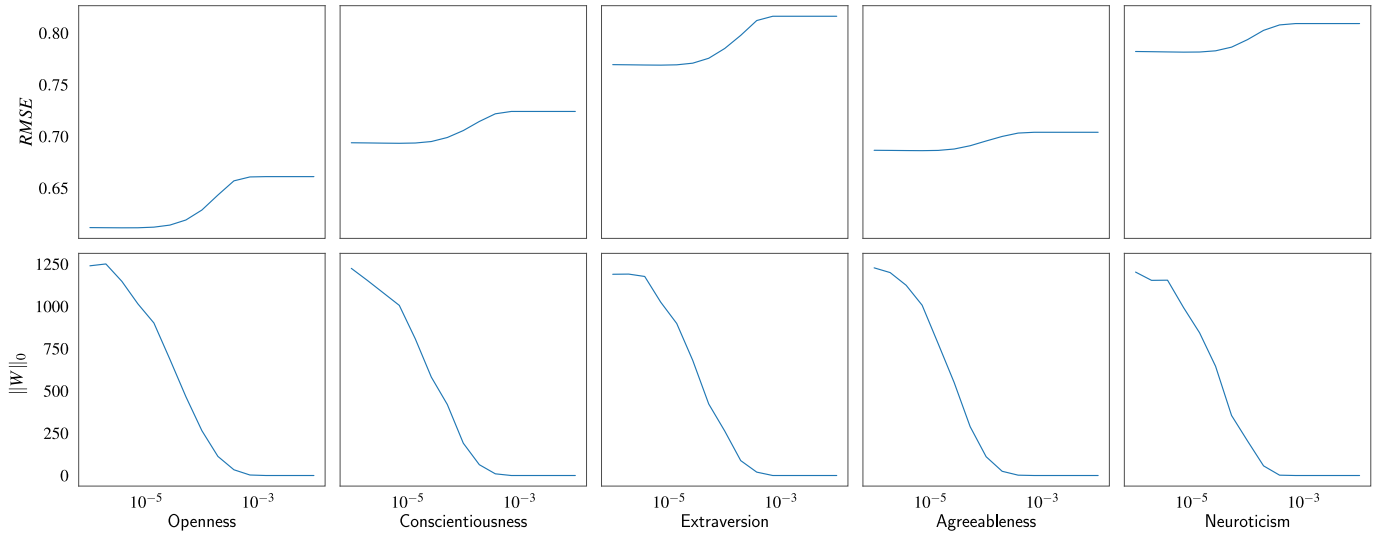


Fig. 5. The average *RMSE* and the corresponding number of non-zero W_i coefficients of Lasso along the cross-validation search space divided by Big Five personality trait.

the algorithms proposed in Sections 4.1 and 4.2 to improve the recommendations for sensitive users and we measured the variations in performance.
- In the fourth experiment (Section 5.4.4), we compare the results obtained from the calibration and re-ranking approaches proposed in Section 4 with the algorithms from the diversity literature (Section 2.1.1).
- In the fifth experiment (Section 5.4.5), we explore the results obtained from the proposed *combination* approach presented in Section 4.3, that combines the outputs of different rankers.

The experiments reported in this section were deployed on Google Cloud Platform instances running Debian 10 OS equipped with 8 vCPUs and 64 GB RAM optimized with the TensorFlow Enterprise 2.3 environment and accelerated with the Intel® MKL-DNN/MKL library. The experiments have been coded in Python 3.9.7 and are based on the RecTorch 0.9.0 [143] and the Scikit-learn 1.0.2 [144] libraries. An NVIDIA Tesla T4 GPU was used for experiments with variational autoencoders [34].

5.4.1. Preliminary analysis of dataset information content

In this Section, we present the results from an experiment designed to evaluate if LDA users' preferences are predictors of Big Five personality traits. In addition, we give a preliminary interpretation of the pages that are most and least correlated with each personality trait.

The experiment was carried out using the following methodology. We randomly split the dataset into training and test sets (80% / 20%). For each personality trait, we trained *Lasso* [145] to predict the personality score from users' LDA preferences. We performed a 5-fold grid search cross-validation on the training set to find the best hyperparameters of the model by optimizing the *Root Mean Square Error* (*RMSE*) [145]. The search space for Lasso was defined by exploring its regularization coefficient in the $[-6, -2]$ logarithmic range. Subsequently, Lasso was re-trained for each personality trait in the full training set with the hyperparameters found in the previous step and *RMSE* was evaluated in the test set. We then exploited the $\mathbf{w}$ coefficients of the fine-tuned models to give a qualitative evaluation of the most and least predictive pages for each trait. In the following, we discuss the results obtained.

In Table 2 we report the results for the fine-tuned models evaluated in the test set. Results indicate that it is particularly difficult to predict Neuroticism and Extraversion. Openness is associated with higher performance, while Conscientiousness and Agreeableness show intermediate results. In Fig. 5 we show the average *RMSE* along the cross-validation search space and the corresponding number of non-zero $\mathbf{w}$ coefficients ($\|\mathbf{w}\|_0$) for the different Big Five personality traits. As expected for all the personality traits the average *RMSE* is increasing as the number of non-zero coefficients decreases until a saturation point is reached. The results are in line with those presented by Liu et al. [142] where, differently from our work, LDA topics are defined on Facebook user status updates.

In Table 3 we show the pages most positively and negatively associated with each personality trait according to the $\mathbf{w}$ coefficients of the models fine-tuned with the experimental procedure described above. Analyzing the results, we can see that interests in acting, drawing, philosophy and poetry are positively associated with Openness and can be interpreted as indicators of creativity and curiosity. TV shows such as *Survivor* and *Cake Boss*, on the other hand, are negatively associated and can be interpreted as indicators of commonplaces and narrow interests. Regarding conscientiousness, some positively associated interests are *Cappex.com* and *QuikTrip* or sports such as running that may be indicators of organization, reliability, and self-control. Interests in marijuana, manga, and video games such as *The Sims 3* are negatively related and can be interpreted as a symptom of irresponsibility. Some of the positive associations with Extraversion are dancing and acting, brands like *Victoria's Secret* or singers like *Lil Wayne*, *Rihanna* and *Michael Jackson*, these may be interpreted as indicators of socialization, energy and activity. Interests in

Table 3
The LDA pages associated with the top-10 highest and lowest values of L_1 coefficients divided by Big Five personality trait.

Openness		Conscientiousness		Extraversion		Agreeableness		Neuroticism	
Score	Page	Score	Page	Score	Page	Score	Page	Score	Page
7.32	Acting	8.94	Running	13.25	Dancing	8.21	The Bible	6.02	Alice in Wonderland
6.04	Philosophy	3.63	Cappex.com	7.79	Acting	5.36	Toy Story	5.07	Juno
5.27	Drawing	3.57	Grey's Anatomy	7.61	Victoria's Secret Pink	3.99	Camping	4.72	Pedigree Adoption
4.99	Writing Poetry	3.04	Criminal Minds	7.36	Lil Wayne	3.00	Friendship	4.36	Glass
4.00	The Princess Bride	2.89	Jesus Daily	5.26	Superbad	2.74	Everything	3.99	The Sims 3
3.88	Singing	2.77	Cooking	4.49	Everything	2.57	Cuddling	3.70	Hot Topic
3.77	The Alchemist	2.70	HGTV	3.90	Wiz Khalifa	2.49	Chris Tomlin	3.49	Evanescence
3.64	The Boondocks	2.65	QuikTrip	3.86	DJ Pauly D	2.31	Hiking	3.25	My Chemical Romance
3.45	Learning	2.65	Victoria's Secret	3.50	Rihanna	2.31	God	3.14	The Vampire Diaries
3.44	Astrology	2.61	Camping	3.49	Michael Jackson	2.27	Chase Community	3.04	Twilight
…	…	…	…	…	…	…	…	…	…
−2.55	Everybody Loves Raymond	−2.57	Glass	−3.78	Everybody Loves Raymond	−2.09	Paintball	−2.87	The Patriot
−2.56	Cake Boss	−2.58	Hot Topic	−3.82	Anime	−2.25	Animal Farm	−2.99	Michael Jordan
−2.62	The Notebook	−3.20	Billy Mays	−4.01	Linkin Park	−2.32	Natalie Portman	−2.99	Everything
−2.63	Sports	−3.22	The Sims 3	−4.05	StumbleUpon	−2.54	Halloween	−3.04	Hip hop music
−2.69	Hockey	−3.35	Manga	−4.10	The Sims 3	−2.55	Kim Kardashian	−3.84	Sports
−2.74	Buffalo Wild Wings	−3.49	My Phrases	−4.39	Manga	−2.69	Best Quotes	−3.85	Soccer
−2.91	Paintball	−3.64	Marijuana is Safer	−4.81	Alice in Wonderland	−2.74	Alice in Wonderland	−4.29	Running
−3.05	Texas Hold'em Poker	−3.64	Food	−5.13	Evanescence	−2.79	Scarface	−4.39	Superbad
−3.16	Dr Pepper	−3.91	Social Interview	−5.47	Zynga RewardVille	−2.90	Urban Dictionary	−4.58	Hiking
−3.79	Survivor	−5.07	Ray William Johnson	−7.66	NCIS	−3.39	Marilyn Manson	−5.03	Snowboarding

Table 4
The dataset distribution of sensitive and non-sensitive users divided by case study.

Case study	Sensitive	Non-sensitive
Depression	1,403 (1.52%)	90,852 (98.48%)
Aggression	5,904 (6.40%)	86,351 (93.60%)

Table 5
The dataset distribution of influential and non-influential items in the ground truth preferences of sensitive and non-sensitive users divided by case study.

Case study	Item set	Sensitive	Non-sensitive
Depression	Influential	3,138 (2.24%)	316,451 (3.48%)
	Non-Infl.	137,162 (97.76%)	8,768,749 (96.52%)
Aggression	Influential	55,554 (9,41%)	683,124 (7,91%)
	Non-Infl.	534,846 (90,59%)	7,951,976 (92,09%)

anime, manga, and video games like *The Sims 3* and *Zynga*, on the other hand, are negatively associated and can be indicators of shyness and introversion. As for Agreeableness, positively associated interests are *Bible*, *God*, friendship, and cuddling, which can be interpreted as indicators of kindness, generosity, and affection. Sports like paintball, movies like *Scarface* and singers like *Marilyn Manson* are negatively related and can be interpreted as indicators of cruelty, harshness, and coldness. Finally, as for Neuroticism, positive associations can be found with singers such as *Evanescence* and *My Chemical Romance*, TV shows such as *The Vampire Diaries* and films such as *Alice in Wonderland*. These associations can be interpreted as symptoms of tension, anxiety, and moodiness. On the other hand, sports such as running, soccer, snowboarding, and hiking are negatively correlated and can be interpreted as indicators of emotional stability and control.

5.4.2. Analysis of influential items recommendations to sensitive users

In this section, we study the performance of recommender systems for two different case studies. We also give an interpretation of the pages that the algorithms recommend most and least frequently.

As a preliminary step, we created a binary user–item interaction matrix from the dataset presented in Section 5.1. For the experiments, we considered the top-k LDA pages for each user as a binary rating measure. This procedure was performed because the RecTorch library recommendation algorithms [143] did not accept real numbers as input, but only binary ratings.[6] After some empirical tests aimed at selecting only the most relevant pages for 92,255 users, we chose $k = 100$ obtaining a total of 9,225,500 ratings.

Then, we selected two subsets of users from the dataset by filtering their Big Five personality profile. Potentially depressed and potentially aggressive users were selected, respectively, based on the correlation of personality traits with depressive disorders and aggressive behaviors discussed in Section 2.2.1. The subset of potentially depressed users was defined considering users with high Neuroticism, low Extraversion, and low Conscientiousness [15]. The subset of potentially aggressive users was defined by considering users with high Neuroticism and low Agreeableness [16]. For the results we present below, we will refer to both subsets of users as sensitive users, while the rest of the users will be defined as non-sensitive. Next, two subsets of items, favorable and controversial, respectively, were manually selected from the available pages based on an analysis of the topics of the pages. Items were chosen arbitrarily by selecting some of the most correlated pages identified in the previous experiment. This choice is meant to be illustrative of our experiments, but in real-world applications it must be regulated according to well-defined criteria. The subset of favorable items was selected from pages related to sports, sports teams, famous sportsmen, and sports channels. The subset of controversial items was selected instead from pages related to violent sports, war games, alcoholic drinks, and death metal bands. For the results that follow, we will refer to both subsets of items as influential items while the rest of the items will be defined as non-influential items. Next, we studied the performance of two recommenders where items are the pages, for two different case studies.

In the first case, we will study the recommendation of favorable items to potentially depressed users, while in the second case, we will study the recommendation of controversial items to potentially aggressive users.

The distribution of sensitive users is shown in Table 4. The set of potentially aggressive users is higher than the set of potentially depressed users compared to the total: 6.40% of 92,255 instead of 1.52%. The distribution of influential items in ground truth user preferences is shown in Table 5. Potentially depressed users tend to put fewer likes on favorable item pages compared to the non-sensitive group: 2.24% instead of 3.48%. Instead, potentially aggressive users tend to put more likes on controversial item pages than the non-sensitive group: 9.41% compared to 7.91%.

The experimental methodology proceeds as follows for each of the two case studies presented above. We randomly split the dataset vertically on the users into training and test sets (60% / 40%). In the vertical split procedure, users who appear in the

[6] This will be discussed in more detail in Section 6.

Table 6
The overall, sensitive and non-sensitive *NDCG*, the sensitive and non-sensitive *AIR*, the absolute difference $|\Delta_{AIR}|$ and the percentage difference $|\Delta_{AIR}\%|$ divided by case study, backbone recommender, algorithmic approach, and number of k recommended items.

Case study	Backbone	k	Approach	Hyper	NDCG			AIR			
					Overall	Sensitive	Non-Sensitive	Sensitive	Non-Sensitive	$\lvert\Delta_{AIR}\rvert$	$\lvert\Delta_{AIR}\%\rvert$
Depression	SLIM	10	Backbone	–	0.8558	0.8726	0.8556	0.0185	0.0318	0.0133	–
			Diversification	$\eta = 0.10$	0.8535	0.8705	0.8532	0.0362	0.0544	0.0044	33.08%
			Re-ranking	$\alpha = 0.00$	0.8558	0.8726	0.8556	0.0185	0.0318	0.0133	100.00%
			Calibration	$\lambda = 0.60$	0.8558	0.8694	0.8556	0.0316	0.0318	0.0002	1.50%
		25	Backbone	–	0.7590	0.7742	0.7588	0.0290	0.0419	0.0129	–
			Diversification	$\eta = 0.10$	0.7584	0.7739	0.7581	0.0365	0.0509	0.0054	41.86%
			Re-ranking	$\alpha = 0.04$	0.7590	0.7708	0.7588	0.0518	0.0419	0.0099	76.74%
			Calibration	$\lambda = 0.60$	0.7590	0.7722	0.7588	0.0419	0.0419	0.0000	0.00%
		50	Backbone	–	0.8290	0.8408	0.8288	0.0315	0.0454	0.0139	–
			Diversification	$\eta = 0.20$	0.8288	0.8407	0.8286	0.0383	0.0554	0.0071	51.08%
			Re-ranking	$\alpha = 0.04$	0.8290	0.8399	0.8288	0.0510	0.0454	0.0056	40.29%
			Calibration	$\lambda = 0.90$	0.8290	0.8399	0.8288	0.0413	0.0454	0.0041	29.50%
		75	Backbone	–	0.8476	0.8583	0.8474	0.0325	0.0463	0.0138	–
			Diversification	$\eta = 0.30$	0.8475	0.8580	0.8473	0.0413	0.0567	0.0050	36.23%
			Re-ranking	$\alpha = 0.02$	0.8476	0.8583	0.8474	0.0425	0.0463	0.0038	27.54%
			Calibration	$\lambda = 0.99$	0.8476	0.8572	0.8474	0.0436	0.0463	0.0027	19.57%
		100	Backbone	–	0.8572	0.8694	0.8570	0.0334	0.0457	0.0123	–
			Diversification	$\eta = 0.40$	0.8571	0.8693	0.8569	0.0419	0.0555	0.0038	30.89%
			Re-ranking	$\alpha = 0.03$	0.8572	0.8692	0.8570	0.0440	0.0457	0.0017	13.82%
			Calibration	$\lambda = 0.99$	0.8572	0.8689	0.8570	0.0446	0.0457	0.0011	8.94%
	Mult-VAE	10	Backbone	–	0.6980	0.7028	0.6979	0.0219	0.0361	0.0142	–
			Diversification	$\eta = 0.20$	0.6905	0.6984	0.6903	0.0546	0.0832	0.0185	130.28%
			Re-ranking	$\alpha = 0.00$	0.6980	0.7028	0.6979	0.0219	0.0361	0.0142	100.00%
			Calibration	$\lambda = 0.80$	0.6979	0.6998	0.6979	0.0348	0.0361	0.0013	9.15%
		25	Backbone	–	0.6127	0.6247	0.6125	0.0235	0.0425	0.0190	–
			Diversification	$\eta = 0.20$	0.6116	0.6246	0.6114	0.0350	0.0613	0.0075	39.47%
			Re-ranking	$\alpha = 0.04$	0.6127	0.6226	0.6125	0.0522	0.0425	0.0097	51.05%
			Calibration	$\lambda = 0.99$	0.6127	0.6214	0.6125	0.0399	0.0425	0.0026	13.68%
		50	Backbone	–	0.6973	0.7040	0.6972	0.0271	0.0422	0.0151	–
			Diversification	$\eta = 0.30$	0.6969	0.7036	0.6968	0.0375	0.0569	0.0047	31.13%
			Re-ranking	$\alpha = 0.02$	0.6973	0.7032	0.6972	0.0380	0.0422	0.0042	27.81%
			Calibration	$\lambda = 0.99$	0.6973	0.7026	0.6972	0.0348	0.0422	0.0074	49.01%
		75	Backbone	–	0.7334	0.7465	0.7332	0.0278	0.0432	0.0154	–
			Diversification	$\eta = 0.40$	0.7331	0.7465	0.7329	0.0388	0.0570	0.0044	28.57%
			Re-ranking	$\alpha = 0.04$	0.7334	0.7462	0.7332	0.0417	0.0432	0.0015	9.74%
			Calibration	$\lambda = 0.99$	0.7334	0.7459	0.7332	0.0359	0.0432	0.0073	47.40%
		100	Backbone	–	0.7486	0.7626	0.7484	0.0272	0.0411	0.0139	–
			Diversification	$\eta = 0.50$	0.7484	0.7625	0.7482	0.0384	0.0550	0.0027	19.42%
			Re-ranking	$\alpha = 0.03$	0.7486	0.7619	0.7484	0.0425	0.0411	0.0014	10.07%
			Calibration	$\lambda = 0.99$	0.7486	0.7622	0.7484	0.0339	0.0411	0.0072	51.80%
Aggression	SLIM	10	Backbone	–	0.8551	0.8547	0.8551	0.1008	0.0833	0.0175	–
			Diversification	$\eta = 0.00$	0.8551	0.8547	0.8551	0.1008	0.0833	0.0175	100.00%
			Re-ranking	$\beta = 0.20$	0.8549	0.8525	0.8551	0.0859	0.0833	0.0026	14.86%
			Calibration	$\lambda = 0.90$	0.8540	0.8381	0.8551	0.1000	0.0833	0.0167	95.43%
		25	Backbone	–	0.7581	0.7603	0.7580	0.0953	0.0790	0.0163	–
			Diversification	$\eta = 0.00$	0.7581	0.7603	0.7580	0.0953	0.0790	0.0163	100.00%
			Re-ranking	$\beta = 0.20$	0.7581	0.7592	0.7580	0.0887	0.0790	0.0097	59.51%
			Calibration	$\lambda = 0.99$	0.7573	0.7469	0.7580	0.0800	0.0790	0.0010	6.13%
		50	Backbone	–	0.8272	0.8303	0.8270	0.0809	0.0692	0.0118	–
			Diversification	$\eta = 0.00$	0.8272	0.8303	0.8270	0.0809	0.0692	0.0118	100.00%
			Re-ranking	$\beta = 0.12$	0.8270	0.8283	0.8270	0.0703	0.0692	0.0011	9.32%
			Calibration	$\lambda = 0.99$	0.8268	0.8242	0.8270	0.0694	0.0692	0.0002	1.69%
		75	Backbone	–	0.8457	0.8497	0.8455	0.0773	0.0673	0.0100	–
			Diversification	$\eta = 0.00$	0.8457	0.8497	0.8455	0.0773	0.0673	0.0100	100.00%
			Re-ranking	$\beta = 0.02$	0.8457	0.8488	0.8455	0.0681	0.0673	0.0008	8.00%
			Calibration	$\lambda = 0.99$	0.8456	0.8471	0.8455	0.0693	0.0673	0.0020	20.00%
		100	Backbone	–	0.8549	0.8557	0.8549	0.0751	0.0668	0.0083	–
			Diversification	$\eta = 0.00$	0.8549	0.8557	0.8549	0.0751	0.0668	0.0083	100.00%
			Re-ranking	$\beta = 0.10$	0.8549	0.8553	0.8549	0.0671	0.0668	0.0003	3.61%
			Calibration	$\lambda = 0.99$	0.8548	0.8544	0.8549	0.0697	0.0668	0.0029	34.94%
	Mult-VAE	10	Backbone	–	0.6957	0.6938	0.6959	0.1088	0.0880	0.0208	–
			Diversification	$\eta = 0.00$	0.6957	0.6938	0.6959	0.1088	0.0880	0.0208	100.00%
			Re-ranking	$\beta = 0.20$	0.6956	0.6916	0.6959	0.0864	0.0880	0.0016	7.69%
			Calibration	$\lambda = 0.99$	0.6948	0.6800	0.6959	0.1000	0.0880	0.0120	57.69%
		25	Backbone	–	0.6117	0.6152	0.6115	0.0982	0.0781	0.0202	–
			Diversification	$\eta = 0.00$	0.6117	0.6152	0.6115	0.0982	0.0781	0.0202	100.00%
			Re-ranking	$\beta = 0.20$	0.6117	0.6147	0.6115	0.0864	0.0781	0.0083	41.09%
			Calibration	$\lambda = 0.99$	0.6111	0.6060	0.6115	0.0798	0.0781	0.0017	8.42%
		50	Backbone	–	0.6985	0.7035	0.6982	0.0947	0.0788	0.0158	–
			Diversification	$\eta = 0.00$	0.6985	0.7035	0.6982	0.0947	0.0788	0.0158	100.00%
			Re-ranking	$\beta = 0.14$	0.6984	0.7023	0.6982	0.0772	0.0788	0.0016	10.13%
			Calibration	$\lambda = 0.99$	0.6983	0.7000	0.6982	0.0801	0.0788	0.0013	8.23%
		75	Backbone	–	0.7339	0.7367	0.7337	0.0922	0.0772	0.0149	–
			Diversification	$\eta = 0.00$	0.7339	0.7367	0.7337	0.0922	0.0772	0.0149	100.00%
			Re-ranking	$\beta = 0.14$	0.7339	0.7364	0.7337	0.0772	0.0772	0.0000	0.00%
			Calibration	$\lambda = 0.99$	0.7339	0.7355	0.7337	0.0801	0.0772	0.0029	19.46%
		100	Backbone	–	0.7521	0.7548	0.7519	0.0878	0.0745	0.0133	–
			Diversification	$\eta = 0.00$	0.7521	0.7548	0.7519	0.0878	0.0745	0.0133	100.00%
			Re-ranking	$\beta = 0.13$	0.7521	0.7542	0.7519	0.0751	0.0745	0.0006	4.51%
			Calibration	$\lambda = 0.99$	0.7521	0.7545	0.7519	0.0794	0.0745	0.0049	36.84%

Table 7
The LDA pages associated with the highest and lowest SIB_i@100 divided by case study and backbone recommender system.

Depression				Aggression			
SLIM		Mult-VAE		SLIM		Mult-VAE	
Score	Page	Score	Page	Score	Page	Score	Page
0.0728	Naruto Shippuuden	0.1111	Naruto Shippuuden	0.0358	Seether	0.0367	Nine Inch Nails
0.0721	Bleach	0.0936	Vocaloid	0.0306	Dimebag Darrell	0.0350	Seether
0.0709	Patrick Star	0.0936	Naruto	0.0290	Metallica	0.0339	Pantera
0.0698	Daft Punk	0.0912	Gaia Online	0.0287	Superbad	0.0338	Breaking Benjamin
0.0635	deadmau5	0.0882	zOMG!	0.0286	Evanescence	0.0326	Godsmack
0.0582	Courage Wolf	0.0880	deviantART.com	0.0259	Linkin Park	0.0319	Tool
0.0560	Naruto	0.0842	Avenged Sevenfold	0.0248	Stephen King	0.0306	Slipknot
0.0557	PlayStation	0.0827	Manga	0.0243	Nine Inch Nails	0.0302	Fight Club
0.0556	The Colbert Report	0.0704	Korn	0.0236	Nirvana	0.0290	Dimebag Darrell
0.0554	Linkin Park	0.0696	Bleach	0.0233	Shawshank Redemption	0.0290	Korn
...	...	...	...	...	...	...	...
−0.0517	Buffalo Wild Wings	−0.0628	Forever 21	−0.0277	Lance Armstrong	−0.0265	Gucci Mane
−0.0521	H&M	−0.0646	Alicia Keys	−0.0283	I love SLEEP	−0.0267	Movies
−0.0538	Rihanna	−0.0662	Wiz Khalifa	−0.0285	Basketball	−0.0271	Unlimited Texting
−0.0559	Nicki Minaj	−0.0685	T.I.	−0.0292	Social Interview	−0.0273	Running
−0.0572	Family Feud	−0.0715	Family Feud	−0.0300	Movies	−0.0283	I Hate Mosquitos
−0.0579	Chick-fil-A	−0.0765	Eminem	−0.0305	Volleyball	−0.0292	Softball
−0.0624	Basketball	−0.0816	Victoria's Secret	−0.0317	I Hate Mosquitos	−0.0325	The Bible
−0.0642	Victoria's Secret	−0.0828	Drake	−0.0354	Starbucks	−0.0332	Bible
−0.0686	Eminem	−0.0829	Nicki Minaj	−0.0360	Soccer	−0.0357	Sports
−0.0753	T.I.	−0.0928	Victoria's Secret Pink	−0.0380	The Bible	−0.0366	Soccer

training set are not included in the test set. The proportion of sensitive users in both sets was balanced using a stratification procedure. We used 20% of the items per user in the test set as known ratings to avoid cold-start, and the remaining 80% to compute the metrics. We trained two state-of-the-art recommendation algorithms to predict the top-{10, 25, 50, 75, 100}[7] items for each user. Respectively, SLIM, and Mult-VAE were selected to investigate both traditional and deep learning based approaches (Section 5.2). We performed a vertical stratified 5-fold grid search cross-validation on the training set to find the best hyperparameters of the models by optimizing *NDCG*. The search space for SLIM was defined by exploring δ in $\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ and γ in $\{10^{-2}, 10^{-3}\}$ as introduced in Eq. (10). As for Mult-VAE we set the batch size to 512, the annealing steps to 10,000, the ψ regularizer introduced in Eq. (11) to 0.2, and we train for 100 epochs, selecting the model with the best validation *NDCG* while searching for different neural architectures in {*n*-100-*n*, *n*-200-*n*, *n*-400-*n*, *n*-200-100-200-*n*, *n*-400-200-400-*n*} with *n* as the total number of items. We then re-trained the models on the full training set with the hyperparameters found in the previous step and evaluated normalized discounted cumulative gain (*NDCG*), average influential ratio (*AIR*) and sensitive imbalance (*SIB*) on the test set (Section 5.3).

In Table 6 we report the results of various experiments divided by case study (i.e., depression, aggression), backbone recommender system (i.e., SLIM, Mult-VAE), algorithmic approach used (i.e., backbone, diversification, re-ranking and calibration), best selected hyperparameter, and number of k recommended items evaluated on the test set. In addition to the *NDCG* (overall, sensitive and non-sensitive) and *AIR* (sensitive and non-sensitive), we also reported the absolute difference in *AIR* between sensitive and non-sensitive user groups ($|\Delta_{AIR}|$) and the absolute percentage difference $|\Delta_{AIR}\%| = \frac{|\Delta_{AIR}|}{|\Delta_{AIR}|_{bkb}}$ calculated between the $|\Delta_{AIR}|$ and its corresponding backbone value $|\Delta_{AIR}|_{bkb}$. In the following, we discuss the results obtained for the *Backbone* algorithms.

As can be observed, in both case studies, SLIM shows superior *NDCG* performance compared to Mult-VAE both overall and measured in the sensitive and non-sensitive user groups for the different top-k settings. Moreover, *AIR* results indicate a tendency for all algorithms to over-recommend controversial items to potentially aggressive users and to under-recommend favorable items to potentially depressed users compared to the non-sensitive groups.

Below we show a representative example with $k = 100$ to observe the most and least frequently recommended pages for the sensitive user group compared to non-sensitive one. Similar behavior is observed also with other top-k settings.

In Table 7 we show the pages associated with the highest and lowest SIB_i for the different case studies and recommendation algorithms. As can be observed, anime and manga pages such as *Naruto* or *Bleach* are recommended most frequently to the group of potentially depressed users. Sports such as basketball or singers such as *Alicia Keys* and *Eminem* are less frequently recommended. As for potentially aggressive users, heavy metal bands such as *Seether* and *Slipknot* are the most commonly recommended. Sports such as volleyball, soccer, or religious pages about *Bible* or *Jesus* are recommended less frequently.

5.4.3. Analysis of proposed methodologies to balance influential items in sensitive user recommendations

As we observed in the previous experiment, SLIM and Mult-VAE show a tendency to over-recommend controversial items and under-recommend favorable items to sensitive users. Given this tendency, we exploited the methodologies introduced in Section 4, respectively, to promote the recommendation of favorable items to potentially depressed users and to discourage the recommendation of controversial items to potentially aggressive users. In this section, we study the variations in performance resulting from the application of these procedures.

The experiment proceeds as follows. We used the procedure presented in Section 4.1, referred to as *Calibration Approach*, to calibrate the recommendations for both groups of sensitive users based on the item distributions of the non-sensitive user groups. For the experiments, we varied the regularization coefficient λ defined in Eq. (5) in the range $[0, 0.99]$ to balance the distributions of the recommended items until they converge with the target distribution. We then compared the results with those obtained from the application of the procedure introduced in Section 4.2, referred to as *Re-ranking Approach*. In this case, for experiments involving potentially depressed users, we varied the coefficient α defined in Eq. (7) in the range $[0, 0.5]$ to promote the recommendation of favorable items. As for experiments with potentially aggressive users, instead, we varied the threshold β defined in Eq. (7) in the range $[0, 0.5]$ to discourage recommendations of controversial items. Both the α and β parameter ranges were selected to be able to analyze the performance of the

[7] The choice to use top-{10, 25, 50, 75, 100} setting for evaluation is compliant with other work proposing recommender systems based on myPersonality [137,138] and with Steck's original work [10].

algorithms at the point where the item distributions of sensitive users converged with those of non-sensitive users.

In Table 6 we present the results of experiments performed using the experimental setting discussed in Section 5.4.2. In the following we discuss the results obtained for *Calibration* and *Re-ranking* algorithms. Since we are not interested in a performance comparison but in a simple analysis of the results as the regularization parameters vary, we reported the results with parameters that minimize the absolute difference in *AIR* between the sensitive and non-sensitive user groups ($|\Delta_{AIR}|$). However, in a real-world conditions a human operator (and possibly an additional validation set) would be required to set the hyperparameters of the post-processing approaches to obtain an acceptable tradeoff between diversity and the quality of recommendations for sensitive users. As can be observed, compared to backbones, the *AIR* results indicate that the distributions of favorable items recommended to potentially depressed users and controversial items recommended to potentially aggressive users are more similar to those of non-sensitive users. The use of algorithmic approaches allowed in almost[8] all cases to optimize $|\Delta_{AIR}|$ without compromising the *NDCG* both overall and measured on the sensitive user group.

5.4.4. Comparison with other approaches from the recommendation diversity literature

To address the problem discussed in Section 3 other methods from the diversity literature (Section 2.1.1) could also have been used. In this section we study pros and cons of these methodologies[9] and compare their performance with that of the approaches proposed in Section 4.

As discussed in a recent survey [7], there are various approaches in the literature that can be used to introduce diversity into recommendations. It is well understood that the diversification of recommendations also results in the reduction of bias. Accordingly, some of the main diversification algorithms may be used to address the problem proposed in this article.

To understand whether some of the methodologies in the literature were applicable, we first analyzed the main Refs. [17,45–49,53,54,56,57] proposed in the work of Kunaver and Požrl [7]. We found that some of these algorithms could not be applied. Specifically, some algorithms are domain-specific [47] and can only be applied in the music domain. Other algorithms require additional information to be applied that is not available in our case, such as item meta-data descriptions or temporal information [46,48,49] or specific run-time user input [53]. Of the remaining approaches, some [17,45,54,56,57] could be applied to our problem. However, some of these [54,56,57] diversify the recommendations based on popularity criteria and do not take into consideration either the item category (i.e., influential vs non-influential) or the type of user (i.e., sensitive vs non-sensitive). Others [17,45] diversify the recommendations by item category, but do not distinguish the type of user. Consequently, although the latters turn out to be applicable, since they make no distinction between sensitive and non-sensitive users the increase in diversity would also lead to a decrease in overall accuracy.

To compare the performance of diversification algorithms with those proposed in our paper, we implemented Ziegler et al. [17] algorithm [17] as it was considered the most meaningful for our context. The algorithm, referred to as *Diversification*, allows user recommendations to be diversified based on item category through a η diversification factor.

The experimental methodology proceeds as follows. We used the algorithm [17] to diversify user recommendations according to the type of item: whether influential or not-influential. For the experiments, we varied the diversification factor η in the range [0, 1]. Tests were performed exploiting the experimental setting discussed in Section 5.4.2. In the following we discuss the results obtained.

In Table 6 we can observe the results of the *Diversification* algorithm [17] divided by case study, backbone recommender system and number of recommended items. As in the previous experiment, we reported the results with the parameters that minimize the difference in *AIR* between sensitive and non-sensitive users. To make a fair comparison in this case we calculated the $|\Delta_{AIR}|$ between the *AIR* of sensitive users its corresponding non-sensitive backbone value. As can be seen by comparing the results with our methods: in the case of potentially depressed users, the diversification algorithm shows on average a higher $|\Delta_{AIR}\%|$ w.r.t. those obtained from Calibration and Re-ranking; in the case of potentially aggressive users, the algorithm does not obtain satisfactory performance, and the best results are obtained when no diversification is made ($\eta = 0.00$). The latter result on potentially aggressive users is due to the fact that, because of the design of the algorithm, the diversification always results in an upward shift of the distribution of influential and non-influential items, and thus the controversial items of sensitive users never fall to the same level as those of non-sensitive users. In addition, the algorithm changes the item rankings also for non-sensitive users. This results in a slightly lower non-sensitive *NDCG* performance when compared to our methods that leave the item rankings of non-sensitive users unaltered.

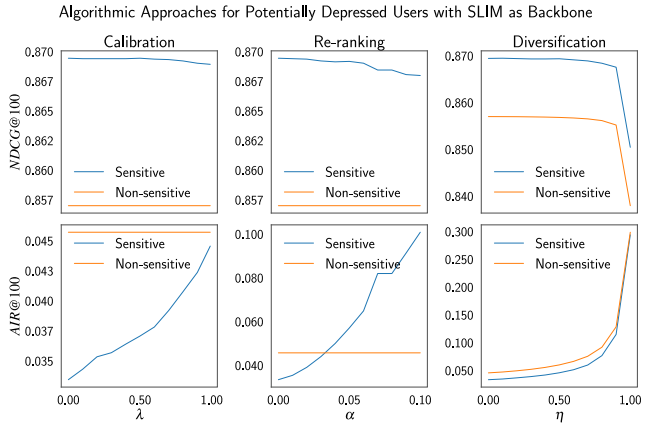
Below we show a representative example with $k = 100$ to observe how the metrics measured for the *Calibration*, *Re-ranking* and *Diversification* approaches vary as the hyperparameters vary. Other top-k settings show similar behavior.

Fig. 6 shows the results of the experiments that compare the different methodologies used to optimize the recommendations for potentially depressed users divided by backbone recommender system. As can be observed, using the calibration approach, as the coefficient λ increases, the *AIR*@100 tends to converge with the average percentage of influential items of non-sensitive users. Exploiting the re-ranking approach, on the other hand, as the α coefficient increases, the *AIR*@100 first converges with those of non-sensitive users and then exceeds them. As for the diversification approach, instead, by increasing the η parameter, the *AIR*@100 of influential items tends to increase and converge with that of non-influential items. All results were expected, as calibration aims to calibrate the item distributions of the recommendations for sensitive users until they converge with the target distribution, while re-ranking is used with the aim of promoting favorable item recommendations for potentially depressed users based on a percentage criterion and diversification is applied to both sensitive and non-sensitive users. In both re-ranking and calibration cases, the *NDCG*@100 of sensitive users decreases slightly without compromising overall performance. As for diversification, on the other hand, the decrease in *NDCG*@100 is much more important than calibration and re-ranking because it is applied also to non-sensitive users.

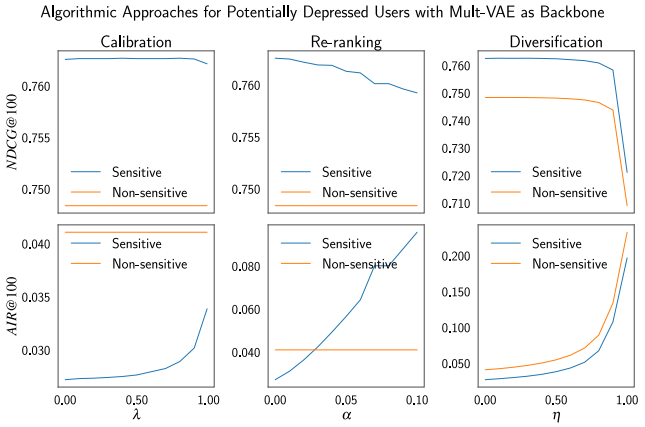
In Fig. 7 we show the results of the experiments for potentially aggressive users. Overall outcomes are in line with previous ones with a small exception in the re-ranking approach. Since constraints tend to discourage rather than promote controversial items in sensitive users recommendations, as β decreases, the

[8] In the case of potentially depressed users with $k = 10$, the re-ranking algorithm do not to improve because, for numerical reasons, the % of influential items it adds to the sensitive users rank as α increases is always too high compared to that of non-sensitive users.

[9] Note that in addition to the diversity algorithms, other fairness algorithms (Section 2.1.2), if repurposed, could also have been used to address the problem. However, since the objective of the paper is to introduce the problem and propose two initial solutions, repurposing other fairness algorithms does not currently result in scope but would certainly be a promising future research direction.

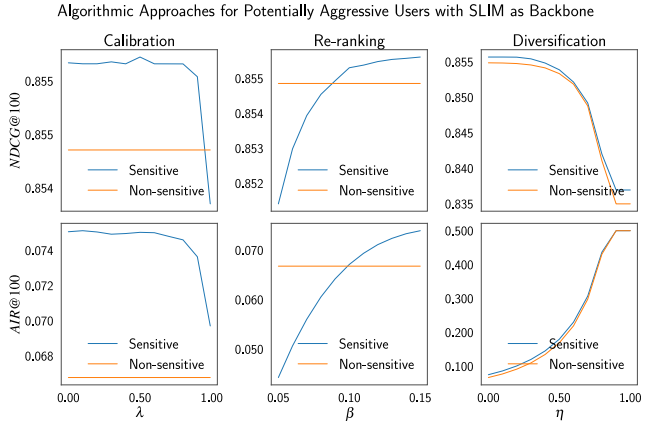


(a) The $NDCG$@100 (upper corners) and AIR@100 (lower corners) obtained using SLIM as backbone recommender. Results obtained from the calibration approach (left column) are compared with those obtained from re-ranking (central column) and diversification (right column).

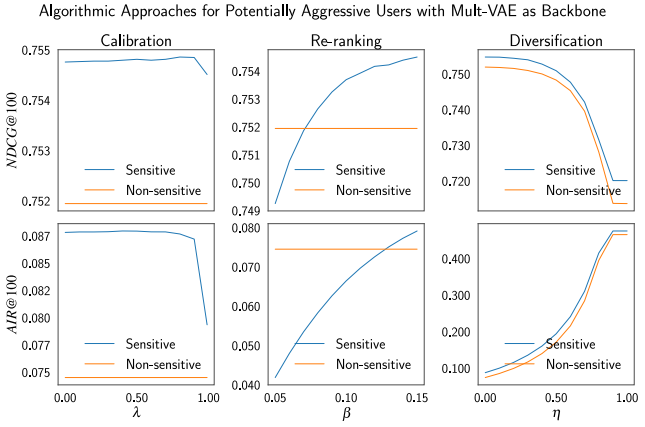


(b) The $NDCG$@100 (upper corners) and AIR@100 (lower corners) obtained using Mult-VAE as backbone recommender. Results obtained from the calibration approach (left column) are compared with those obtained from re-ranking (central column) and diversification (right column).

Fig. 6. The results of the experiments for potentially depressed users divided by backbone recommender system (SLIM left, Mult-VAE right) and algorithmic approach.



(a) The $NDCG$@100 (upper corners) and AIR@100 (lower corners) obtained using SLIM as backbone recommender. Results obtained from the calibration approach (left column) are compared with those obtained from re-ranking (central column) and diversification (right column).



(b) The $NDCG$@100 (upper corners) and AIR@100 (lower corners) obtained using Mult-VAE as backbone recommender. Results obtained from the calibration approach (left column) are compared with those obtained from re-ranking (central column) and diversification (right column).

Fig. 7. The results of the experiments for potentially aggressive users divided by backbone recommender system (SLIM left, Mult-VAE right) and algorithmic approach.

behavior we observe is the opposite of the depression case. In this case, as β decreases, the *AIR*@100 of the sensitive users first converges with that of the non-sensitive users and then diverges both in the case where the backbone is based on SLIM and Mult-VAE. Overall, as expected, sensitive users *NDCG*@100 decreases as λ and η increases and as β decreases. The *NDCG* performance are not compromised in the case of re-ranking and calibration but there is a greater impact in the case of diversification since is applied also to non-sensitive users. Moreover, the diversification algorithm does not achieve satisfactory results in the case of potentially aggressive users because it never results in a reduction in the *AIR*@100 of sensitive users but always in its increase.

5.4.5. Analysis of proposed methodology to combine different approaches together

In this section we study the performance of a method that combines the output of different algorithms simultaneously with the goal of further improving the results and to relieve a practitioner from the need of selecting a-priori a single approach.

The experiment proceeds as follows. We used the methodology presented in Section 4.3, referred to as the *Combination Approach* to combine the rankings obtained from *Calibration*, *Re-ranking* and *Diversification* algorithms discussed in previous sections. For each experimental scenario (i.e., the tuple consisting of case study, backbone recommender and number of recommender items), the algorithm was executed by combining the output of the techniques and hyper-parameters reported in Table 6.

In Table 8, we present the results of the *Combination* method divided by case study, backbone recommender system and number of recommended items, exploiting the experimental setting discussed in Section 5.4.2. The results are compared with those of the techniques that had obtained the best outcomes in previous experiments. As can be observed, the combination approach

Table 8
The overall, sensitive and non-sensitive *NDCG*, the sensitive and non-sensitive *AIR*, the absolute difference $|\Delta_{AIR}|$ and the percentage difference $|\Delta_{AIR}\%|$ divided by case study, backbone recommender, and number of k recommended items for the combination approach compared to the approaches that best performed in the previous experiments.

Case study	Backbone	k	Approach	Hyper	NDCG			AIR			
					Overall	Sensitive	Non-sensitive	Sensitive	Non-sensitive	$\|\Delta_{AIR}\|$	$\|\Delta_{AIR}\%\|$
Depression	SLIM	10	Calibration	$\lambda = 0.60$	0.8558	0.8694	0.8556	0.0316	0.0318	0.0002	1.50%
			Combination	–	0.8558	0.8700	0.8556	0.0310	0.0318	0.0008	6.02%
		25	Calibration	$\lambda = 0.60$	0.7590	0.7722	0.7588	0.0419	0.0419	0.0000	0.00%
			Combination	–	0.7589	0.7700	0.7588	0.0418	0.0419	0.0001	0.78%
		50	Calibration	$\lambda = 0.90$	0.8290	0.8399	0.8288	0.0413	0.0454	0.0041	29.50%
			Combination	–	0.8290	0.8395	0.8288	0.0452	0.0454	0.0002	1.44%
		75	Calibration	$\lambda = 0.99$	0.8476	0.8572	0.8474	0.0436	0.0463	0.0027	19.57%
			Combination	–	0.8476	0.8572	0.8474	0.0463	0.0463	0.0000	0.00%
		100	Calibration	$\lambda = 0.99$	0.8572	0.8689	0.8570	0.0446	0.0457	0.0011	8.94%
			Combination	–	0.8572	0.8689	0.8570	0.0457	0.0457	0.0000	0.00%
	Mult-VAE	10	Calibration	$\lambda = 0.80$	0.6979	0.6998	0.6979	0.0348	0.0361	0.0013	9.15%
			Combination	–	0.6979	0.7004	0.6979	0.0357	0.0361	0.0004	2.82%
		25	Calibration	$\lambda = 0.99$	0.6127	0.6214	0.6125	0.0399	0.0425	0.0026	13.68%
			Combination	–	0.6127	0.6216	0.6125	0.0424	0.0425	0.0001	0.53%
		50	Re-ranking	$\alpha = 0.02$	0.6973	0.7032	0.6972	0.0380	0.0422	0.0042	27.81%
			Combination	–	0.6973	0.7019	0.6972	0.0421	0.0422	0.0001	0.66%
		75	Re-ranking	$\alpha = 0.04$	0.7334	0.7462	0.7332	0.0417	0.0432	0.0015	9.74%
			Combination	–	0.7334	0.7455	0.7332	0.0431	0.0432	0.0001	0.65%
		100	Re-ranking	$\alpha = 0.03$	0.7486	0.7619	0.7484	0.0425	0.0411	0.0014	10.07%
			Combination	–	0.7486	0.7619	0.7484	0.0409	0.0411	0.0002	1.44%
Aggression	SLIM	10	Re-ranking	$\beta = 0.20$	0.8549	0.8525	0.8551	0.0859	0.0833	0.0026	14.86%
			Combination	–	0.8543	0.8426	0.8551	0.0833	0.0833	0.0000	0.00%
		25	Calibration	$\lambda = 0.99$	0.7573	0.7469	0.7580	0.0800	0.0790	0.0010	6.13%
			Combination	–	0.7573	0.7470	0.7580	0.0790	0.0790	0.0000	0.00%
		50	Calibration	$\lambda = 0.99$	0.8268	0.8242	0.8270	0.0694	0.0692	0.0002	1.69%
			Combination	–	0.8268	0.8246	0.8270	0.0691	0.0692	0.0001	0.85%
		75	Re-ranking	$\beta = 0.02$	0.8457	0.8488	0.8455	0.0681	0.0673	0.0008	8.00%
			Combination	–	0.8456	0.8471	0.8455	0.0673	0.0673	0.0000	0.00%
		100	Re-ranking	$\beta = 0.10$	0.8549	0.8553	0.8549	0.0671	0.0668	0.0003	3.61%
			Combination	–	0.8548	0.8544	0.8549	0.0668	0.0668	0.0000	0.00%
	Mult-VAE	10	Re-ranking	$\beta = 0.20$	0.6956	0.6916	0.6959	0.0864	0.0880	0.0016	7.69%
			Combination	–	0.6950	0.6822	0.6959	0.0880	0.0880	0.0000	0.00%
		25	Calibration	$\lambda = 0.99$	0.6111	0.6060	0.6115	0.0798	0.0781	0.0017	8.42%
			Combination	–	0.6112	0.6062	0.6115	0.0781	0.0781	0.0000	0.00%
		50	Calibration	$\lambda = 0.99$	0.6983	0.7000	0.6982	0.0801	0.0788	0.0013	8.23%
			Combination	–	0.6983	0.7005	0.6982	0.0788	0.0788	0.0000	0.00%
		75	Re-ranking	$\beta = 0.14$	0.7339	0.7364	0.7337	0.0772	0.0772	0.0000	0.00%
			Combination	–	0.7339	0.7356	0.7337	0.0772	0.0772	0.0000	0.00%
		100	Re-ranking	$\beta = 0.13$	0.7521	0.7542	0.7519	0.0751	0.0745	0.0006	4.51%
			Combination	–	0.7521	0.7543	0.7519	0.0745	0.0745	0.0000	0.00%

achieved better results in almost[10] all cases, further decreasing the $|\Delta_{AIR}\%|$ and keeping the overall, sensitive and non-sensitive *NDCG* high. The algorithm in particular seems to perform very well in the case of potentially aggressive users. This can be explained intuitively based on the design of the algorithm. In fact, as previously discussed in Section 4.3 since the algorithm is incremental and iterates on all sensitive users: for the first sensitive users processed the $|\Delta_{AIR}|$ tends to vary; as the number of users processed increases, the $|\Delta_{AIR}|$ tends to stabilize and decrease. Consequently, since the number of potentially aggressive users is higher than the number of potentially depressed users (Table 4), the algorithm obtains better results in the former case, because the initial instability phase is absorbed by the high number of users.

6. Discussion and limitations of the study

In the previous experiments, we proposed two different cases to study the recommendation of favorable items (e.g. sports and hobbies pages) to potentially depressed users and the recommendation of controversial items (e.g. alcohol, weapons, and death metal pages) to potentially aggressive users. As we have seen (see Section 5.4.2), in some cases recommender systems tend to over-recommend controversial items and to under-recommend favorable items to sensitive users. The techniques introduced in Section 4 were used to promote the recommendation of favorable items to potentially depressed users and discourage the recommendation of controversial items to potentially aggressive users (Experiment 5.4.3). These techniques were also compared with one of the most significant recommendation diversification algorithms [17] for our context in Section 5.4.4. In Section 5.4.5

[10] The only case where *Combination* does not get better results or very close to the baseline results is for potentially depressed users and $k = 10$. In this case, the number of sensitive users is not high enough (Table 4) to perfectly stabilize the $|\Delta_{AIR}|$. Consequently, as the algorithm iterates over all sensitive users by choosing the best performing rank at each iteration, it does not select only the ranks generated by calibration (that is the best performing method), but also the ones from re-ranking and diversification that did not perform well. As a result, it still gets good results, but suffers from the variability of the ranks selected in the first iterations.

finally we also tried to combine the outputs from the various approaches used together to further improve the results. All the proposed techniques proved to be valid to address the problem, but with some differences. In this section, we discuss the pros and cons of the two approaches and the limitations of the present study.

The calibration algorithm proved to be a valid approach, but sometimes the distributions of influential items in sensitive and non-sensitive user recommendations did not converge. As λ in Eq. (5) increases it was expected that sensitive users *NDCG* would decrease and the absolute difference in *AIR* would decrease. If $\lambda = 0.99$ is set, the difference between the distributions should be minimized. However, the distributions often did not converge because the KL-based distance function tends to suffer from numerical instability when the difference between the distributions is small. Moreover, the computational complexity was higher compared to the constrained re-ranking approach and consequently also the execution time. Regarding the constraint-based approach, it proved to be effective and fast. Varying the parameters α and β in Eq. (7) allowed to easily balance the distribution of influential items. However, it was necessary to study the performance of the algorithm in the α and β ranges to obtain results that maximize *NDCG* and minimize the absolute difference in *AIR*. In addition, analyzing the performance of the diversity algorithm that could best be exploited in our context [17] as comparison approach, although it allowed for the diversification of recommendations based on item category, it showed lower overall performance than our methods. Furthermore, the algorithm could not be satisfactorily applied to the case of potentially aggressive users because, by algorithm design, as η increases, *AIR* always tends to rise and that of sensitive users never falls to the level of that of non-sensitive ones. As for the combination approach, on the other hand, the algorithm presented some interesting features. Although the order in which the algorithm iterated over the sensitive users may have varied the final results, as different rankers were used to generate the candidate lists, the recommendations addressed to the single sensitive user still enjoyed the same properties as the original rankers used. Since rankers were selected based on the hyperparameters maximizing certain diversity indicators, the recommendations addressed to the single sensitive user always presented a certain degree of diversification of influential items. Moreover, iterating over the users, as the number of sensitive users increased, the overall $|\Delta_{AIR}|$ tended by design to stabilize and decrease.

We identified some limitations of the present work regarding the data behind the experiments, the way in which sensitive users and influential items were selected, and the algorithms used to address the problem.

Regarding the data, as far as the authors know, the current literature lacked datasets compatible with our experimental settings, i.e., containing information about the users that would have allowed the identification of sensitive ones and item descriptions that could have been used to identify influential items. In particular, although there were other datasets in the literature besides myPersonality that reported Big Five personality traits such as Personality2018 [146], ADS [147], and PsychoFlickr [148], these did not have the necessary characteristics to be used for experiments. In particular, Personality2018 [146] lacked contextual features to identify influential items and was based on a small set of only 1800 users. ADS [147], while being compatible with our experimental setting, exhibited an excessively small size to be used for experiments, i.e. 120 users and 300 items only. Assuming the same percentage of sensitive users as in our considered dataset, the identified number of sensitive users would have been too small (e.g., less than 10 users) to obtain meaningful results. Finally, PsychoFlickr [148], presented the dimensionality problems observed in previous datasets (i.e., 300 users), and was not suitable for use in collaborative filtering scenarios since each user was associated with a different set of items.

Considering instead the data used for experiments based on myPersonality, the pages that users liked were determined by breaking down the LDA topics of the most popular pages. Moreover, to train recommender systems, it was necessary to binarize the user–item interaction matrix considering only the top-k items for each user because the algorithms in the RecTorch library [143] did not accept real values as input. Furthermore, consistently with other work proposing recommendation systems based on myPersonality [137,138] and the work of Steck [10] it was chosen to evaluate the results of the experiments using the top-$\{10, 25, 50, 75, 100\}$ settings. Together, these factors could have potentially influenced the experiments and interpretations of the results.

Some other points concerned the selection of sensitive users and influential items. As for sensitive users, in the experiments, it was proposed to select them based on certain correlations known in the literature with the Big Five personality traits. To define user groups, correlations with depressive disorders [15] and aggressive behavior [16] were exploited. Although initially relying on correlation might be a valid approach, this could lead to type 1 and 2 errors, selecting users who do not really suffer from the disorder or excluding others who do. Moreover, as for the influential items, these have been selected manually from an analysis of the topics of available pages. For both of the previous points, the selection procedure may have introduced errors and influenced the experiments.

As for the algorithms, the experiments were based exclusively on SLIM [33] and Mult-VAE [34] as backbones. Furthermore, optimization approaches were designed using exclusively post-processing methodologies taking inspiration from algorithmic fairness literature (Section 2.1.2). Since the purpose of this article was to introduce the problem and propose two solutions to address it, it was not studied how to adapt other fairness algorithms based on post-processing and in-processing approaches. Moreover, although we have also proposed an initial methodology to combine the outputs of different rankers simultaneously to further improve results, there are many studies in the literature addressing hybrid recommendation [12] and rank fusion [13] and new approaches could be developed or other existing algorithms repurposed. Together, these elements represent promising future research directions.

7. Conclusion and future works

In this paper, we addressed the problem of recommending influential items to sensitive users. We defined as sensitive, users whose behavior can be influenced by specific types of items. Similarly we referred to influential, as those items that can influence sensitive users' behavior. In our study, we formalized the problem and proposed two techniques to maximize the performance of a recommender system that aimed to diversify the item distribution to positively affect sensitive users' behavior: mitigating potentially dangerous societal consequences and promoting healthier lifestyles. The first technique was a calibration approach that aimed to balance sensitive users' recommendations based on the distribution of non-sensitive users' influential items. The second technique was a re-ranking approach that aimed to optimize the performance of a recommender system under influential items' constraints. We also proposed a joint approach to combine the outputs of any technique together to achieve better results. We considered a real-world dataset to test the proposed techniques in two different case studies that involved potentially aggressive and depressive users. All techniques proved

effective in allowing high performance to be maintained while diversifying influential items.

We identified several future research directions. An initial research direction could focus on building more datasets to use for experiments. Furthermore, designing automatic methods to select sensitive users and influential items could also be a worthwhile research path. In fact, although the Big Five model provided a theoretical starting point to select sensitive users, asking platform users to fill out a questionnaire is not feasible in most real-world circumstances. In addition, even manually selecting the influential items would not be feasible. Another research direction might focus on diversification algorithms. The algorithms proposed in this paper took inspiration from approaches in the literature of algorithmic fairness applied to recommender systems. A promising research direction could aim to re-adapt other fairness algorithms to address the proposed problem as well. Moreover, the approaches proposed in this article were based on post-processing techniques and considered only one set of influential items for each sensitive user group. Research could investigate the use of in-processing methodologies and extend the problem to manage multiple influential item sets simultaneously. It could also be further investigated how to combine several algorithms simultaneously to achieve superior performance using methodologies from the literature on hybrid recommendation and rank fusion. Finally, another research direction could seek to study how user behavior is affected by the proposed techniques through the use of simulations.

CRediT authorship contribution statement

Alvise De Biasio: Conceptualization, Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Merylin Monaro:** Conceptualization, Resources, Writing – review & editing. **Luca Oneto:** Resources, Writing – review & editing. **Lamberto Ballan:** Supervision. **Nicolò Navarin:** Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgment

This work was partially funded by estilos srl.

References

[1] F. Ricci, L. Rokach, B. Shapira, Introduction to recommender systems handbook, in: Recommender Systems Handbook, Springer, 2011, pp. 1–35, http://dx.doi.org/10.1007/978-0-387-85820-3_1.

[2] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, Recommender systems survey, Knowl.-Based Syst. 46 (2013) 109–132, http://dx.doi.org/10.1016/j.knosys.2013.03.012.

[3] D. Bollen, B.P. Knijnenburg, M.C. Willemsen, M. Graus, Understanding choice overload in recommender systems, in: Proceedings of the Fourth ACM Conference on Recommender Systems, 2010, pp. 63–70, http://dx.doi.org/10.1145/1864708.1864724.

[4] M. Cinelli, G.D.F. Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, The echo chamber effect on social media, Proc. Natl. Acad. Sci. 118 (9) (2021) http://dx.doi.org/10.1073/pnas.2023301118.

[5] C.R. Sunstein, The Law of Group Polarization. Vol. 91, University of Chicago Law School, John M. Olin Law & Economics Working Paper, 1999, http://dx.doi.org/10.7551/mitpress/11974.003.0005.

[6] J. Esteban, G. Schneider, Polarization and conflict: Theoretical and empirical issues, J. Peace Res. 45 (2) (2008) 131–141, http://dx.doi.org/10.1177/0022343307087168.

[7] M. Kunaver, T. Požrl, Diversity in recommender systems–A survey, Knowl.-Based Syst. 123 (2017) 154–162, http://dx.doi.org/10.1016/j.knosys.2017.02.009.

[8] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, A. Galstyan, A survey on bias and fairness in machine learning, ACM Comput. Surv. 54 (6) (2021) 1–35, http://dx.doi.org/10.1145/3457607.

[9] E. Pitoura, K. Stefanidis, G. Koutrika, Fairness in rankings and recommendations: An overview, VLDB J. (2022) 1–28, http://dx.doi.org/10.1007/s00778-021-00697-y.

[10] H. Steck, Calibrated recommendations, in: Proceedings of the 12th ACM Conference on Recommender Systems, 2018, pp. 154–162, http://dx.doi.org/10.1145/3240323.3240372.

[11] M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed, R. Baeza-Yates, Fa* ir: A fair top-k ranking algorithm, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 1569–1578, http://dx.doi.org/10.1145/3132847.3132938.

[12] R. Burke, Hybrid recommender systems: Survey and experiments, User Model. User-Adapt. Interact. 12 (4) (2002) 331–370, http://dx.doi.org/10.1023/A:1021240730564.

[13] O. Kurland, J.S. Culpepper, Fusion in information retrieval: Sigir 2018 half-day tutorial, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 1383–1386, http://dx.doi.org/10.1145/3209978.3210186.

[14] E.C. Tupes, R.E. Christal, Recurrent personality factors based on trait ratings, J. Personal. 60 (2) (1992) 225–251, http://dx.doi.org/10.1111/j.1467-6494.1992.tb00973.x.

[15] R. Kotov, W. Gamez, F. Schmidt, D. Watson, Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis, Psychol. Bull. 136 (5) (2010) 768, http://dx.doi.org/10.1037/a0020327.

[16] C.P. Barlett, C.A. Anderson, Direct and indirect relations between the Big 5 personality traits and aggressive and violent behavior, Pers. Individ. Differ. 52 (8) (2012) 870–875, http://dx.doi.org/10.1016/j.paid.2012.01.029.

[17] C.-N. Ziegler, S.M. McNee, J.A. Konstan, G. Lausen, Improving recommendation lists through topic diversification, in: Proceedings of the 14th International Conference on World Wide Web, 2005, pp. 22–32, http://dx.doi.org/10.1145/1060745.1060754.

[18] Z. Li, D. Amagata, Y. Zhang, T. Maekawa, T. Hara, K. Yonekawa, M. Kurokawa, HML4Rec: Hierarchical meta-learning for cold-start recommendation in flash sale e-commerce, Knowl.-Based Syst. 255 (2022) 109674, http://dx.doi.org/10.1016/j.knosys.2022.109674.

[19] N. Qiu, B. Gao, H. Tu, F. Huang, Q. Guan, W. Luo, LDGC-SR: Integrating long-range dependencies and global context information for session-based recommendation, Knowl.-Based Syst. 248 (2022) 108894, http://dx.doi.org/10.1016/j.knosys.2022.108894.

[20] Y. Chen, J. Wang, Z. Wu, Y. Lin, Integrating User-Group relationships under interest similarity constraints for social recommendation, Knowl.-Based Syst. 249 (2022) 108921, http://dx.doi.org/10.1016/j.knosys.2022.108921.

[21] A. De Biasio, A. Montagna, F. Aiolli, N. Navarin, A systematic review of value-aware recommender systems, Expert Syst. Appl. (2023) 120131, http://dx.doi.org/10.1016/j.eswa.2023.120131, URL: https://www.sciencedirect.com/science/article/pii/S0957417423006334.

[22] D. Jannach, M. Jugovac, Measuring the business value of recommender systems, ACM Trans. Manag. Inform. Syst. (TMIS) 10 (4) (2019) 1–23, http://dx.doi.org/10.1145/3370082.

[23] D. Kotkov, S. Wang, J. Veijalainen, A survey of serendipity in recommender systems, Knowl.-Based Syst. 111 (2016) 180–192, http://dx.doi.org/10.1016/j.knosys.2016.08.014.

[24] M. Deshpande, G. Karypis, Item-based top-n recommendation algorithms, ACM Trans. Inform. Syst. (TOIS) 22 (1) (2004) 143–177, http://dx.doi.org/10.1145/963770.963776.

[25] H. Ko, S. Lee, Y. Park, A. Choi, A survey of recommendation systems: Recommendation models, techniques, and application fields, Electronics 11 (1) (2022) 141, http://dx.doi.org/10.3390/electronics11010141.

[26] X. Su, T.M. Khoshgoftaar, A survey of collaborative filtering techniques, Adv. Artif. Intell. 2009 (2009) http://dx.doi.org/10.1155/2009/421425.

[27] N. Nassar, A. Jafar, Y. Rahhal, A novel deep multi-criteria collaborative filtering model for recommendation system, Knowl.-Based Syst. 187 (2020) 104811, http://dx.doi.org/10.1016/j.knosys.2019.06.019.

[28] P. Lops, M.d. Gemmis, G. Semeraro, Content-based recommender systems: State of the art and trends, in: Recommender Systems Handbook, Springer, 2011, pp. 73–105, http://dx.doi.org/10.1007/978-0-387-85820-3_3.

[29] D. Wang, Y. Liang, D. Xu, X. Feng, R. Guan, A content-based recommender system for computer science publications, Knowl.-Based Syst. 157 (2018) 1–9, http://dx.doi.org/10.1016/j.knosys.2018.05.001.

[30] C. Desrosiers, G. Karypis, A comprehensive survey of neighborhood-based recommendation methods, Recomm. Syst. Handb. (2011) 107–144, http://dx.doi.org/10.1007/978-0-387-85820-3_4.

[31] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, Computer 42 (8) (2009) 30–37, http://dx.doi.org/10.1109/MC.2009.263.

[32] E. Yang, Y. Huang, F. Liang, W. Pan, Z. Ming, FCMF: Federated collective matrix factorization for heterogeneous collaborative filtering, Knowl.-Based Syst. 220 (2021) 106946, http://dx.doi.org/10.1016/j.knosys.2021.106946.

[33] X. Ning, G. Karypis, Slim: Sparse linear methods for top-n recommender systems, in: 2011 IEEE 11th International Conference on Data Mining, IEEE, 2011, pp. 497–506, http://dx.doi.org/10.1109/ICDM.2011.134.

[34] D. Liang, R.G. Krishnan, M.D. Hoffman, T. Jebara, Variational autoencoders for collaborative filtering, in: Proceedings of the 2018 World Wide Web Conference, 2018, pp. 689–698, http://dx.doi.org/10.1145/3178876.3186150.

[35] G. Shani, A. Gunawardana, Evaluating recommendation systems, in: Recommender Systems Handbook, Springer, 2011, pp. 257–297, http://dx.doi.org/10.1007/978-0-387-85820-3_8.

[36] M. Kaminskas, D. Bridge, Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems, ACM Trans. Interact. Intell. Syst. (TiiS) 7 (1) (2016) 1–42, http://dx.doi.org/10.1145/2926720.

[37] N. Hurley, M. Zhang, Novelty and diversity in top-n recommendation–analysis and evaluation, ACM Trans. Internet Technol. (TOIT) 10 (4) (2011) 1–30, http://dx.doi.org/10.1145/1944339.1944341.

[38] N. Tintarev, M. Dennis, J. Masthoff, Adapting recommendation diversity to openness to experience: a study of human behaviour, in: International Conference on User Modeling, Adaptation, and Personalization, Springer, 2013, pp. 190–202, http://dx.doi.org/10.1007/978-3-642-38844-6_16.

[39] M. Ge, F. Gedikli, D. Jannach, Placing high-diversity items in top-n recommendation lists, in: ITWP@ IJCAI, 2011.

[40] M.D. Ekstrand, F.M. Harper, M.C. Willemsen, J.A. Konstan, User perception of differences in recommender algorithms, in: Proceedings of the 8th ACM Conference on Recommender Systems, 2014, pp. 161–168, http://dx.doi.org/10.1145/2645710.2645737.

[41] K. Bradley, B. Smyth, Improving recommendation diversity, in: Proceedings of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science, Maynooth, Ireland, Vol. 85, Citeseer, 2001, pp. 141–152.

[42] C.L. Clarke, M. Kolla, G.V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, I. MacKinnon, Novelty and diversity in information retrieval evaluation, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2008, pp. 659–666, http://dx.doi.org/10.1145/1390334.1390446.

[43] S. Vargas, New approaches to diversity and novelty in recommender systems, in: Fourth BCS-IRSG Symposium on Future Directions in Information Access, Vol. 4, FDIA 2011, 2011, pp. 8–13, http://dx.doi.org/10.14236/ewic/FDIA2011.2.

[44] S. Vargas, L. Baltrunas, A. Karatzoglou, P. Castells, Coverage, redundancy and size-awareness in genre diversity for recommender systems, in: Proceedings of the 8th ACM Conference on Recommender Systems, 2014, pp. 209–216, http://dx.doi.org/10.1145/2645710.2645743.

[45] W. Premchaiswadi, P. Poompuang, N. Jongswat, N. Premchaiswadi, Enhancing diversity-accuracy technique on user-based top-n recommendation algorithms, in: 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops, IEEE, 2013, pp. 403–408, http://dx.doi.org/10.1109/COMPSACW.2013.68.

[46] Z. Abbassi, V.S. Mirrokni, M. Thakur, Diversity maximization under matroid constraints, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2013, pp. 32–40, http://dx.doi.org/10.1145/2487575.2487636.

[47] M. Slaney, W. White, Measuring playlist diversity for recommendation systems, in: Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia, 2006, pp. 77–82, http://dx.doi.org/10.1145/1178723.1178735.

[48] S.-M. Choi, Y.-S. Han, A content recommendation system based on category correlations, in: 2010 Fifth International Multi-Conference on Computing in the Global Information Technology, IEEE, 2010, pp. 66–70, http://dx.doi.org/10.1109/ICCGI.2010.31.

[49] N. Lathia, S. Hailes, L. Capra, X. Amatriain, Temporal diversity in recommender systems, in: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2010, pp. 210–217, http://dx.doi.org/10.1145/1835449.1835486.

[50] G. Adomavicius, Y. Kwon, Improving aggregate recommendation diversity using ranking-based techniques, IEEE Trans. Knowl. Data Eng. 24 (5) (2011) 896–911, http://dx.doi.org/10.1109/TKDE.2011.15.

[51] S. Wang, M. Gong, H. Li, J. Yang, Multi-objective optimization for long tail recommendation, Knowl.-Based Syst. 104 (2016) 145–155, http://dx.doi.org/10.1016/j.knosys.2016.04.018.

[52] C. Ren, P. Zhu, H. Zhang, A new Collaborative Filtering technique to improve recommendation diversity, in: 2016 2nd IEEE International Conference on Computer and Communications, ICCC, IEEE, 2016, pp. 1279–1282, http://dx.doi.org/10.1109/CompComm.2016.7924910.

[53] D. Bridge, J.P. Kelly, Ways of computing diverse collaborative recommendations, in: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Springer, 2006, pp. 41–50, http://dx.doi.org/10.1007/11768012_6.

[54] Y.-C. Ho, Y.-T. Chiang, J.Y.-J. Hsu, Who likes it more? Mining worth-recommending items from long tails by modeling relative preference, in: Proceedings of the 7th ACM International Conference on Web Search and Data Mining, 2014, pp. 253–262, http://dx.doi.org/10.1145/2556195.2566589.

[55] L. Chen, G. Zhang, H. Zhou, Improving the diversity of top-N recommendation via determinantal point process, in: Large Scale Recommendation Systems Workshop, 2017.

[56] R. Boim, T. Milo, S. Novgorodov, Diversification and refinement in collaborative filtering recommender, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, 2011, pp. 739–744, http://dx.doi.org/10.1145/2063576.2063684.

[57] K. Lee, K. Lee, Escaping your comfort zone: A graph-based recommender system for finding novel recommendations among relevant items, Expert Syst. Appl. 42 (10) (2015) 4851–4858, http://dx.doi.org/10.1016/j.eswa.2014.07.024.

[58] A.L. Zanon, L.C.D. da Rocha, M.G. Manzato, Balancing the trade-off between accuracy and diversity in recommender systems with personalized explanations based on Linked Open Data, Knowl.-Based Syst. 252 (2022) 109333, http://dx.doi.org/10.1016/j.knosys.2022.109333.

[59] G. Liang, J. Wen, W. Zhou, Individual diversity preference aware neural collaborative filtering, Knowl.-Based Syst. 258 (2022) 109730, http://dx.doi.org/10.1016/j.knosys.2022.109730.

[60] L. Wang, X. Zhang, R. Wang, C. Yan, H. Kou, L. Qi, Diversified service recommendation with high accuracy and efficiency, Knowl.-Based Syst. 204 (2020) 106196, http://dx.doi.org/10.1016/j.knosys.2020.106196.

[61] Q. Liu, A.H. Reiner, A. Frigessi, I. Scheel, Diverse personalized recommendations with uncertainty from implicit preference data with the Bayesian Mallows model, Knowl.-Based Syst. 186 (2019) 104960, http://dx.doi.org/10.1016/j.knosys.2019.104960.

[62] M.D. Ekstrand, M. Tian, I.M. Azpiazu, J.D. Ekstrand, O. Anuyah, D. McNeill, M.S. Pera, All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness, in: Conference on Fairness, Accountability and Transparency, PMLR, 2018, pp. 172–186.

[63] J. Kleinberg, J. Ludwig, S. Mullainathan, A. Rambachan, Algorithmic fairness, in: Aea Papers and Proceedings, Vol. 108, 2018, pp. 22–27, http://dx.doi.org/10.1257/pandp.20181018.

[64] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. Zemel, Fairness through awareness, in: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 2012, pp. 214–226, http://dx.doi.org/10.1145/2090236.2090255.

[65] M. Hardt, E. Price, N. Srebro, Equality of opportunity in supervised learning, Adv. Neural Inf. Process. Syst. 29 (2016) 3315–3323.

[66] Z. Jiang, X. Han, C. Fan, F. Yang, A. Mostafavi, X. Hu, Generalized demographic parity for group fairness, in: International Conference on Learning Representations, 2021.

[67] M. Zehlike, K. Yang, J. Stoyanovich, Fairness in ranking, Part I: Score-based ranking, ACM Comput. Surv. 55 (6) (2022) 1–36, http://dx.doi.org/10.1145/3533379.

[68] M. Zehlike, K. Yang, J. Stoyanovich, Fairness in ranking, Part II: Learning-to-rank and recommender systems, ACM Comput. Surv. 55 (6) (2022) 1–41, http://dx.doi.org/10.1145/3533380.

[69] R. Burke, N. Sonboli, A. Ordonez-Gauger, Balanced neighborhoods for multi-sided fairness in recommendation, in: Conference on Fairness, Accountability and Transparency, PMLR, 2018, pp. 202–214, http://dx.doi.org/10.18122/b2gq53.

[70] B. Xia, J. Yin, J. Xu, Y. Li, WE-Rec: A fairness-aware reciprocal recommendation based on Walrasian equilibrium, Knowl.-Based Syst. 182 (2019) 104857, http://dx.doi.org/10.1016/j.knosys.2019.07.028.

[71] C. Castillo, Fairness and transparency in ranking, in: ACM SIGIR Forum, Vol. 52, ACM, New York, NY, USA, 2019, pp. 64–71, http://dx.doi.org/10.1145/3308774.3308783.

[72] G.K. Patro, L. Porcaro, L. Mitchell, Q. Zhang, M. Zehlike, N. Garg, Fair ranking: a critical review, challenges, and future directions, in: 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, pp. 1929–1942, http://dx.doi.org/10.1145/3531146.3533238.

[73] S. Yao, B. Huang, Beyond parity: Fairness objectives for collaborative filtering, Adv. Neural Inf. Process. Syst. 30 (2017).

[74] K. Yang, J. Stoyanovich, Measuring fairness in ranked outputs, in: Proceedings of the 29th International Conference on Scientific and Statistical Database Management, 2017, pp. 1–6, http://dx.doi.org/10.1145/3085504.3085526.

[75] A. Singh, T. Joachims, Fairness of exposure in rankings, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 2219–2228, http://dx.doi.org/10.1145/3219819.3220088.

[76] A.J. Biega, K.P. Gummadi, G. Weikum, Equity of attention: Amortizing individual fairness in rankings, in: The 41st International Acm Sigir Conference on Research & Development in Information Retrieval, 2018, pp. 405–414, http://dx.doi.org/10.1145/3209978.3210063.

[77] D. Serbos, S. Qi, N. Mamoulis, E. Pitoura, P. Tsaparas, Fairness in package-to-group recommendations, in: Proceedings of the 26th International Conference on World Wide Web, 2017, pp. 371–379, http://dx.doi.org/10.1145/3038912.3052612.

[78] L. Xiao, Z. Min, Z. Yongfeng, G. Zhaoquan, L. Yiqun, M. Shaoping, Fairness-aware group recommendation with pareto-efficiency, in: Proceedings of the Eleventh ACM Conference on Recommender Systems, 2017, pp. 107–115, http://dx.doi.org/10.1145/3109859.3109887.

[79] Y. Xiao, Q. Pei, L. Yao, S. Yu, L. Bai, X. Wang, An enhanced probabilistic fairness-aware group recommendation by incorporating social activeness, J. Netw. Comput. Appl. 156 (2020) 102579, http://dx.doi.org/10.1016/j.jnca.2020.102579.

[80] Y. Xiao, Q. Pei, T. Xiao, L. Yao, H. Liu, MutualRec: Joint friend and item recommendations with mutualistic attentional graph neural networks, J. Netw. Comput. Appl. 177 (2021) 102954, http://dx.doi.org/10.1016/j.jnca.2020.102954.

[81] R.B. Nozari, H. Koohi, A novel group recommender system based on members' influence and leader impact, Knowl.-Based Syst. 205 (2020) 106296, http://dx.doi.org/10.1016/j.knosys.2020.106296.

[82] H. Liu, N. Zhao, X. Zhang, H. Lin, L. Yang, B. Xu, Y. Lin, W. Fan, Dual constraints and adversarial learning for fair recommenders, Knowl.-Based Syst. 239 (2022) 108058, http://dx.doi.org/10.1016/j.knosys.2021.108058.

[83] H. Liu, H. Lin, B. Xu, N. Zhao, D. Wen, X. Zhang, Y. Lin, Perceived individual fairness with a molecular representation for medicine recommendations, Knowl.-Based Syst. 247 (2022) 108755, http://dx.doi.org/10.1016/j.knosys.2022.108755.

[84] J. Bobadilla, F. Ortega, A. Hernando, J. Bernal, A collaborative filtering approach to mitigate the new user cold start problem, Knowl.-Based Syst. 26 (2012) 225–238, http://dx.doi.org/10.1016/j.knosys.2011.07.021.

[85] Y. Wei, X. Wang, Q. Li, L. Nie, Y. Li, X. Li, T.-S. Chua, Contrastive learning for cold-start recommendation, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 5382–5390, http://dx.doi.org/10.1145/3474085.3475665.

[86] N.M. Villegas, C. Sánchez, J. Díaz-Cely, G. Tamura, Characterizing context-aware recommender systems: A systematic literature review, Knowl.-Based Syst. 140 (2018) 173–200, http://dx.doi.org/10.1016/j.knosys.2017.11.003.

[87] J. Li, C. Chen, H. Chen, C. Tong, Towards context-aware social recommendation via individual trust, Knowl.-Based Syst. 127 (2017) 58–66, http://dx.doi.org/10.1016/j.knosys.2017.02.032.

[88] X. Li, C.-H. Chen, P. Zheng, Z. Jiang, L. Wang, A context-aware diversity-oriented knowledge recommendation approach for smart engineering solution design, Knowl.-Based Syst. 215 (2021) 106739, http://dx.doi.org/10.1016/j.knosys.2021.106739.

[89] C. Musto, P. Lops, M. de Gemmis, G. Semeraro, Context-aware graph-based recommendations exploiting personalized PageRank, Knowl.-Based Syst. 216 (2021) 106806, http://dx.doi.org/10.1016/j.knosys.2021.106806.

[90] S. Dhelim, N. Aung, M.A. Bouras, H. Ning, E. Cambria, A survey on personality-aware recommendation systems, Artif. Intell. Rev. (2022) 1–46, http://dx.doi.org/10.1007/s10462-021-10063-7.

[91] M. Tkalcic, L. Chen, Personality and recommender systems, in: Recommender Systems Handbook, Springer, 2015, pp. 715–739, http://dx.doi.org/10.1007/978-1-4899-7637-6_21.

[92] O.P. John, S. Srivastava, et al., The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives, Vol. 2, University of California Berkeley, 1999.

[93] M. Tkalcic, M. Kunaver, A. Košir, J. Tasic, Addressing the new user problem with a personality based user similarity measure, in: First International Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems, Vol. 106, DEMRA 2011, 2011, pp. 106–123.

[94] R. Hu, P. Pu, Using personality information in collaborative filtering for new users, in: Recommender Systems and the Social Web, Vol. 17, 2010.

[95] M. Elahi, M. Braunhofer, F. Ricci, M. Tkalcic, Personality-based active learning for collaborative filtering recommender systems, in: Congress of the Italian Association for Artificial Intelligence, Springer, 2013, pp. 360–371, http://dx.doi.org/10.1007/978-3-319-03524-6_31.

[96] L. Chen, W. Wu, L. He, How personality influences users' needs for recommendation diversity? in: CHI'13 Extended Abstracts on Human Factors in Computing Systems, 2013, pp. 829–834, http://dx.doi.org/10.1145/2468356.2468505.

[97] W. Wu, L. Chen, L. He, Using personality to adjust diversity in recommender systems, in: Proceedings of the 24th ACM Conference on Hypertext and Social Media, 2013, pp. 225–229, http://dx.doi.org/10.1145/2481492.2481521.

[98] H.-C. Yang, Z.-R. Huang, Mining personality traits from social messages for game recommender systems, Knowl.-Based Syst. 165 (2019) 157–168, http://dx.doi.org/10.1016/j.knosys.2018.11.025.

[99] H. Wang, Y. Zuo, H. Li, J. Wu, Cross-domain recommendation with user personality, Knowl.-Based Syst. 213 (2021) 106664, http://dx.doi.org/10.1016/j.knosys.2020.106664.

[100] P.T. Costa, R.R. McCrae, The NEO Personality Inventory, Psychological Assessment Resources Odessa, FL, 1985.

[101] L.R. Goldberg, The structure of phenotypic personality traits, Am. Psychol. 48 (1) (1993) 26, http://dx.doi.org/10.1037/0003-066x.48.1.26.

[102] L.R. Goldberg, J.A. Johnson, H.W. Eber, R. Hogan, M.C. Ashton, C.R. Cloninger, H.G. Gough, The international personality item pool and the future of public-domain personality measures, J. Res. Personal. 40 (1) (2006) 84–96, http://dx.doi.org/10.1016/j.jrp.2005.08.007.

[103] Y. Mehta, N. Majumder, A. Gelbukh, E. Cambria, Recent trends in deep learning based personality detection, Artif. Intell. Rev. (2019) 1–27, http://dx.doi.org/10.1007/s10462-019-09770-z.

[104] M. Kosinski, D. Stillwell, T. Graepel, Private traits and attributes are predictable from digital records of human behavior, Proc. Natl. Acad. Sci. 110 (15) (2013) 5802–5805, http://dx.doi.org/10.1073/pnas.1218772110.

[105] S. Adali, J. Golbeck, Predicting personality with social behavior, in: 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, IEEE, 2012, pp. 302–309, http://dx.doi.org/10.1109/ASONAM.2012.58.

[106] Y. Amichai-Hamburger, G. Vinitzky, Social network use and personality, Comput. Hum. Behav. 26 (6) (2010) 1289–1295, http://dx.doi.org/10.1016/j.chb.2010.03.018.

[107] D. Garcia, S. Sikström, The dark side of Facebook: Semantic representations of status updates predict the Dark Triad of personality, Pers. Individ. Differ. 67 (2014) 92–96, http://dx.doi.org/10.1016/j.paid.2013.10.001.

[108] T.C. Marshall, K. Lefringhausen, N. Ferenczi, The Big Five, self-esteem, and narcissism as predictors of the topics people write about in Facebook status updates, Pers. Individ. Differ. 85 (2015) 35–40, http://dx.doi.org/10.1016/j.paid.2015.04.039.

[109] T. Yarkoni, Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers, J. Res. Personal. 44 (3) (2010) 363–373, http://dx.doi.org/10.1016/j.jrp.2010.04.001.

[110] S. Winter, G. Neubaum, S.C. Eimler, V. Gordon, J. Theil, J. Herrmann, J. Meinert, N.C. Krämer, Another brick in the Facebook wall–How personality traits relate to the content of status updates, Comput. Hum. Behav. 34 (2014) 194–202, http://dx.doi.org/10.1016/j.chb.2014.01.048.

[111] C. Sumner, A. Byers, R. Boochever, G.J. Park, Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets, in: 2012 11th International Conference on Machine Learning and Applications, Vol. 2, IEEE, 2012, pp. 386–393, http://dx.doi.org/10.1109/ICMLA.2012.218.

[112] W. Youyou, M. Kosinski, D. Stillwell, Computer-based personality judgments are more accurate than those made by humans, Proc. Natl. Acad. Sci. 112 (4) (2015) 1036–1040, http://dx.doi.org/10.1073/pnas.1418680112.

[113] A. Eftekhar, C. Fullwood, N. Morris, Capturing personality from Facebook photos and photo-related activities: How much exposure do you need? Comput. Hum. Behav. 37 (2014) 162–170, http://dx.doi.org/10.1016/j.chb.2014.04.048.

[114] S. Han, H. Huang, Y. Tang, Knowledge of words: An interpretable approach for personality recognition from social media, Knowl.-Based Syst. 194 (2020) 105550, http://dx.doi.org/10.1016/j.knosys.2020.105550.

[115] K. Biswas, P. Shivakumara, U. Pal, T. Chakraborti, T. Lu, M.N.B. Ayub, Fuzzy and genetic algorithm based approach for classification of personality traits oriented social media images, Knowl.-Based Syst. 241 (2022) 108024, http://dx.doi.org/10.1016/j.knosys.2021.108024.

[116] C. Suman, S. Saha, A. Gupta, S.K. Pandey, P. Bhattacharyya, A multi-modal personality prediction system, Knowl.-Based Syst. 236 (2022) 107715, http://dx.doi.org/10.1016/j.knosys.2021.107715.

[117] J. Lin, W. Mao, D.D. Zeng, Personality-based refinement for sentiment classification in microblog, Knowl.-Based Syst. 132 (2017) 204–214, http://dx.doi.org/10.1016/j.knosys.2017.06.031.

[118] S. Dhelim, N. Aung, H. Ning, Mining user interest based on personality-aware hybrid filtering in social networks, Knowl.-Based Syst. 206 (2020) 106227, http://dx.doi.org/10.1016/j.knosys.2020.106227.

[119] A.P. Association, et al., Diagnostic and Statistical Manual of Mental Disorders (DSM-5®), American Psychiatric Pub, 2013, http://dx.doi.org/10.1176/appi.books.

[120] M. Monaro, A. Toncini, S. Ferracuti, G. Tessari, M.G. Vaccaro, P. De Fazio, G. Pigato, T. Meneghel, C. Scarpazza, G. Sartori, The detection of malingering: a new tool to identify made-up depression, Front. Psychiatry 9 (2018) 249, http://dx.doi.org/10.3389/fpsyt.2018.00249.

[121] L. Eraslan, A. Kukuoglu, Social relations in virtual world and social media aggression, World J. Educ. Technol. Curr. Issues 11 (2) (2019) 1–11, http://dx.doi.org/10.18844/wjet.v11i2.4145.
[122] F. Mishna, C. Regehr, A. Lacombe-Duncan, J. Daciuk, G. Fearing, M. Van Wert, Social media, cyber-aggression and student mental health on a university campus, J. Ment. Health 27 (3) (2018) 222–229, http://dx.doi.org/10.1080/09638237.2018.1437607.
[123] M. Appel, C. Marker, T. Gnambs, Are social media ruining our lives? A review of meta-analytic evidence, Rev. Gen. Psychol. 24 (1) (2020) 60–74, http://dx.doi.org/10.1177/1089268019880891.
[124] L.M. Saulsman, A.C. Page, The five-factor model and personality disorder empirical literature: A meta-analytic review, Clin. Psychol. Rev. 23 (8) (2004) 1055–1085, http://dx.doi.org/10.1016/j.cpr.2002.09.001.
[125] C. Hakulinen, M. Elovainio, L. Pulkki-Råback, M. Virtanen, M. Kivimäki, M. Jokela, Personality and depressive symptoms: Individual participant meta-analysis of 10 cohort studies, Depress. Anxiety 32 (7) (2015) 461–470, http://dx.doi.org/10.1002/da.22376.
[126] K.S. Douglas, G.M. Vincent, J.F. Edens, Risk for Criminal Recidivism: The Role of Psychopathy, The Guilford Press, 2018.
[127] E. Mitsopoulou, T. Giovazolias, Personality traits, empathy and bullying behavior: A meta-analytic approach, Aggress. Viol. Behav. 21 (2015) 61–72, http://dx.doi.org/10.1016/j.avb.2015.01.007.
[128] P.C. Heaven, Personality and self-reported delinquency: Analysis of the "Big Five" personality dimensions, Pers. Individ. Differ. 20 (1) (1996) 47–54, http://dx.doi.org/10.1016/0191-8869(95)00136-T.
[129] K.A. Gleason, L.A. Jensen-Campbell, D. South Richardson, Agreeableness as a predictor of aggression in adolescence, Aggress. Behav. Official J. Int. Soc. Res. Aggress. 30 (1) (2004) 43–61, http://dx.doi.org/10.1002/ab.20002.
[130] J. Sharpe, S. Desai, The revised Neo Personality Inventory and the MMPI-2 Psychopathology Five in the prediction of aggression, Pers. Individ. Differ. 31 (4) (2001) 505–518, http://dx.doi.org/10.1016/S0191-8869(00)00155-0.
[131] K.W. Reardon, J.L. Tackett, D. Lynam, The personality context of relational aggression: A Five-Factor Model profile analysis, Personal. Disord. Theory Res. Treat. 9 (3) (2018) 228, http://dx.doi.org/10.1037/per0000231.
[132] J. Carvalho, P.J. Nobre, Five-factor model of personality and sexual aggression, Int. J. Offender Therap. Comp. Criminol. 63 (5) (2019) 797–814, http://dx.doi.org/10.1177/0306624X13481941.
[133] T. Donkers, J. Ziegler, The dual echo chamber: Modeling social media polarization for interventional recommending, in: Proceedings of the 15th ACM Conference on Recommender Systems, 2021, pp. 12–22, http://dx.doi.org/10.1145/3460231.3474261.
[134] R. Jiang, S. Chiappa, T. Lattimore, A. György, P. Kohli, Degenerate feedback loops in recommender systems, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 2019, pp. 383–390, http://dx.doi.org/10.1145/3306618.3314288.
[135] D. Franco, N. Navarin, M. Donini, D. Anguita, L. Oneto, Deep fair models for complex data: Graphs labeling and explainable face recognition, Neurocomputing 470 (2022) 318–334, http://dx.doi.org/10.1016/j.neucom.2021.05.109.
[136] C.D. Manning, Introduction to Information Retrieval, Syngress Publishing, 2008.
[137] V. Moscato, A. Picariello, G. Sperli, An emotional recommender system for music, IEEE Intell. Syst. 36 (5) (2020) 57–68, http://dx.doi.org/10.1109/MIS.2020.3026000.
[138] M. Polignano, F. Narducci, M. de Gemmis, G. Semeraro, Towards emotion-aware recommender systems: an affective coherence model based on emotion-driven behaviors, Expert Syst. Appl. 170 (2021) 114382, http://dx.doi.org/10.1016/j.eswa.2020.114382.
[139] M.B. Donnellan, F.L. Oswald, B.M. Baird, R.E. Lucas, The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality, Psychol. Assess. 18 (2) (2006) 192, http://dx.doi.org/10.1037/1040-3590.18.2.192.
[140] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[141] K. Järvelin, J. Kekäläinen, IR evaluation methods for retrieving highly relevant documents, in: ACM SIGIR Forum, Vol. 51, No. 2, ACM, New York, NY, USA, 2017, pp. 243–250, http://dx.doi.org/10.1145/3130348.3130374.
[142] Y. Liu, J. Wang, Y. Jiang, PT-LDA: A latent variable model to predict personality traits of social network users, Neurocomputing 210 (2016) 155–163, http://dx.doi.org/10.1016/j.neucom.2015.10.144.
[143] M. Polato, Rectorch: Pytorch-Based Framework for Top-N Recommendation, Zenodo, 2020, http://dx.doi.org/10.5281/zenodo.3841898.
[144] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[145] T. Hastie, R. Tibshirani, J.H. Friedman, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Vol. 2, Springer, 2009, http://dx.doi.org/10.1007/978-0-387-21606-5.
[146] T.T. Nguyen, F. Maxwell Harper, L. Terveen, J.A. Konstan, User personality and user satisfaction with recommender systems, Inform. Syst. Front. 20 (2018) 1173–1189, http://dx.doi.org/10.1007/s10796-017-9782-y.
[147] G. Roffo, A. Vinciarelli, Personality in computational advertising: A benchmark, 2016.
[148] S.C. Guntuku, S. Roy, L. Weisi, Personality modeling based image recommendation, in: MultiMedia Modeling: 21st International Conference, MMM 2015, Sydney, NSW, Australia, January 5-7, 2015, Proceedings, Part II 21, Springer, 2015, pp. 171–182, http://dx.doi.org/10.1007/978-3-319-14442-9_15.