Privacy-Preserving Collaborative Filtering: A New Approach based on Variable-Group-Size Microaggregation

Fran Casino[a], Constantinos Patsakis[a], Agusti Solanas[b]

[a]*Department of Informatics, University Piraeus, Piraeus, Greece*
[b]*Smart Technologies Research Group. Dept. of Computer Engineering and Mathematics. Universitat Rovira i Virgili. Catalonia. Spain*

Abstract

The amount of available information is growing steadily and, as a result, Internet users benefit from recommender systems, which help them find information, services and products that best fit their needs. A common technique used in recommender systems is Collaborative Filtering, which is based on users' collaboration to make recommendations. However, users are getting more concerned about their privacy and can be reluctant to disclose their interests and other personal information. Hence, with the aim to foster users collaboration, the development of privacy-aware collaborative filtering methods has become a hot topic in the field.

In this article we recall the concept of Privacy-Preserving Collaborative Filtering (PPCF) and introduce a novel approach based on variable-group-size microaggregation, which provides k-anonymity to the users. Also, we introduce several new metrics based on users' behaviour that overcome the drawbacks of traditional metrics. Extensive experiments show that our approach can provide more accurate recommendations than well-known methods while, at the same time, preserving users' privacy.

Keywords: Privacy, Collaborative Filtering, Recommender Systems, Variable-Group-Size Microaggregation, Behavioural Metrics

Email addresses: francasino@unipi.gr (Fran Casino), kpatsak@unipi.gr (Constantinos Patsakis), agusti.solanas@urv.cat (Agusti Solanas)

1. Introduction

Nowadays, recommender systems (Schafer et al., 2007) play an active role on the Internet through the advances in data mining and artificial intelligence. Furthermore, the potential and opportunities of the semantic web (*i.e. Web 3.0*), which enforces the user – computer interaction, have radically changed the way we use recommender systems. Therefore, recommender systems are now integrated into systems that we use in our everyday life transparently and efficiently. Recommendations based on the Internet are especially relevant for certain types of industries, such as e-commerce, which is one of the fastest growing industry worldwide (Han and Kim, 2019; Chou, 2019). In this regard, RS are a useful alternative to search engines since they help users to discover items they might not have found by themselves. Internet users are increasing their participation in multiple contexts such as the gastronomic field (*e.g.* with sites like Tripadvisor or Foursquare), the e-commerce field (*e.g.* Amazon, eBay, AliExpress), or in the audiovisual context, with sites like Netflix or IMDb. It is worth to note that nowadays, recommendations from personal acquaintances and online consumer's evaluations are the most trusted forms of advertising (The Nielsen Company, 2009; Mackinnon, 2012).

Collaborative Filtering (CF) (Goldberg et al., 1992) is a family of techniques introduced to provide automatic recommendations in a digital environment. The main idea behind of CF is to suggest/recommend items (*e.g.* books, movies or routes), based on the preferences of users that have already acquired and/or rated these items. These relationships between users and items are stored in the form of matrices. More concretely, data take the form of $n \times m$ matrices (i.e. n users and m items) and each cell (i, j) stores the evaluation of user i on item j. Moreover, binary rating data schemes which store whether an item has been rated/bought or not are also considered.

Therefore, recommendations provided by CF methods are based on the assumption that similar users would be interested in the same products. Therefore, items liked by user u_a can be recommended to user u_b, if u_a and u_b exhibit a high degree of similarity according to some metric.

The literature proposes a classification of CF methods into three main categories depending on the data they use(Su and Khoshgoftaar, 2009; Shi et al., 2014): (i) *memory-based methods*, which use the raw data matrix with all entries, ratings and relationships; (ii) *model-based methods*, which use data mining, data modelling and statistical techniques that receive the data

matrix as input; and (iii) *hybrid methods*, which may combine the previous methods with other information sources (Shi et al., 2014).

CF methods have several inherent limitations including *sparseness*, *scalability*,*shilling*, *cold start*, *synonymy*, *bribing*, *copy-profile attacks*, and the lack of privacy (Su and Khoshgoftaar, 2009; Shi et al., 2014; Casino et al., 2013b; Gunes et al., 2014). For more on CF, we point the interested reader to (Shi et al., 2014; Kluver et al., 2018; Bobadilla et al., 2013) for a review of the state-of-the-art and the most relevant advances and trends.

CF systems and the quality of their recommendations rely on the collection of private behavioural information in a wide variety of contexts (*e.g.* places to go, things to do, or products to buy) (Shi et al., 2014). While such information collection provides great opportunities and benefits for both companies and users, the users end up losing their privacy (Jeckmans et al., 2013). Moreover, careless management of personal information, besides being illegal, could lead to severe consequences for both users, whose information is stored and processed, as well as companies, whose role is to process and use it responsively according to users' preferences. In fact, the recent introduction of the EU General Data Protection Regulation (GDPR) tries to regulate the data collection and processing management and provide more control to the users (General Data Protection Regulation; Politou et al., 2018). However, companies risk something more than just being fined for improper storing and processing user information. They risk their reputation for delivering their core services accurately.

Even if someone disregards the privacy of individuals, he would definitely appreciate the advantages of Privacy-Preserving Collaborative Filtering (PPCF) methods in the Internet industry. For instance, the existence of large monopolies on the Internet (*e.g.* Google, Amazon) enables the sharing of data between different entities managed by such companies, without the user's awareness or consent. Therefore, the lack of privacy could result in massive data exploitation, orchestrated by companies which could collect the profiles of many users in a given market sector, getting a huge advantage over their competitors. Undoubtedly, the fusion of multiple sources of users' data could lead to better recommendations; it could also lead to harmful consequences if non-trusted parties access the data.

Despite the extensive range and amount of data available on the Internet, people are reluctant to disclose personal information and their interests, which makes privacy one of the most relevant problems. Thus, users' privacy concerns (*e.g.* the disclosure of their profile information and the transfer of

personal data to third parties) affect their behaviour (Cranor et al., 1999) resulting in a reduction of both the number of given assessments as well as their quality. Therefore, two relevant aspects need to be taken into account (Friedman et al., 2015): (i) the privacy of the data; and (ii) the users' perception and reasoning about privacy. To address such privacy issues, current research focuses on Privacy-Preserving Collaborative Filtering (PPCF) methods.

1.1. Main contributions

As already discussed, one of the main limitations and risks of CF is the lack of users' privacy. Thus, amongst the open problems in CF (Shi et al., 2014; Casino et al., 2013b), in this article, we concentrate on the protection of the privacy of the users involved in the CF processes. We present a novel method for Privacy-Preserving Collaborative Filtering (PPCF) based on variable-sized group microaggregation which guarantees k-anonymity. Moreover, we show that our method provides both more privacy protection and quality of recommendations than other well-known methods such as the Alternating Least Squares (ALS) method (Koren, 2010), the non-negative matrix factorization approach (NMF) (Luo et al., 2014), the Regularized Singular Value Decomposition method (RSVD), other clustering approaches such as Co-clustering (George and Merugu, 2005), previous microaggregation approaches (Casino et al., 2013a, 2015) which are based on MDAV which is a fixed-size microaggregation method and, by extension, than the widely used Gaussian noise addition (GNA) method (Polat, 2003). Note that despite its used in numerous contexts, to the best of our knowledge, this is the first time that V-MDAV is used to protect the privacy of the users in a recommender system context. The experiments are performed with two well-known datasets which have different characteristics in both data sparsity and number of dimensions. In addition, we conduct a novel data analysis that includes a comparison of the quality of the recommendations before and after data obfuscation for both datasets. Next, we provide a thorough analysis of the privacy protection provided by the privacy-preserving microaggregation approaches using well-known metrics as well as a novel efficiency score, which measures the trade-off between privacy and recommendation's accuracy. Finally, we provide an extensive analysis of the recommendations by using a novel behavioural-based metric which succeeds in providing more information about the quality of the recommendations than well-known metrics such as precision and recall and the mean absolute error. These new metrics over-

come some of the weaknesses and limitations of current metrics and enable a more in-depth analysis of the recommendations.

1.2. Plan of the article

The rest of this article is organised as follows. In Section 2, we recall the basic concepts of PPCF and discuss the most relevant methods. Moreover, in Section 2.1 we provide a background on Statistical Disclosure Control (SDC) and microaggregation. Next, in Section 3, we describe our approach and the new metrics. In Section 4, we describe the benchmarks and the set of experiments performed to support our hypothesis, whose outcomes are thoroughly discussed in Section 5. Finally, Section 6 concludes the article and provides future research directions.

2. Related work

As previously stated in Section 1, one of the main requirements of recommender systems and CF is information. In this regard, privacy-preserving collaborative filtering methods are designed to solve the privacy issues raised by the systematic collection of private information. The relevance of such systems is justified by the numerous information leakage cases described in the literature. In (Minkus and Ross, 2014), the authors show how the history of purchases and opinions of user profiles in eBay can disclose their interests and the types of products that they have purchased. Moreover, serious privacy issues can be raised if the usernames and information about users are correlated with other sources. Another example is shown in (Hannak et al., 2014) where authors perform a study about how the e-commerce websites characterise users and personalise their searches to perform price discrimination (*i.e.* discounts for specific groups of users) or steering (*i.e.* reorder recommendations to influence searches or purchases). The creation of private and trustable communities, as well as profile obfuscation methods, are attractive solutions for privacy-concerned users. Nevertheless, similar user's communities (*e.g.* trust or social networks) may affect the recommendations, since the homogeneity of the groups may diminish the diversity and originality of recommendations (McPherson et al., 2001). Notwithstanding, privacy-preserving techniques should not be an obstacle in the Internet industry, since companies may share information to enhance their recommendations. Nevertheless, due to privacy and business concerns, unprotected user data should not be disclosed between companies.

In (Casino et al., 2013b), the authors proposed a classification of PPCF methods according to their organisational structure (i.e. centralised and decentralised). Typically, centralised methods exhibit higher efficiency than their decentralised counterparts since computations avoid several communication overheads. However, in centralised methods, data are managed by a single entity which has total control over them, with the implied privacy issues if data have a low protection level (Aggarwal, 2007; Huang et al., 2005). We can find examples of perturbation-based centralised PPCF in (Batmaz and Polat, 2016). Other works that include PPCF dimensionality reduction approaches can be found in (Bilge and Polat, 2013; Casino et al., 2015; Luo et al., 2014; Funk, 2006; George and Merugu, 2005; Wei et al., 2018).

On the contrary, decentralised methods use distributed networks of users to perform intermediate calculations and recommendations using their private data. Such schemes generally involve less information disclosure than their centralised counterparts, but they require the use of cryptographic protocols and more complex calculations, and also need the active participation of the users in most cases. Some examples with partitioned market basket databases (*i.e.* binary datasets which store whether an item is relevant to a user or not) (Polat and Du, 2008; Yakut and Polat, 2012) have been proposed in the literature. Randomised perturbation techniques have also been implemented using a distributed architecture, as showed in (Boutet et al., 2016). There are also several approaches with partitioned data schemes (Polat and Du, 2008; Okkalioglu et al., 2016). Finally, approaches in which users store their ratings can be found in (Canny and Canny, 2002; Berkovsky et al., 2007; Kaleli and Polat, 2010; Basu et al., 2011). Recently, differential privacy is an active research field (Li et al., 2017; Shin et al., 2018; Xian et al., 2017); however, the trade-off between efficiency and accuracy of such methods is still to be discussed. Interestingly enough, a survey on distributed and parallel techniques in CF can be found in (Karydi and Margaritis, 2016). Most of these approaches could be used to enhance the performance of PPCF approaches if data is privately distributed utilising secure protocols and cryptographic tools. Figure 1 shows a classification structure to provide a comprehensive overview of PPCF's state-of-the-art methods.

For more on PPCF, the interested reader may refer to (Casino et al., 2013b; Bilge et al., 2013; Batmaz and Kaleli, 2017).

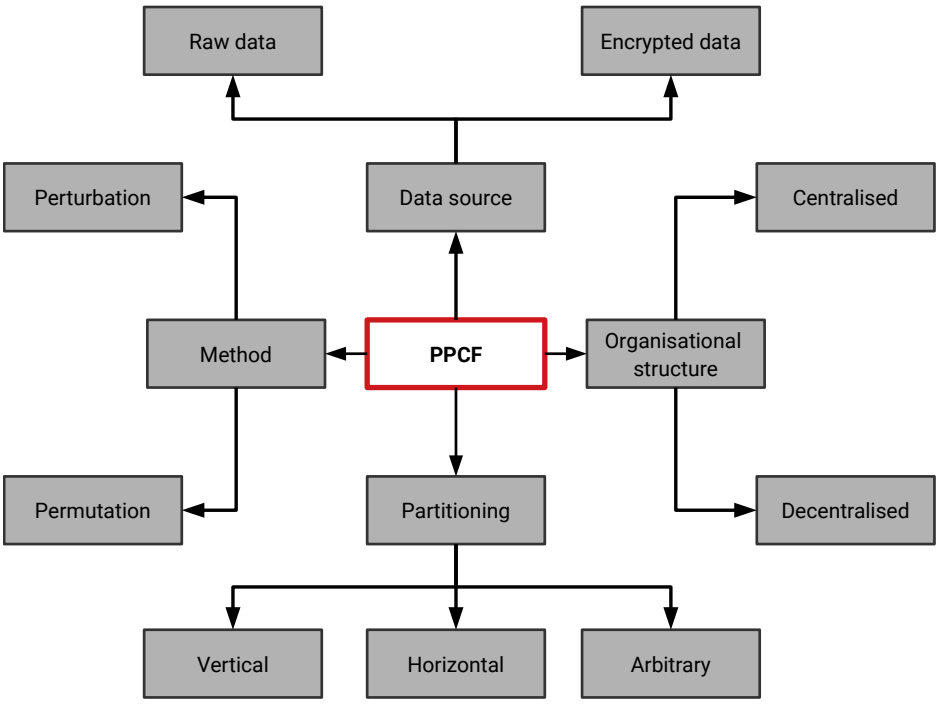


Figure 1: Classification of PPCF in terms of obfuscation methods, data partitioning schemes, organisational structures and data sources.

2.1. Background

The goal of Statistical Disclosure Control (SDC, (Hundepool et al., 2012)), also known as data anonymisation, is to transform microdata sets (*i.e.* datasets consisting of records corresponding to individual respondents). This transformation is performed prior to publication to prevent possible re-identification of the respondent corresponding to any particular record in the anonymised published microdata set —identity disclosure— and to prevent the disclosure of the value of any confidential attribute (*e.g.* salary) for a *specific* respondent —attribute disclosure.

Before any anonymisation process, direct identifiers (ID and passport number, phone number etc.) are removed from the dataset. Nevertheless, a subset of the attributes that remain in the anonymised dataset may be *quasi-identifiers*, that is, attributes which may enable indirect re-identification of respondents through their combination with external data sources (available as adversaries' background knowledge).

In fact, we consider the preferences of users on different items as at-

tributes, so that each record contains the preferences of a particular user. Therefore, we consider all attributes to be *quasi-identifiers* since a large number of preferences has been shown to lead to user re-identification (*e.g.* Netflix users were identified based on their preferences in (Narayanan and Shmatikov, 2008)).

Microaggregation algorithms work in two stages:

1. The set of records in a dataset is clustered in such a way that each cluster contains at least k records and records within a cluster are as similar as possible.
2. A representative of the cluster, typically the centroid record (*i.e.* the average of the cluster) replaces the records within each cluster.

When microaggregation is applied to a set of records, the resulting dataset is k-anonymous, that is, to an adversary, each record in the dataset is indistinguishable within a group of k records (i.e. belonging to the same cluster) in terms of the quasi-identifiers. One of the prevalent multivariate microaggregation methods is MDAV (Domingo-Ferrer and Mateo-Sanz, 2002; Casino et al., 2015). However, being a fixed-size heuristic, there are cases in which it yields a k-partition far from the optimal one (Domingo-Ferrer et al., 2006). Therefore, such heuristics lack the flexibility of adapting the group size to the distribution of the records in the dataset, which may result in poor within-group homogeneity. The Variable-size Maximum Distance to Average Vector (V-MDAV) (Solanas and Martinez-Balleste, 2006) algorithm was designed to overcome the limitations of the methods which compute groups of fixed-size cardinality (*e.g.* MDAV) by computing a variable size k-partition with a similar computational cost.

The V-MDAV approach follows a similar strategy to MDAV and its complexity is almost the same except for two main differences:

1. While MDAV computes a centroid in each iteration, V-MDAV only computes the dataset centroid at the beginning, enhancing its efficiency.
2. Contrary to V-MDAV, MDAV does not build a matrix of distances; on the contrary, it computes distances as many times as needed.

Moreover, once a group is formed (Algorithm 1, *line 6*) V-MDAV applies a heuristic (Algorithm 1, *line 7*) which allows it to fit the dataset distribution better by generating variable-sized groups. The expansion of the group is computed as follows:

Algorithm 1 Variable microaggregation algorithm

```
1: function V-MDAV(DataSet D, Integer k)
2:     ComputeDistancesMatrix(D);
3:     C = ComputeCentroid(D);
4:     while (RecordsToAssign> k − 1) do
5:         e = SelectMostDistantRecordToCentroid (D,C);
6:         g_i = BuildGroupFromRecord(e,D,k);
7:         g_i = ExtendGroup(g_i,D,k);
8:     end while
9:     g_1 ... g_s = RemainingUnassignedRecords(D,g_1 ... g_s);
10:    M = BuildMicroaggregatedDataSet(D,g_1 ... g_s);
11: return M                                  ▷ microaggregatedSet M
12: end function
```

Given a group g with p records, the record e_{min} among unassigned records outside g nearest to g and the minimum distance d_{in} between e_{min} and g are defined by Equation (1) and (2):

$$d_{in} = \min_{j \in [1, N_{un}]} d(e_i^g, e_j), \forall_i \in [1, p] \tag{1}$$

$$e_{min} = \arg \min_{j \in [1, N_{un}]} d(e_i^g, e_j), \forall_i \in [1, p] \tag{2}$$

where e_i^g denotes the i-th record in group g, e_j denotes the j-th record in the unassigned set of records and N_{un} is the number of unassigned records, that is, the number of records which not belong to any group. Thereafter, we randomly designate one of the records that satisfy Equation (2) as e_{min}. Next, we obtain the minimum distance d_{out} between e_{min} and any of the remaining unassigned records using the following equation:

$$d_{out} = \min_{j \in [1, N_{un}], e_{min} \neq e_j} [d(e_{min}, e_j)], \tag{3}$$

Finally, we compute the distance between e_{min} and g (denoted as d_{in}) and compare it with the distance between e_{min} and the closest unassigned neighbour (denoted as d_{out}) to decide on its inclusion into group g. Therefore, we include it if and only if $d_{in} < \gamma < d_{out}$, where γ is a *gain* factor that has to be tuned to improve the adaptability of V-MDAV. In the original approach (Solanas and Martinez-Balleste, 2006), γ was designed to generate

the groups that better fit the data distribution to minimise the information loss. However, in the V-MDAV inspired method presented in this paper, γ will act as a gain factor that will be tuned to optimise the trade-off between the accuracy of the recommendations and the level of privacy.

The extension process is repeated until the group size reaches $2k-1$ or $d_{in} < \gamma(d_{out})$, because as shown in (Domingo-Ferrer and Mateo-Sanz, 2002), in an optimal k-partition, each group consists of k to $2k-1$ records.

Similarly to MDAV, the proposed method may leave some records unassigned at the end of the main loop. Hence, the remaining records are assigned to their closest group (Algorithm 1, *line 9*). Finally, a microaggregated dataset M is generated from the resulting k-partition represented in Algorithm 1, *line 10*) as $(g_1 \dots g_s)$.

3. Our proposal

In what follows, we first detail our proposed recommendation system in Section 3.1. Next, we describe a new set of metrics with which provide a more detailed analysis of recommendations quality as well as obfuscation efficiency (i.e. the obfuscation needed to achieve a desired level of privacy) in Section 3.2.

3.1. Method description

The architecture of our method relies on a central server which computes the personalised recommendations according to the classical centralised architecture, as depicted in Figure 2. Following this scheme, our method is applied in the anonymisation server. In principle, we assume that users trust the anonymisation server; therefore, they use a secure channel to tunnel their data to it (*e.g.* a TLS connection). Nonetheless, users may partially trust the server regarding their privacy. The recent cases of, e.g. Facebook, have illustrated that big service providers may provide high quality services, yet they may arbitrarily share sensitive user data. Therefore, users may opt to route their preferences to the server through other users they trust, blinding him of who submitted the values by, e.g. using blind signatures (Chaum, 1983) to validate their submissions. For the sake of simplicity, we assume that the users have correctly submitted their data to the anonymisation server using their preferred method.

The general outline of our system is illustrated in Figure 3. First, we need to ensure that the dataset is not missing any value for any attribute

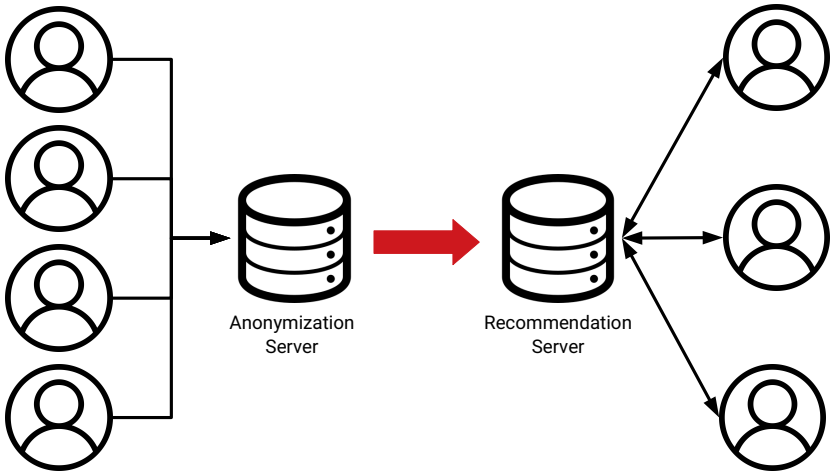


Figure 2: Centralised PPCF scheme. First, the information of users is collected and anonymised in a secure server. Next, the obfuscated information is stored in the recommendations server. Finally, users establish bidirectional communication with the server (*i.e.* query & recommendation). Note that the obfuscation is performed in the anonymisation server.

in its records. This is necessary to guarantee k-anonymity (i.e. unique values could lead to disclosure). We may use the central value or the overall mean, amongst others (Brick and Kalton, 1996; Dempster et al., 1977), to fill/impute the empty fields of the dataset matrix. Once the matrix is filled, we compute the z-scores of each column (item) of the dataset to standardise the data (*i.e.* give the same statistical weight to each dimension), using the following expression:

$$\text{z-score} = \frac{x_i - \mu}{\sigma} \tag{4}$$

where x_i is the i-th value of item x, μ and σ are the mean and the standard deviation of item x, respectively. Apparently, the mean and the standard deviation of the transformed items are 0 and 1, respectively.

Next, we apply the V-MDAV clustering algorithm, described in Section 2.1, to microaggregate the data. Afterwards, users are grouped into n clusters, with each cluster C_i containing the k most similar users. The cardinality of each group, denoted by k, is computed using a heuristic which is detailed in Section 2.1. Once the group relationships are established, the mean values of each C_i, denoted as M_i, are computed. Then, each value of C_i is replaced by the corresponding M_i. This way, the V-MDAV clustering pro-

cess results to a new dataset where members of the same cluster C_i have the same profiles, becoming indistinguishable within their group. Thus, after applying V-MDAV, this dataset satisfies, at least, k-anonymity. Finally, to compute recommendations, the results are de-standardised to obtain the final microaggregated dataset.

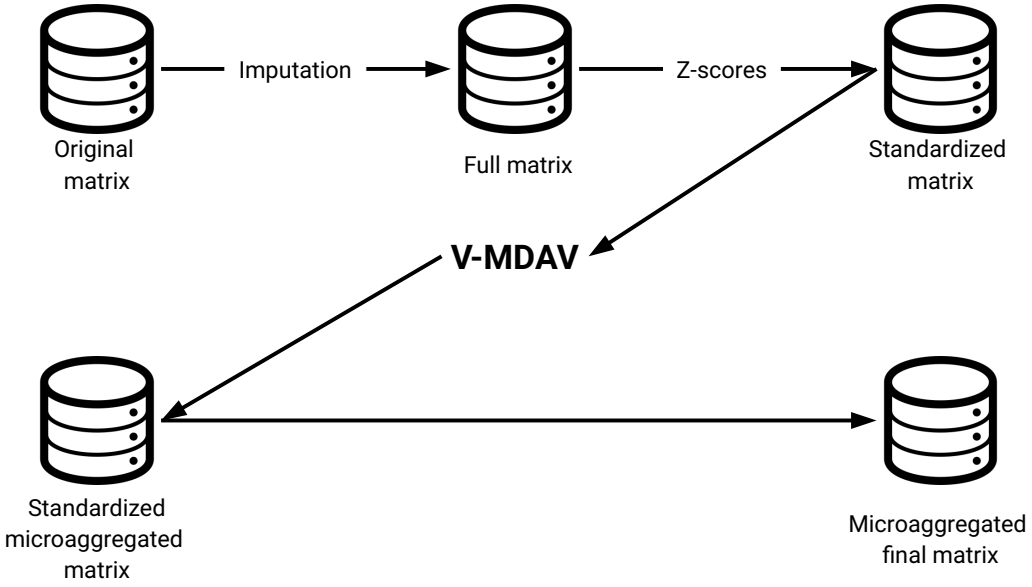


Figure 3: Overview of the steps performed in our proposal.

3.2. New metrics and recommendations analysis

Typically, the error between the original dataset values and the recommended values is measured using the mean absolute error (MAE), defined as:

$$MAE = \frac{\sum_{i=1}^{n} |p_i - r_i|}{n} \tag{5}$$

where n is the number of recommended elements, p_i is the recommended value over the element i, and r_i is the real value of i.

With regard to the recommendation quality measures, MAE provides an overall value which does not reflect the recommendation quality in an individualised and precise form. For instance, a low MAE could indicate

highly accurate or nearly perfect recommendations for many users while poor and practically useless recommendations for many others. Measures such as precision and recall focus on what percentage of recommended items (*i.e.* in a ranked list) are of interest to the user. Nevertheless, we propose the use of *behavioural precision* metrics to characterise the profile of the user taking into account both positive and negative assessments, hence obtaining information about the users' interests in a holistic manner. Therefore, we introduce two measures, namely *binary* precision and *four-level* (4L) precision, to gauge how recommendations fit the users' interests. To measure such behavioural precision, the value range of a given dataset is divided into two (*i.e.* binary) or into four (*i.e.* 4L) sub-ranges and we ascertain in which sub-range is the active user u_a's recommendation. Figure 4 describes the range level classification using a tree structure. Due to the particularity of our analysis procedure, which evaluates recommendations according to their sub-range (*i.e.* given an original value v, we first check if a recommendation r belongs to the same sub-range and then we evaluate the rest of sub-ranges, considering first the closest ones), we define all ranges as closed (cf. Figure 4).

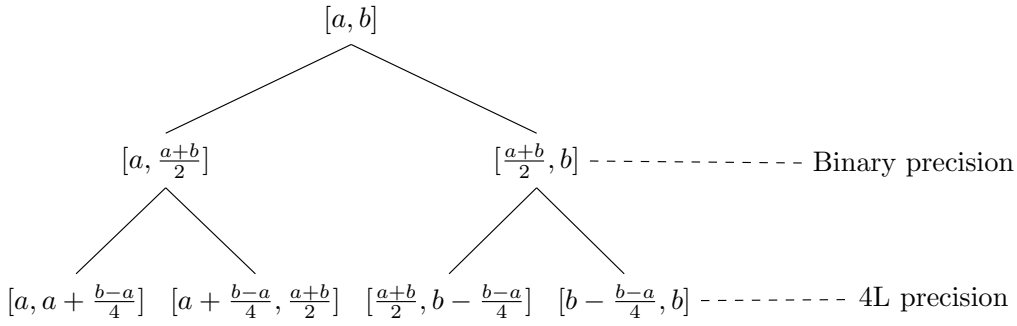


Figure 4: Level classification given a generic range of values $[a, b]$. Note that each level of the tree corresponds to a behavioural precision implementation.

Once the different sub-ranges are created, the next step is to compute their precision by comparing each original value v with the computed recommendation r. Tables 1 and 2 describe how behavioural precision is computed in the binary and the 4L cases, respectively.

Although we may divide the ratings range into n sub-ranges, we selected only two strategies, namely binary and 4L, for the sake of clarity. Note that the more sub-ranges, the more complicated it is to find users with the

r \ v	$[a, \frac{a+b}{2}]$	$[\frac{a+b}{2}, b]$
$[a, \frac{a+b}{2}]$	Match	Reversal
$[\frac{a+b}{2}, b]$	Reversal	Match

Table 1: Detail of binary behavioural precision computation. A reversal is a recommendation which is opposite of the user's behaviour.

r \ v	$[a, a + \frac{b-a}{4}]$	$[a + \frac{b-a}{4}, \frac{a+b}{2}]$	$[\frac{a+b}{2}, b - \frac{b-a}{4}]$	$[b - \frac{b-a}{4}, b]$
$[a, a + \frac{b-a}{4}]$	Match	Slight Match	Slight Reversal	Reversal
$[a + \frac{b-a}{4}, \frac{a+b}{2}]$	Slight Match	Match	Slight Match	Slight Reversal
$[\frac{a+b}{2}, b - \frac{b-a}{4}]$	Slight Reversal	Slight Match	Match	Slight Match
$[b - \frac{b-a}{4}, b]$	Reversal	Slight Reversal	Slight Match	Match

Table 2: Detail of 4L behavioural precision computation.

same behaviours, especially if we analyse sparse data. Moreover, users have different vote tendencies which may change according to the possible range of values (O'Mahony et al., 2006). Despite that, a methodology to select the proper number of sub-ranges considering the sparseness and the range of the ratings of a dataset is left for future research.

Other well-known metrics are precision and recall (Su and Khoshgoftaar, 2009; Phan et al., 2017). In this regard, we consider two top-N recommendation sets where $n = 10$ and $n = 20$. Note that the lower the n the more affected the recall measure is, especially when the total number of relevant items for a user is high compared to n. Therefore, precision at n is the proportion of correctly recommended items in the top-N set that are relevant:

$$Precision@n = \frac{\text{recommended items @n that are relevant}}{\text{recommended items @n}} \tag{6}$$

where relevant items are these that have a $rating \geq 3$.

Recall@n is the proportion of relevant items found in the top-n recommendations, that is:

$$Recall@n = \frac{\text{recommended items @n that are relevant}}{\text{number of relevant items}} \tag{7}$$

We also use the F1-score, which is the harmonic mean of the precision and the recall and is computed as follows:

$$F_1 = 2 * \frac{precision * recall}{precision + recall} \tag{8}$$

Typically, two factors are considered to measure the quality of the privacy provided by a perturbation method, namely the information loss and the disclosure risk. Notwithstanding, here we focus on the quality of the recommendations (i.e. MAE) since we need to measure the real data usability when computing recommendations, being far less significant the value of the information loss.

The disclosure risk (DR) measures the probability of correlation between a record of the obfuscated/protected data matrix with a record of the original dataset. DR is also known as re-identification probability or re-identification risk. From an attacker perspective, the re-identification procedure consists of computing the distances (*e.g.* the Euclidean distance) between a given protected record p_i (corresponding to user i), and the target records o_j that could be obtained from external sources. In our case, we assume the "oracle" scenario, which is the best scenario for an attacker, as he has access to both the original dataset O and the obfuscated dataset P, and he tries to link each record p_i in P with at least one record o_j in O.

For each record p_i in P the attacker finds the closest record o_j in O. If o_j is actually the original record belonging to p_i, the attacker succeeds, and we say that p_i has been re-identified. Therefore, DR represents the percentage of correct re-identifications.

Clearly, in terms of privacy and utility of the data, both the MAE and the DR should be low. To exploit the combination of these measures, we propose the use of a score, which is a well-known procedure in the SDC field (Domingo-Ferrer and Torra, 2001; Domingo-ferrer et al., 2001). Hence, we combine such measures in a single score, namely T-score, defined as:

$$\text{T-score} = \frac{MAE + DR}{2} \tag{9}$$

In this case, the lower the percentage, the better the obtained score. Therefore, we compare the score achieved by different approaches to determine their efficiency in terms of accuracy and privacy.

4. Experimental setup

In this section, we describe the set of strategies and tools used to perform our experiments. The benchmarks used in our comparisons are introduced

in Section 4.1, and we describe the tests and summarise their main characteristics in a comprehensive table in Section 4.2.

4.1. Benchmark datasets

We use two well-known CF datasets to perform the experiments. Grouplens developed the Movielens dataset (Resnick et al., 1994), and it is one of the prevalent CF datasets. Here, we focus on Movielens 100k, which contains 100,000 ratings (ranging from 1 to 5) of 943 users on 1,682 movies. This database is highly sparse since more than 90% of the fields are empty. Jester (Goldberg et al., 1999) is a joke recommendation system developed at the University of California, Berkeley. The database contains 100 jokes and ratings of 73,421 users. As a result, this dataset, once filled, contains a total of 7,342,100 values. However, Jester database is less sparse than Movielens 100k since approximately 44% of cells are empty.

The well-known "leave-one-out experiment", in which the values of the closest neighbour of the active user u_a are selected as the recommended values has been conducted to evaluate the accuracy of recommendations. The error between the original dataset values and the recommended values is computed using Equation (5).

Nevertheless, we have slightly modified the leave-one-out experiment to enable the analysis of raw databases. Since raw matrices contain non-rated items, the Euclidean distance between common-rated items is used to compute the closest neighbour of each user. Thus, we may need to select more than one neighbour to obtain the predictes values each time that a user is evaluated. Moreover, the evaluated user u_a might be the only one who rated a specific item. In such a case, we would not be able to provide a prediction for that item, since nobody else rated it. To address such issues, we use the central value as a recommendation. The problem above does not affect microaggregation methods (i.e. MDAV and V-MDAV) because the matrices are filled before data obfuscation. Note that we use the Euclidean distance in this work instead of other well-known similarity approaches such as Pearson Correlation because the data sparseness makes it difficult to find users with common assessments. The latter is known to affect correlation-based measurements (Schafer et al., 2007; da Silva et al., 2016), especially for computations over raw data.

Method	Description	Parameters
Co-Clustering	Compute a k partition of users and items and use their data to compute predictions	k_u and k_i
RSVD Koren (2010)	Iterative method for minimizing an objective similarity function	n_factors=100, n_epochs=20, init_std_dev=0.1
ALS Koren (2010)	Alternate optimization for the L2 norm	$\lambda_i = 5$, $\lambda_u = 12$, n_epochs=5
NMF Luo et al. (2014)	Non-negative low-rank matrix factorization	n_factors=15, n_epochs=50, $\lambda_u = 0.6$, $\lambda_i = 0.6$

Table 3: Main characteristics and features of the literature methods used in our comparison.

4.2. Test description

To showcase the efficacy of our proposal, we perform a set of experiments. First, we compare our approach with a set of well-known state-of-the-art methods: (i) the Alternating Least Squares (ALS) method, as described in (Koren, 2010), (ii) the non-negative matrix factorization approach (NMF) (Luo et al., 2014), (iii) the Regularized Singular Value Decomposition method (RSVD), proposed by Simon Funk for the famous Netflix Prize (Funk, 2006), (iv) the co-clustering method (Co-Clus), as described in (George and Merugu, 2005) and (v) the MDAV method (Casino et al., 2015). Table 3 summarises the main characteristics of such methods. We use a Python implementation (Hug, 2017) for RSVD, ALS and NMF. In the case of V-MDAV and MDAV, we follow a pre-processing step in which benchmark datasets are filled with the central value of their corresponding value range, to alleviate sparseness and to compute similarities accurately. Due to the possible configurations of the clustering approaches, we depict the best results of MDAV and V-MDAV for k=2 and k=3. In the case of V-MDAV, our optimisation strategy is set to obtain the lowest MAE, and thus, we test different values of γ between 0.1 and 3.0 and select the optimal one. In this case, we use a 5-fold test in Movielens 100k to compare the accuracy of the methods. We discuss the outcomes of this comparison in Section 5.1.

Second, we perform an extensive comparison between our proposal and the original MDAV approach (Casino et al., 2015). We analyse the data utility and privacy provided by our proposal and the MDAV method to ascertain which method achieves better outcomes. In this case, we evaluate our proposal with different values of γ (*i.e.* ranging from 0.1 to 3.0, being $\gamma = 0$ a value which indicates no extension of the groups and, hence, a fixed-size

cardinality V-MDAV) for every k, to generate the groups which better fit the data distribution according to a given optimisation strategy. Therefore, since our V-MDAV approach has been designed to optimise the trade-off between the accuracy and privacy, the best γ is selected according to the best T-score value. Both in this and the following tests, we also use Jester, so that we can study how data are affected considering fixed and variable cardinality groups in two different sets of data. The outcomes of such analysis are discussed in Section 5.2.

Third, in Section 5.3.1, we compute the quality of the recommendations using the datasets prior to data obfuscation (i.e. between raw and filled datasets). The aim is to study how data are later affected by V-MDAV. Finally, we assess the quality of the recommendations of the V-MDAV method in Section 5.3.2. Table 4 summarises the main characteristics of each test.

	Prediction comparison	**Top-N recommendation**	**MDAV & V-MDAV obfuscation efficiency**	**Raw data recommendations**	**V-MDAV recommendations**
Methods	RSVD, ALS, NMF Co-Clus, MDAV, V-MDAV	RSVD, ALS, NMF Co-Clus, MDAV, V-MDAV	MDAV, V-MDAV	Central value	V-MDAV
Benchmark datasets	ML 100k	ML 100k	ML 100k, Jester	ML 100k, Jester	ML 100k, Jester
Metrics	MAE, Behavioural	Precision@n, Recall@n, F1-score	MAE, DR, T-score	MAE, Behavioural	MAE, Behavioural
Optimisation strategy	MAE	T-score	T-score	-	T-score

Table 4: Summary of tests and their main characteristics. The optimisation strategy is only applicable to V-MDAV.

Note that, in all tests, the datasets are filled with the central value before using V-MDAV and MDAV (cf Section 3.1 and (Casino et al., 2015)). Moreover, the 5-fold tests are repeated 100 times, and the average of the outcomes is considered as the recommended value.

5. Discussion

5.1. Comparison with SOA methods

The MAE comparison between all methods is depicted in Table 5. We can observe that both microaggregation approaches can obfuscate data and provide k-anonymity while guaranteeing highly accurate predictions, outperforming well-known state of the art methods. Note that V-MDAV obtained the best MAE when $(k, \gamma) = (2, 0.6)$ and $(k, \gamma) = (3, 0.7)$. The minor differences between V-MDAV and MDAV are due to their design (i.e. MDAV

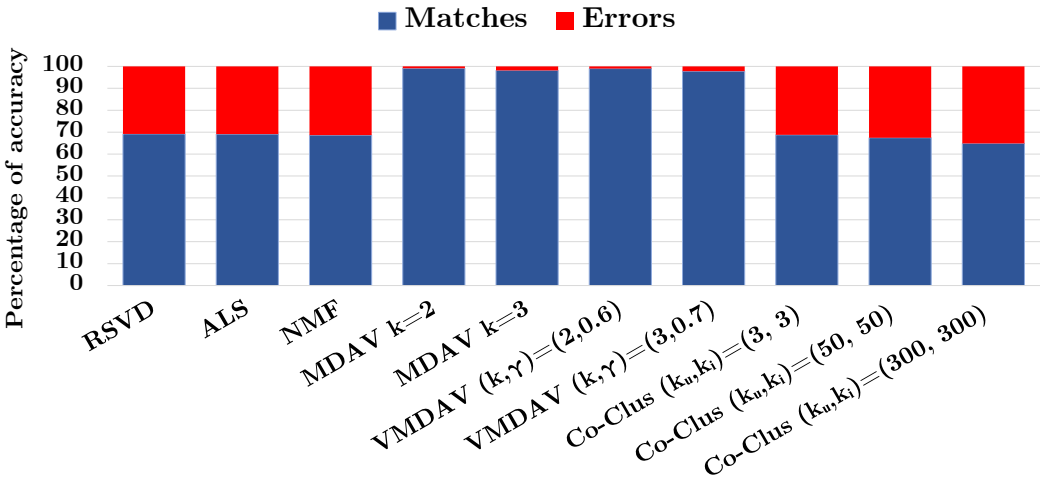
recomputes the centroid at each iteration, while V-MDAV does not), grouping similar profiles into possibly different clusters. See Section 2.1 for more details.

MAE%			
MDAV $k = 2$	11.71	V-MDAV $(k, \gamma) = (2, 0.6)$	11.78
MDAV $k = 3$	14.89	V-MDAV $(k, \gamma) = (3, 0.7)$	15.18
Co-Clus $(k_u, k_i) = (3, 3)$	18.88	RSVD	18.45
Co-Clus $(k_u, k_i) = (50, 50)$	20.13	NMF	18.92
Co-Clus $(k_u, k_i) = (300, 300)$	23.21	ALS	18.63

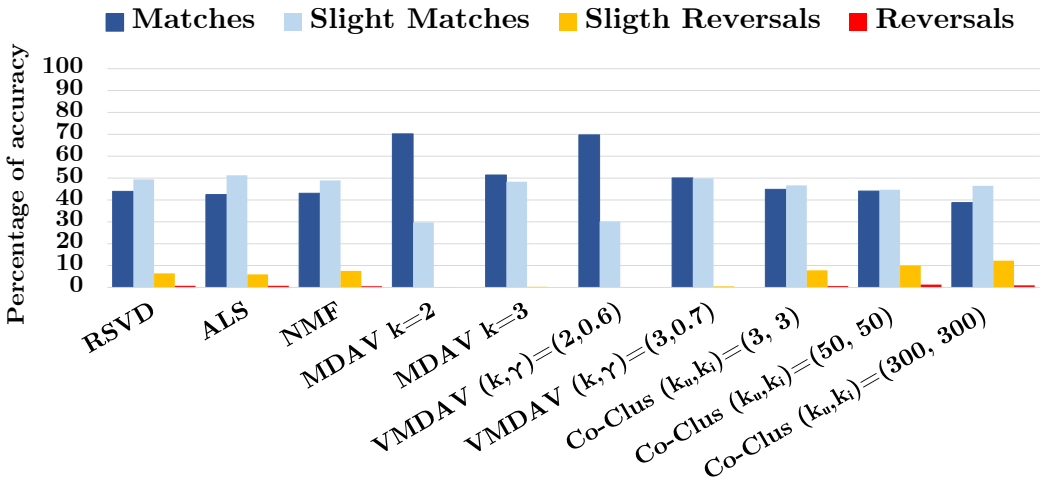
Table 5: MAE outcomes for all methods.

The behavioural precision outcomes are depicted in Figure 5. We observe that the behavioural precision (both binary and 4L) of the evaluated dimensionality reduction methods is similar, as seen in other works (Niemann and Wolpers, 2013; Mehta et al., 2007; Rendle and Schmidt-Thieme, 2008; George and Merugu, 2005). Nevertheless, the behavioural precision of MDAV and V-MDAV is better in all cases. One of the main reasons is because we fill the datasets with the central value, so that we cluster non-sparse data, decreasing possible bias in group creation. Note that using the central value is a feature that is valid for our approach but may not achieve the desired results when applied to dimensionality reduction methods such as RSVD or NMF. Therefore, as already explored in Koren (2010), imputation methods increase the amount of information and, thus, the computational time of such methods. Moreover, the bias introduced by such imputation may bring noise to the data Brick and Kalton (1996); Dempster et al. (1977), which translates into biased factorisation. In the case of the Co-clustering method, we observe that the higher the number of clusters, the lower the accuracy is, contrary to the behaviour of our method. The latter indicates that the co-clustering creation methodology can generate better outcomes when the number of referrals (i.e. neighbours in a cluster) is higher. Nevertheless, the 4L behavioural precision, which gives us a more fine-grained view than its binary counterpart, indicates that the percentage of errors of all methods are below 10%, except for two co-clustering configurations (i.e. Co-clus (50,50) and Co-clus (300,300)).

The outcomes of the precision, recall and F1 metrics are depicted in Table 6. We can observe that the highest precision@10 is obtained by the MDAV



(a) Binary behavioural accuracy - Movielens 100k



(b) 4L behavioural accuracy - Movielens 100k

Figure 5: MAE comparison between state of the art methods and our approach.

and V-MDAV methods, followed by the Co-Clus(3,3) method. In this case, we used the T-score optimization of the V-MDAV method, which slightly affects the recommendation accuracy, since we wanted to observe how much it would hinder the precision and recall compared to a the MAE strategy

Movielens 100k	P@10	R@10	F1@10	P@20	R@20	F1@20
RSVD	0.908	0.635	0.747	0.895	0.772	0.829
NMF	0.904	0.610	0.728	0.893	0.741	0.809
ALS	0.901	0.650	0.755	0.892	0.784	0.834
Co-Clus $(k_u, k_i) = (3, 3)$	0.911	0.605	0.727	0.900	0.739	0.811
Co-Clus $(k_u, k_i) = (50, 50)$	0.890	0.590	0.709	0.888	0.716	0.792
Co-Clus $(k_u, k_i) = (300, 300)$	0.870	0.580	0.696	0.868	0.717	0.785
MDAV $k = 2$	0.992	0.723	0.836	0.990	0.861	0.921
MDAV $k = 3$	0.985	0.727	0.837	0.980	0.866	0.919
V-MDAV $(k, \gamma) = (2, 0.6)$	0.990	0.724	0.836	0.987	0.862	0.920
V-MDAV $(k, \gamma) = (3, 0.7)$	0.983	0.729	0.837	0.978	0.868	0.919

Table 6: Precision, Recall and F1-score of the evaluated methods for the different values of n.

(where results would be almost the same than with MDAV). In terms of F1-score, the best outcomes are by MDAV and V-MDAV, followed by RSVD and ALS. Since the F1-score computes a harmonic mean, low values are penalised (i.e. a low recall has more weight than a high precision, see MDAV with $k = 2$). In the case of the outcomes when $n = 20$, we observe a similar pattern, in this case with much better recall values. The accurate results obtained by the microaggregation approaches can be related with these of binary behavioural precision in terms of correct items recommended in the top-n set. One of the keys of such success is the use of the central value imputation. Nevertheless, binary schemes fail to precisely identify to which extent the recommendation is adequate (i.e. moreover, in the case of precision and recall, this is usually done only for the a top-n set) and a more accurate scheme is needed. In this regard, the multilevel adaptability of the behavioural metrics provides enough resolution to analyse the predictions in a more extensive way, contrary to precision, recall or the F1-score.

5.2. Obfuscation efficiency analysis

As stated in Section 3, we want to study the amount of obfuscation that needs to be applied to data to achieve a desired level of privacy. Apparently, the lower the obfuscation needed, the more efficient a method is.

In this regard, we performed two experiments to analyse the obfuscation efficiency of MDAV and V-MDAV, since the rest of the methods do not guarantee privacy. First, we compared their MAE and DR outcomes, which are depicted in Figure 6. One may observe that the outcomes of both methods are very similar for Jester (*cf* Figure 6b). These results are not surprising because MDAV is known to perform very well on scattered datasets (Solanas and

Martinez-Balleste, 2006). In the case of Movielens 100k dataset (*cf* Figure 6a), a huge percentage of the data has been filled with the central value (*i.e.* the sparseness of this dataset is more than 90%). Such a procedure transforms Movielens 100k dataset into a clustered dataset, in which V-MDAV has been shown to outperform MDAV (Solanas and Martinez-Balleste, 2006). Second, we used the novel T-score metric presented in Section 3.2 to compare the values obtained by both methods for every k, which are depicted in Figure 7. Although the degrowth pace is nearly the same for both approaches, the T-score values tend to stabilise for high values of k. This occurs because, as the value of k increases, the gain factor γ of the V-MDAV method reaches better T-score values when the cardinality of the group is closer to k. Therefore, when the group cardinality is drastically increased, the groups generated by both methods become almost identical. The most extreme case would be reached when $n/2 < k \leq n$ (n stands for the total number of users) and, consequently, both methods would generate the same single group. Note that the maximum value of k achieved by V-MDAV is far higher than the one obtained by MDAV. This occurs because the γ values that optimise the T-score generate groups close to $2k-1$, since accuracy decreased at a slower pace than the level of privacy did. In the case of Movielens 100k (cf Figure 7a), V-MDAV outperforms MDAV for every value of k. However, in the case of Jester (cf Figure 7a), the efficiency results are quite similar for both methods. As previously stated, such outcomes are related to the distribution of data in each dataset (*i.e.* Movielens 100k is a clustered dataset and Jester is a scattered one).

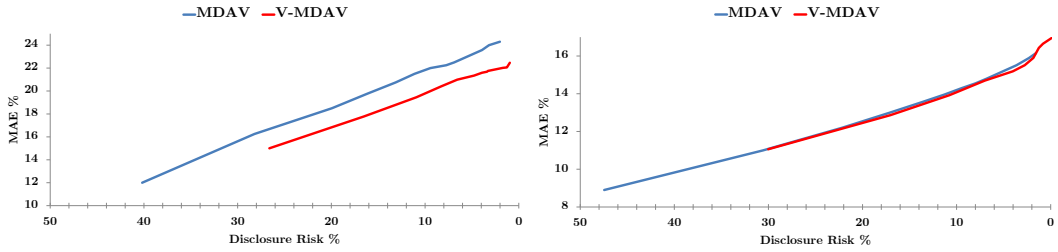
In general, we observe that our method obfuscates data more efficiently than the MDAV approach and, by extension than Gaussian noise addition (Casino et al., 2015).

5.3. Recommendation analysis

In what follows, we use the metrics presented in Section 3.2 to provide a comprehensive and holistic analysis of the quality of the recommendations for the datasets without privacy (i.e. using raw data) in section 5.3.1 and for the V-MDAV approach in Section 5.3.2.

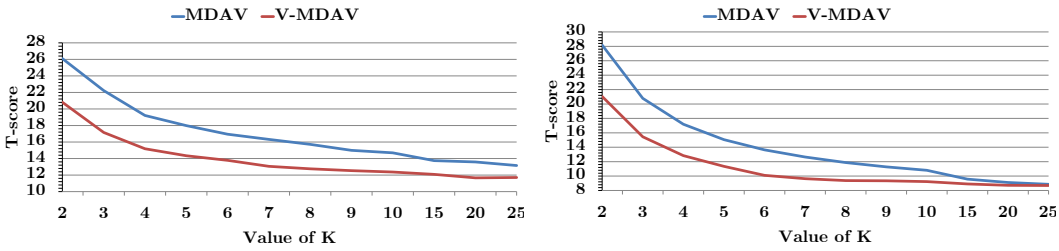
5.3.1. Raw data recommendations

The MAE obtained for the experiments without privacy, which compares raw data with data fulfilled using the central value, is shown in Table 7. Clearly, such results indicate that filling the databases with the central value



(a) MAE and DR comparison - Movielens 100k (b) MAE and DR comparison - Jester

Figure 6: MAE and DR comparison. The lower, the better. Note that the highest DR value achieved by V-MDAV is far lower than the one obtained by MDAV since V-MDAV heuristics tend to aggregate elements in each group, which increases the cardinality (*i.e.* always between k and $2k-1$) and decreases the DR value.



(a) Efficiency of the applied noise - Movielens 100k (b) Efficiency of the applied noise - Jester

Figure 7: Efficiency of the applied noise in both datasets regarding the T-score obtained by the evaluated approaches. The lower, the better.

leads to better accuracy than using the raw values. This is justified by the fact that; as discussed in Section 3.2, many neighbours may be required to obtain a valid value for every user when using raw matrices. Therefore, such evaluations can be provided by users who are not neighbours at all and, hence, such values become meaningless. Regarding behavioural precision, the outcomes depicted in Figure 8 indicate that filling the datasets with their corresponding central value provides both higher binary and 4L precision than using raw data. More precisely, in the case of binary precision (Figures 8a and 8b), we achieve more than 95% accuracy for the Movielens 100k dataset and more than 75% accuracy precision for the Jester dataset, results which clearly outperform the ones obtained with raw databases. In the case

of 4L precision analysis (Figures 8c and 8d), the majority of predictions conducted using the filled version of Movielens 100k have the same behaviour. Thus the error is close to 0%, as observed with binary precision. In contrast, the raw Movielens 100k dataset generates approximately 10% slight errors and almost a 5% reversals (although the % of matches is higher). When we analyse the results of the filled version of Jester, we observe that the sum of slight reversals and reversals is below 10%, while the total sum of errors obtained by raw Jester is more than 20%. Moreover, the behavioural matches of the raw Jester database are about 15% lower than those achieved with the filled version. In summary, filling the matrices with the central value gives us both more accuracy and precision in comparison with raw matrices.

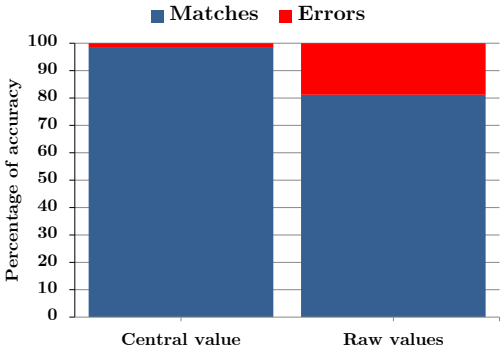
MAE %	**Movielens 100k**	**Jester**
Central value	23.519	19.022
Raw data	26.205	22.913

Table 7: MAE of the evaluated datasets, without privacy.

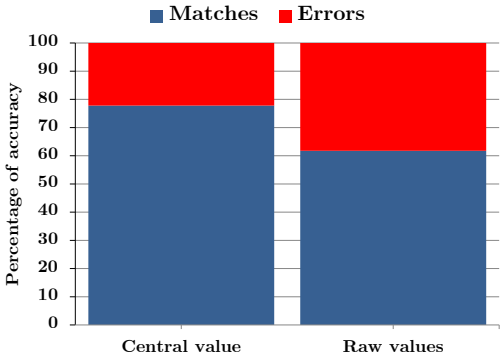
5.3.2. V-MDAV results

As shown in Section 5.2, our experiments indicate that V-MDAV outperforms the results of MDAV. For the sake of simplicity, in what follows we focus only on the analysis of the V-MDAV outcomes. The MAE results of our proposal for every value of k are depicted in Figure 9, and the behavioural precision of our method is illustrated in Figure 10.

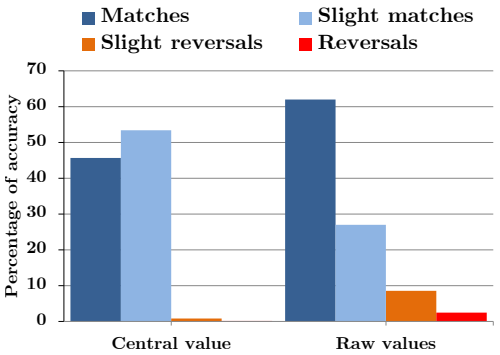
Movielens 100k and Jester datasets are very different in terms of sparseness, as already discussed in Section 4. As a result, Jester dataset achieves lower (*i.e.* and thus, better) MAE values due to its density, see Figure 9. Notwithstanding, all MAE values are lower than those from the datasets without privacy. More concretely, if we observe the MAE values shown in Table 5, we appreciate that they are above 19% (Jester) and above 23.5% (Movielens 100k). Nevertheless, the values obtained after using V-MDAV are below 16% (Jester) and 22% (Movielens 100k). Therefore, from the perspective of the accuracy of recommendations, it is better to obfuscate the data with V-MDAV (e.g. especially for $k = 2$), than using the original matrices without obfuscation. The measurements about the behavioural precision of our method, depicted in Figure 10, also achieve better accuracy than the



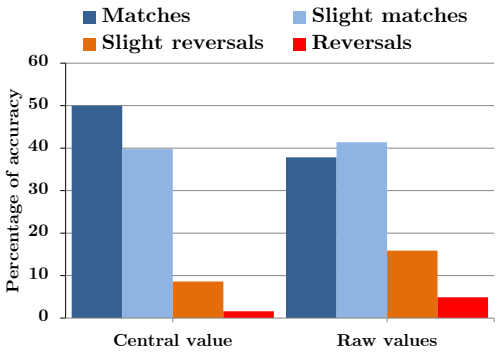
(a) Binary behavioural accuracy - Movielens 100k



(b) Binary behavioural accuracy - Jester



(c) 4L behavioural accuracy - Movielens 100k



(d) 4L behavioural accuracy - Jester

Figure 8: Recommendations of the datasets without privacy (*i.e.* before data obfuscation).

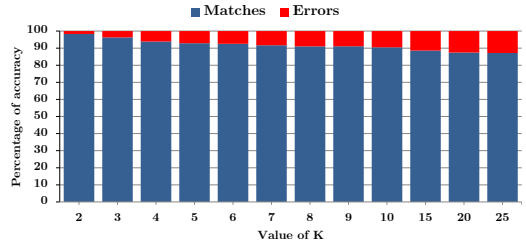
ones without privacy. Hence, our proposal achieves less percentage of errors and reversals in all cases.

In Figure 10, we observe that the outcomes of precision are better for Movielens 100k dataset than for Jester dataset, which initially seems to contradict the MAE values depicted in Figure 9. However, this precision accuracy can be explained since 90% of Movielens 100k's assessments are the respective central values, which may have both beneficial and detrimental effects. One the one hand, it is beneficial because the recommendations will not be biased; but on the other hand, it is detrimental because even if they are not biased, they will not be highly accurate. This fact is reflected in Figures 10c and 10d, where the Jester dataset reaches a little higher percentage

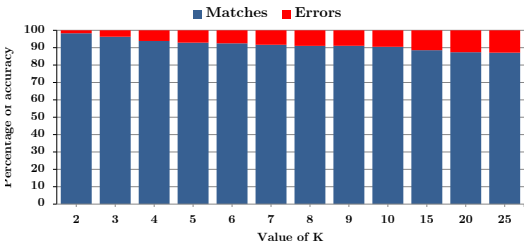
Figure 9: MAE% obtained after obfuscating the datasets using V-MDAV.

of errors than Movielens 100k (although in most cases they are lower than 5%), but also a higher percentage of matches because Jester dataset is far less sparse.

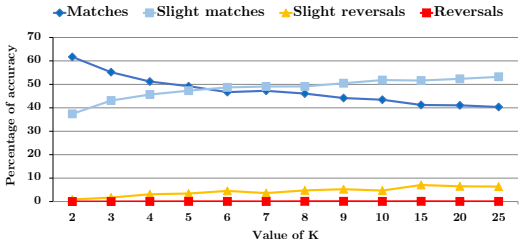
The aforementioned precision results could also be affected by the way the users perform their assessments, regarding the range values of the evaluated data (*e.g.* the higher the range, the more precise the users can be). Moreover, our proposed metrics unveil more information and thus, provide a better understanding of the outcomes than traditional metrics. For instance, although the MAE obtained with the Jester dataset experiments is better, the binary precision of Movielens 100k dataset is higher. The latter means that, in general, a higher amount of users receive meaningful (although not highly accurate) recommendations in the case of Movielens 100k. We may also observe that, although the 4L precision accuracy (cf Figure 10d) achieved a higher percentage of matches in the Jester dataset, it also generated more slight reversals. The aforementioned facts could not be discovered only with MAE nor precision and recall metrics. It is worth to recall that other metrics such as precision and recall (Su and Khoshgoftaar, 2009) only focus on true/false positives and do not provide a holistic analysis of the recommendation's quality. Therefore, the results of our experiments fully support our precision heuristics. The latter means that behavioural metrics can be considered a necessary tool to perform a more robust analysis of a recommender



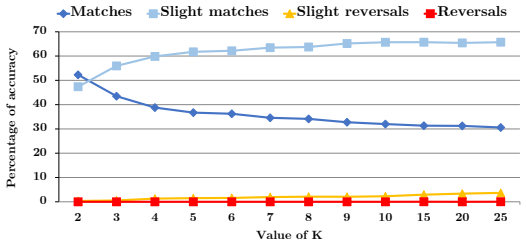
(a) Binary behavioural accuracy - Movielens 100k



(b) Binary behavioural accuracy - Jester



(c) 4L behavioural accuracy - Movielens 100k



(d) 4L behavioural accuracy - Jester

Figure 10: Recommendations after applying V-MDAV to the evaluated datasets.

system and assess its quality.

6. Conclusion and future work

PPCF methods aim to protect the users' privacy without sacrificing the quality of the recommendations. In this paper, we analysed how our proposal provides both more privacy and data usability (*i.e.* T-score) than other well-known methods, such as MDAV (Casino et al., 2015). Therefore, grouping the elements using a variable-sized cardinality strategy better fits the data distribution of the sparse, high-dimensional studied datasets. Moreover, we showed that the quality of the recommendations achieved by our method was even better than those obtained without obfuscating the datasets.

In addition, we proposed a new metric to compute the quality of the recommendations. We claim that the MAE metric is not enough, since it is subject to outliers, and in many occasions might not give accurate information about the recommendation's quality, (*e.g.* lots of users could receive nearly-perfect recommendations, and many others could receive reversals

while achieving an average MAE value). In order to address such cases, we introduced new precision measures which provide a better analysis of the recommendations, regardless of whether obfuscation is performed. Hence, the usability of recommender systems can be assessed more thoroughly, reflecting the behavioural precision of their recommendations. Future work will focus on two streams. First, the use of imputation techniques to improve the recommendation's quality. Second, to overcome dimensionality issues and detect outlier/malicious participants to obtain less biased results.

Acknowledgments and disclaimer

F. Casino and C. Patsakis were supported by the European Commission under the Horizon 2020 Programme (H2020), as part of the projects OPERANDO (Grant Agreement no. 653704) and YAKSHA (Grant Agreement no. 780498). Agusti Solanas is supported by the Government of Catalonia (GC) with grant 2017-DI-002, and with project 2017-SGR-896, and by Fundació PuntCAT with the Vinton Cerf Distinction, and by the Spanish Ministry of Science & Technology with project RTI2018-095499-B-C32.

The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.

References

Aggarwal, C.C., 2007. On randomization, public information and the curse of dimensionality, in: 2007 IEEE 23rd International Conference on Data Engineering, Instanbul, Turkey. pp. 136–145.

Basu, A., Vaidya, J., Kikuchi, H., Dimitrakos, T., 2011. Privacy-preserving collaborative filtering for the cloud, in: 2011 IEEE Third International Conference on Cloud Computing Technology and Science, Athens, Greece. pp. 223–230.

Batmaz, Z., Kaleli, C., 2017. Methods of privacy preserving in collaborative filtering, in: 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey. pp. 261–266.

Batmaz, Z., Polat, H., 2016. Randomization-based privacy-preserving frameworks for collaborative filtering. Procedia Comput. Sci. 96, 33–42.

Berkovsky, S., Eytani, Y., Kuflik, T., Ricci, F., 2007. Enhancing privacy and preserving accuracy of a distributed collaborative filtering, in: Proceedings of the 2007 ACM Conference on Recommender Systems, ACM, Minneapolis, MN, USA. pp. 9–16.

Bilge, A., Kaleli, C., Yakut, I., Gunes, I., Polat, H., 2013. A survey of privacy-preserving collaborative filtering schemes. Int. J. Softw. Eng. Know. 23, 1085–1108.

Bilge, A., Polat, H., 2013. A comparison of clustering-based privacy-preserving collaborative filtering schemes. Appl. Soft. Comput. 13, 2478–2489.

Bobadilla, J., Ortega, F., Hernando, A., Gutirrez, A., 2013. Recommender systems survey. Knowl-Based Syst. 46, 109 – 132.

Boutet, A., Frey, D., Guerraoui, R., Jégou, A., Kermarrec, A.M., 2016. Privacy-preserving distributed collaborative filtering. Computing 98, 827–846.

Brick, J., Kalton, G., 1996. Handling missing data in survey research. Stat. Methods Med. Res. 5, 215–238.

Canny, J., Canny, J., 2002. Collaborative filtering with privacy via factor analysis, in: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, New York, NY, USA. pp. 238–245.

Casino, F., Domingo-Ferrer, J., Patsakis, C., Puig, D., Solanas, A., 2013a. Privacy preserving collaborative filtering with k-anonymity through microaggregation, in: IEEE 10th International Conference on e-Business Engineering (ICEBE), Coventry, UK. pp. 490–497.

Casino, F., Domingo-Ferrer, J., Patsakis, C., Puig, D., Solanas, A., 2015. A k-anonymous approach to privacy preserving collaborative filtering. J. Comput. Syst. Sci. 81, 1000–1011.

Casino, F., Patsakis, C., Puig, D., Solanas, A., 2013b. On privacy preserving collaborative filtering: Current trends, open problems, and new issues, in: e-Business Engineering (ICEBE), 2013 IEEE 10th International Conference on, Coventry, UK. pp. 244–249.

Chaum, D., 1983. Blind signatures for untraceable payments, in: Advances in Cryptology, Springer. pp. 199–203.

Chou, H.Y., 2019. Units of time do matter: How countdown time units affect consumers intentions to participate in group-buying offers. Electron. Commer. Res. Appl. 35, 100839.

Cranor, L.F., Reagle, J., Ackerman, M.S., 1999. Beyond concern: Understanding net users attitudes about online privacy, in: The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy, MITP. pp. 47–70.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc., series B 39, 1–38.

Domingo-Ferrer, J., Martínez-Ballesté, A., Mateo-Sanz, J.M., Sebé, F., 2006. Efficient multivariate data-oriented microaggregation. The VLDB Journal 15, 355–369.

Domingo-Ferrer, J., Mateo-Sanz, J.M., 2002. Practical data-oriented microaggregation for statistical disclosure control. IEEE T. Knowl. Data. En. 14, 189–201.

Domingo-ferrer, J., Mateo-sanz, J.M., Torra, V., 2001. Comparing sdc methods for microdata on the basis of information loss and disclosure, in: Proceedings of ETK-NTTS 2001, Luxemburg: Eurostat, Eurostat. pp. 807–826.

Domingo-Ferrer, J., Torra, V., 2001. A quantitative comparison of disclosure control methods for microdata, in: Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies, Elsevier. pp. 111–133.

Friedman, A., et al., 2015. Privacy aspects of recommender systems, in: Recommender Systems Handbook, Springer US, Boston, MA. pp. 649–688.

Funk, S., 2006. Netflix update: Try this at home. http://sifter.org/simon/journal/20061211.html (accessed on 10 March 2018).

General Data Protection Regulation, 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Official Journal of the European Union, L 119 (4 May 2016), pp. 1-88 .

George, T., Merugu, S., 2005. A scalable collaborative filtering framework based on co-clustering, in: Fifth IEEE International Conference on Data Mining (ICDM'05), Houston, TX, USA. pp. 1–4.

Goldberg, D., Nichols, D., Oki, B.M., Terry, D., 1992. Using collaborative filtering to weave an information tapestry. Commun. ACM 35, 61–70.

Goldberg, K., Gupta, D., Digiovanni, M., Narita, H., 1999. Jester 2.0: Evaluation of a new linear time collaborative filtering algorithm, in: 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, CA, USA.

Gunes, I., Kaleli, C., Bilge, A., Polat, H., 2014. Shilling attacks against recommender systems: a comprehensive survey. Artif. Intell. Rev. 42, 767–799.

Han, J.H., Kim, H.M., 2019. The role of information technology use for increasing consumer informedness in cross-border electronic commerce: An empirical study. Electron. Commer. Res. Appl. 34, 100826.

Hannak, A., Soeller, G., Lazer, D., Mislove, A., Wilson, C., 2014. Measuring price discrimination and steering on e-commerce web sites, in: Proceedings of the 2014 Conference on Internet Measurement Conference, ACM, New York, NY, USA. pp. 305–318.

Huang, Z., Du, W., Chen, B., 2005. Deriving private information from randomized data, in: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA. pp. 37–48.

Hug, N., 2017. Surprise, a Python library for recommender systems. http://surpriselib.com (accessed on 13 January 2018).

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte-Nordholt, E., Spicer, K., de Wolf, P.P., 2012. Statistical Disclosure Control. Wiley.

Jeckmans, A.J., Beye, M., Erkin, Z., Hartel, P., Lagendijk, R., Tang, Q., 2013. Privacy in Recommender Systems. Computer Communications and Networks, Springer London.

Kaleli, C., Polat, H., 2010. P2P collaborative filtering with privacy. Turk. J. Electr. Eng. Co. 18, 101–116.

Karydi, E., Margaritis, K., 2016. Parallel and distributed collaborative filtering: A survey. ACM Comput. Surv. 49, 37:1–37:41.

Kluver, D., Ekstrand, M.D., Konstan, J.A., 2018. Rating-based collaborative filtering: algorithms and evaluation, in: Social Information Access. Springer, pp. 344–390.

Koren, Y., 2010. Factor in the neighbors: Scalable and accurate collaborative filtering. ACM Trans. Knowl. Discov. Data 4, 1:1–1:24.

Li, J., Yang, J., Zhao, Y., Liu, B., Zhou, M., Bi, J., Wang, Q., 2017. Enforcing differential privacy for shared collaborative filtering. IEEE Access 5, 35–49.

Luo, X., Zhou, M., Xia, Y., Zhu, Q., 2014. An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. IEEE T. Ind. Inform. 10, 1273–1284.

Mackinnon, K.A., 2012. User generated content vs. advertising: Do consumers trust the word of others over advertisers?. The Elon Journal of Undergraduate Research in Communications 3, 14–22.

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: Homophily in social networks. Annu. Rev. Sociol. 27, 415–444.

Mehta, B., Hofmann, T., Nejdl, W., 2007. Robust collaborative filtering, in: Proceedings of the 2007 ACM Conference on Recommender Systems, ACM, New York, NY, USA. pp. 49–56.

Minkus, T., Ross, K.W., 2014. I know what you're buying: Privacy breaches on ebay, in: Privacy Enhancing Technologies, Springer. pp. 164–183.

Narayanan, A., Shmatikov, V., 2008. Robust de-anonymization of large sparse datasets, in: 2008 IEEE Symposium on Security and Privacy (sp 2008), Oakland, CA, USA. pp. 111–125.

Niemann, K., Wolpers, M., 2013. A new collaborative filtering approach for increasing the aggregate diversity of recommender systems, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA. pp. 955–963.

Okkalioglu, M., Koc, M., Polat, H., 2016. A privacy review of vertically partitioned data-based ppcf schemes. Int. J. Inform. Sec. Sci. 5, 51–68.

O'Mahony, M.P., Hurley, N.J., Silvestre, G.C., 2006. Detecting noise in recommender system databases, in: Proceedings of the 11th International Conference on Intelligent User Interfaces, ACM, New York, NY, USA. pp. 109–115.

Phan, L.P., Huynh, H.H., Huynh, H.X., 2017. User based recommender systems using implicative rating measure. Int. J. Adv. Comput. Sci. Appl. 8, 37–43.

Polat, H., 2003. Privacy-preserving collaborative filtering using randomized perturbation techniques, in: Third IEEE International Conference on Data Mining, IEEE Comput. Soc, Melbourne, FL, USA. pp. 625–628.

Polat, H., Du, W., 2008. Privacy-preserving top-n recommendation on distributed data. J Am. Soc. Inform. Sci. 59, 1093–1108.

Politou, E., Alepis, E., Patsakis, C., 2018. Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. J. Cybersecurity 4, tyy001.

Rendle, S., Schmidt-Thieme, L., 2008. Online-updating regularized kernel matrix factorization models for large-scale recommender systems, in: Proceedings of the 2008 ACM Conference on Recommender Systems, ACM, New York, NY, USA. pp. 251–258.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J., 1994. Grouplens: an open architecture for collaborative filtering of netnews, in: Proceedings of the 1994 ACM conference on Computer supported cooperative work, ACM. pp. 175–186.

Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S., 2007. Collaborative filtering recommender systems, in: The Adaptive Web, Springer. pp. 291–324.

Shi, Y., Larson, M., Hanjalic, A., 2014. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. ACM Comput. Surv. (CSUR) 47, 1–45.

Shin, H., Kim, S., Shin, J., Xiao, X., 2018. Privacy enhanced matrix factorization for recommendation with local differential privacy. IEEE T. Knowl. Data En. 30, 1770–1782.

da Silva, E.Q., Camilo-Junior, C.G., Pascoal, L.M.L., Rosa, T.C., 2016. An evolutionary approach for combining results of recommender systems techniques based on collaborative filtering. Expert Syst. Appl. 53, 204 – 218.

Solanas, A., Martinez-Balleste, A., 2006. V-mdav: a multivariate microaggregation with variable group size, in: 17th COMPSTAT Symposium of the IASC, Rome, pp. 917–925.

Su, X., Khoshgoftaar, T.M., 2009. A Survey of Collaborative Filtering Techniques. Adv. in Artif. Intell. 2009, 1–19.

The Nielsen Company, 2009. Personal recommendations and consumer opinions posted online are the most trusted forms of advertising globally. http://www.nielsen.com/eu/en/press-room/2015/recommendations-from-friends-remain-most-credible-form-of-advertising.html (accessed on 6 February 2018.

Wei, R., Tian, H., Shen, H., 2018. Improving k-anonymity based privacy preservation for collaborative filtering. Comput. Electr. Eng. 67, 509 – 519.

Xian, Z., Li, Q., Huang, X., Li, L., 2017. New svd-based collaborative filtering algorithms with differential privacy. J. Intell. Fuzzy. Syst. 33, 2133–2144.

Yakut, I., Polat, H., 2012. Estimating nbc-based recommendations on arbitrarily partitioned data with privacy. Knowl-Based Syst. 36, 353–362.