ABSTRACT
Recommenders are becoming one of the main ways to navigate the Internet. They recommend appropriate items to users based on their clicks, i.e., likes, ratings, purchases, etc. These clicks are key to providing relevant recommendations and, in this sense, have a significant utility. Since the user clicks reflect the preferences of users, they may raise privacy concerns. Intuitively, the effects on utility and privacy are not the same for all clicks.

For the first time, we quantify the exact utility and privacy effects of each user click. Interestingly enough, we show that against the common belief of a trade-off between utility and privacy, they are not always antinomic regarding clicks. An appealing application of our measures is what we call a click-advisor diagram, enabling users to decide whether it is worth clicking on an item or not.

1. METHODS
1.1 The utility of a Click
We quantify the precise effect of a click on recommender utility by introducing the notion of commonality of a user profile i.e. how close the taste of a user is to that of other users (which helps a recommender suggest new items to that user). Intuitively, commonality captures the precision of recommenders. The difference between the commonality of a user profile before and after a click corresponds to the notion of the utility of that click.

1.2 The disclosure Risk of a Click
To compute the privacy effect of a click, we introduce the concept of disclosure degree of a user profile, using an information-theoretic approach. The disclosure degree corresponds to the amount of information stored in a user profile, also known as entropy [3]. Roughly speaking, the larger the amount of information in a user profile, the higher the disclosure degree of the user. Intuitively, if the disclosure degree of a user profile is low, then the user is not easily distinguishable from others. We then capture the disclosure risk of a click, implying its privacy effect, as the difference between the disclosure degree of a user profile before and after the click. The click-advisor diagram We demonstrate that, the traditional belief of a trade-off between utility and privacy [5, 6], (i.e.) a user necessarily improves recommendation utility at the expense of compromising privacy (or vice versa), does not always hold. On the contrary, we show that there are some (safe) clicks that can improve both utility and privacy on the one hand and (dangerous and deleterious) clicks that do not improve utility, yet they hamper privacy on the other hand. The rest of the clicks are said to be in the trade-off zone. We precisely compute the conditions for a click to be in each zone. We show how to depict the zones of all clicks, at a click time, in what we call the click-advisor diagram.

A click-advisor diagram can be viewed as a tool to help users decide whether or not to actually click on an item. The process is as follows: a user (a) pre-clicks on an item, (b) gets the feedback on the utility and privacy effects of all possible clicks for rating based on the real preference of the user shown in that pre-click and then (c) decides to confirm, change or cancel the pre-click.
<table>
<thead>
<tr>
<th>Recommended</th>
<th>Relevant</th>
<th>Irrelevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_p(u)</td>
<td>t_p(u)</td>
<td>f_p(u)</td>
</tr>
<tr>
<td>Not Recommended</td>
<td>f_n(u)</td>
<td>t_n(u)</td>
</tr>
</tbody>
</table>

Table 1: Confusion Matrix for True/False Positive/Negative Recommendations

2. RESULTS

2.1 Commonality as a utility measure

We show here empirically that the notion of commonality of a user indeed expresses the quality of a recommender for that user; therefore commonality captures the recommender utility for the users. We study the correlation between commonality and the classical notion of precision of a recommender as defined in [8].

We measure the recommendation quality for each user as follows: we divide the dataset into a training set and a test set. We put each rating from the original dataset into the test set with probability 20% and hide all the ratings in the test set from the training set. We determine the top recommendations for each user based on the state-of-the-art recommender algorithms, namely a matrix reconstruction approach (as described in [9]) and a SVD-based neighboring recommender (as described in [4]). We denote the number of recommendations for a user by N = 20.

The quality is measured in terms of standard classification accuracy metrics (CAM). More precisely, we evaluate how well a recommender can predict the context of the test set from the training set. We determine the top recommendations for each user based on the state-of-the-art recommender algorithms, namely a matrix reconstruction approach (as described in [9]) and a SVD-based neighboring recommender (as described in [4]). We denote the number of recommendations for a user by N = 20.

The quality is measured in terms of standard classification accuracy metrics (CAM). More precisely, we evaluate how well a recommender can predict the context of the test set of a user u using each of the mentioned algorithms on the training set. For this purpose, we determine N-top recommendations for user u and compare them with the test set part of user profile u. We have four different cases for hidden and recommended items as shown in Table 1. We compute Precision(u) as the classification accuracy metric used on top-N recommendations computed in [2]. Precision(u) is the ratio of the number of relevant recommended items to the total number of recommended items to user u. Actually, Precision(u) computes the recommendation quality for u:

\[
\text{Precision}(u) = \frac{t_p(u)}{t_p(u) + f_p(u)}
\]

We report here our results for both MovieLens. As seen in Figures 2 and 3, the commonality and precision follow linear fashion in MovieLens for SVD-based recommenders using matrix reconstruction and neighboring. Mathematically evaluated, the correlation between commonality and precision is 0.6557 for SVD-based recommender using matrix reconstruction and the correlation increases to 0.6838 for a neighboring SVD-based recommender.

2.2 Disclosure degree as a privacy measure

As discussed earlier, different clicks have different effects on user privacy. Also, we introduced the disclosure degree of a user profile as a privacy parameter. In this section, we use the MovieLens and Jester datasets to show that interestingly the disclosure degree captures other well-known privacy concepts such as differential privacy and k-anonymity.

Differential privacy is a very strong privacy criteria for an algorithm which guarantees that the presence or absence of a record in the dataset will not affect the final output of the algorithm significantly for any possible dataset. However, in a real system, we only ought to protect the information of users in the current dataset from being disclosed. Hence, in this evaluation we focus on a specific dataset, namely MovieLens. Disclosure degree of users in a system is computed only based on user profiles in the system. To justify the disclosure degree as a privacy parameter of user profiles, we survey the relation between the differential privacy and disclosure degree. We employ the approach proposed in [7] to provide differential privacy to recommenders by adding different levels of Laplacian noise to MovieLens dataset. Each noisy dataset corresponds to a level of differential privacy. As a comparison, Figure 4 represents the expected value of disclosure degree of users in each of these noisy datasets. We observe that the users in a dataset with a high level of noise (a low epsilon differential private dataset) have lower disclosure degree comparing to the users in a lower differential private dataset. Intuitively, the presence or absence of a rare click (i.e. with high disclosure risk) can highly affect the output of the recommender comparing to the presence or absence of a regular (expected) click.

F. Casino et.al. applied the concept of k-anonymity to the content of collaborative filtering in [1]. Figure 5 presents the expected value of disclosure degree of user profiles\(^1\) in MovieLens.

\(^1\)Note that as k-anonymity is a privacy parameter for a dataset, not a user profile, to compare k-anonymity and disclosure degree, we use the expected value of disclosure degree
lens as well as Jester with different levels of k-anonymity. As depicted in Figure 5, increasing $k$ (i.e., anonymity for users) results in decreasing the expected value of disclosure degree of user profiles. Our concept of disclosure degree of a user profile captures how unique a user profile is in the system. Intuitively, the higher the disclosure degree of a user profile, the more distinguishable the user is among others (the higher the risk of identifying the corresponding user among all the users in the system).

2.3 Click Zones
We compute the effect of a click on utility and privacy using disclosure risk and utility of a click for Movielens. We analyze the Movielens dataset and figure out the percentages of different types of clicks in Table 2. Figure 6 represents the distribution of the different types of click zones in Movielens for a limited number of users and items (because of the sake of clarity). We observe that most of the clicks induce a trade-off between utility and privacy. Yet, there are some clicks which have the same effect on both utility and privacy (e.g., safe, dangerous and deleterious clicks)$^2$.

3. CONCLUDING REMARKS
We show that clicks do not all have the same effect, neither in terms of privacy nor in terms of utility. Click-advisor of all user profiles as an overall users privacy measure. $^2$Risky clicks are either dangerous or deleterious.

<table>
<thead>
<tr>
<th>Type of Click</th>
<th>Safe</th>
<th>Trade-off</th>
<th>Dangerous</th>
<th>Deleterious</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>5.12</td>
<td>78.40</td>
<td>15.05</td>
<td>1.43</td>
</tr>
</tbody>
</table>

Table 2: Distribution of Different Types of Clicks for Movielens

Figure 4: The Disclosure Degree vs Differential Privacy

Figure 5: The Disclosure Degree and K-anonymity

Figure 6: The Distribution of Different Types of Clicks

enables users to provide their informed opinions about their clicks.

4. REFERENCES