Regret Bounds and Regimes of Optimality for User-User and Item-Item Collaborative Filtering

Guy Bresler* and Mina Karzand†

November 8, 2017

Abstract

We consider an online model for recommendation systems, with each user being recommended an item at each time-step and providing ‘like’ or ‘dislike’ feedback. A latent variable model specifies the user preferences: both users and items are clustered into types. All users of a given type have identical preferences for the items, and similarly, items of a given type are either all liked or all disliked by a given user. The model captures structure in both the item and user spaces, and in this paper we assume that the type preference matrix is randomly generated.

We describe two algorithms inspired by user-user and item-item collaborative filtering (CF), modified to explicitly make exploratory recommendations, and prove performance guarantees in terms of their expected regret. For two regimes of model parameters, with structure only in item space or only in user space, we prove information-theoretic lower bounds on regret that match our upper bounds up to logarithmic factors. Our analysis elucidates system operating regimes in which existing CF algorithms are nearly optimal.

1 Introduction

Options are good, but if there are too many options, we need help. It is increasingly the case that our interaction with content is mediated by recommendation systems. There are two main approaches taken in recommendation systems: content filtering and collaborative filtering. Content filtering makes use of features associated with items and users (e.g., age, location, gender of users and genre, director of movies). In contrast, collaborative filtering is based on observed user preferences. Thus, two users are thought of as similar if they have revealed similar preferences irrespective of their profile. Likewise, two items are thought of as similar if most users have similar preferences for them. More generally, collaborative filtering (CF) makes use of structure in the matrix of preferences, as in low-rank matrix formulations [1, 4, 8, 13, 14, 19, 20, 25].

An important aspect of most recommendation systems is that each recommendation influences what is learned about the users and items, which in turn determines the possible accuracy of future recommendations. This introduces a tension between exploring to obtain information and exploiting existing knowledge to make good recommendations. The tension between exploring and exploiting is exactly the phenomenon of interest in the substantial literature on the multi-armed bandit (MAB) problem and its variants [7, 16, 21]. In the multi-armed bandit setup, optimal

*Guy Bresler is with the Department of Electrical Engineering and Computer Science at MIT and a member of LIDS and IDSS. guy@mit.edu
†Mina Karzand is with the Department of Electrical Engineering and Computer Science at MIT and a member of LIDS and IDSS. mkarzand@mit.edu
‡The list of authors is in alphabetical order.
algorithms necessarily converge to repeated play of the same arm; in contrast, a recommendation
system that repeatedly recommends the same movie, even if it is a very good movie, is surely
problematic! For this reason we will allow each item to be recommended at most once to each user
(as done in [5, 6]).

It is common to think of recommendation systems as a matrix completion problem. Given a
subset of observed entries, the matrix completion problem is to estimate the rest of matrix, where it
is assumed that the matrix satisfies some properties. This criterion does not capture the experience
of users in a recommendation system: a more appropriate measure of performance is the proportion
of good recommendations made by the algorithm.

The first contribution of this work is to provide a framework for evaluating the performance
of various recommendation system algorithms. We focus on a latent variable model in which each
user is associated with a user type and each item is associated with an item type. Users who belong
to the same user type share their preferences for all items and items belonging to the same type are
rated by everyone similarly.1 The measure of performance is expected regret, equal to the expected
number of bad recommendations made per user made over a time horizon of interest.

In the literature there are two categories of collaborative filtering (CF) algorithms: user-user
algorithms [3, 5, 12] use structure in the user space to predict user preferences. Here, the preference
of user u for item i is estimated from the preference of other users u′ believed to be similar to u
based on their previous ratings. Alternatively, item-item algorithms [6, 17, 22] use structure in the
item space. This time, the preference of user u for item i is estimated from the preference of the
same user u for other items i′ believed to be similar to i based on previous ratings from users that
have rated both i and i′. We develop versions of user-user and item-item CF algorithms tailored
to our online recommendation system model and prove performance guarantees.

Next, we focus on two extreme parameter regimes of interest in the latent variable model. In
the user structure only scenario, the model parameters are such that there is no structure in the
item space. Alternatively, in the item structure only scenario, the parameters are such that there
is no structure in the user space. We provide information-theoretic lower bounds for performance
of any algorithm in the user-structure only model and item-structure only model. This shows that
our proposed user-user and item-item online CF algorithms are almost information-theoretically
optimal (within multiplicative logarithmic factors).

One of the main contributions of the present paper is the development of techniques for proving
information-theoretic lower bounds on the performance of online recommendation algorithms. The
methods are elementary in nature and do not require sophisticated tools from information theory.
As an example, in the user-structure only model, we make two observations: we cannot be confident
in recommending any item to user u at time t if there is no user u′ that has rated enough items
in common, with user u by time t − 1. In this situation, the similarity of u to any other user is uncertain and so too is the outcome of any recommendation. The outcome of
recommending item i to user u at time t is also uncertain if none of the users similar to u have
rated item i by time t − 1. These observations imply a lower bound on the necessary number of
exploratory recommendations before it is possible to recommend with much better likelihood of
success than chance. Similar observations are made for the model based on item-structure only.

A few papers including [5, 6, 11] have provided theoretical guarantees for online collaborative
filtering. The paper [5] analyzes a user-user CF algorithm in a similar setting to ours and [6]
analyzes an item-item CF algorithm in a somewhat different and more flexible model. Relative to
these, our main distinction is obtaining nearly matching lower bounds showing optimality of our

1A similar model of data to ours, in which there is an underlying clustering of rows and columns, has been studied
in other settings [23, 27].
algorithms and analysis. The model studied by Dabeer and coauthors [11, 1, 2] is quite similar to our setup, but their objective is different: their goal is to maximize the probability of recommending a liked item at finite horizon starting with some historical knowledge of preference of users for items. In particular, they provide an algorithm that exploits in a provably optimal fashion asymptotically, but their approach does not reveal how to explore.

Hybrid algorithms exploiting both structure in user space and item space have been studied before in [24, 26, 15]. Song et al. [24] studies a more flexible latent variable model in the offline (matrix completion style) setting and propose collaborative filtering algorithms using both item and user space. In a forthcoming paper we analyze a hybrid algorithm within the same framework studied here.

1.1 Outline

The model and performance metric are described in Section 2. Section 3 overviews the main results of this paper. Our version of user-user CF is introduced and analyzed in Section 4. In Section 5 we prove that the proposed algorithm is almost information-theoretically optimal in the setup with user structure only, introduced in Definition 2.1. Our version of item-item CF is described and analyzed in Section 6, and the corresponding lower bound in the setting with item structure only is given in Section 7. Appendices A and B contain a couple of useful lemmas.

1.2 Notation

For an integer $a$ we write $[a] = \{1, \cdots, a\}$ and for real-valued $x$ let $(x)_+ = \max\{x, 0\}$. All logarithms are base 2. The set of natural numbers (positive integers) is denoted by $\mathbb{N}$. We note here that variables or parameters in Figure 1 have the same meaning throughout the paper, but any others may take different values in each section. For real-valued $x$, $\lfloor x \rfloor$ denotes the greatest integer less than or equal to $x$. Numerical constants (e, c, c1, c2 and so forth) may take different values in different theorem statements unless explicitly stated otherwise.

2 Model

2.1 Problem setup

There is a fixed set of users $\{1, \cdots, N\}$. At each time $t = 1, 2, 3, \ldots$ the algorithm recommends an item $a_{u,t} \in \mathbb{N}$ to each user $u$ and receives feedback $L_{u,a_{u,s}} \in \{+1, -1\}$ (‘like’ or ‘dislike’). For the reasons noted in the introduction, we impose the condition that each item may be recommended at most once to each user. In order that the algorithm never run out of items to recommend, we suppose there are infinitely many items to draw from and identify them with the natural numbers.

The history at time $t$, $\mathcal{H}_{t-1}$, is the collection of actions and feedback up to time $t-1$, i.e., $\mathcal{H}_{t-1} = \{a_{u,s}, L_{u,a_{u,s}}\}$ for $u \in [N], s \in [t-1]$. We are interested in online learning algorithms, in which the action $a_{u,t}$ is a (possibly random) function of $\mathcal{H}_{t-1}$. This additional randomness is encoded in a random variable $\zeta_{u,t}$, assumed to be independent of all other variables. In this way, $a_{u,t} = f_{u,t}(H_{t-1}, \zeta_{u,t})$, for some deterministic function $f_{u,t}$.

Algorithm performance will be evaluated after some arbitrary number of time-steps $T$. The performance metric we use is expected regret (simply called regret in what follows), defined as the expected number of disliked items recommended per user:

$$\text{regret}(T) = \mathbb{E}\left[\sum_{t=1}^{T} \frac{1}{N} \sum_{u=1}^{N} 1[L_{u,a_{u,t}} = -1]\right].$$ (1)
| Symbol | Description |
|--------|-------------|
| \(N\)  | Number of users |
| \(T\)  | Time horizon |
| \(q_U\) | Number of user types |
| \(q_I\) | Number of item types |
| \(\tau_U(u)\) | User type of user \(u\) |
| \(\tau_I(i)\) | Item type of item \(i\) |
| \(a_{u,t}\) | Item recommended to user \(u\) at time \(t\) |
| \(L_{u,i}\) | Rating of user \(u\) for item \(i\) |
| \(\xi_{k,j}\) | Preference of user type \(k\) for item type \(j\) |
| \(\Xi\)  | Preference matrix |

Figure 1: Notation for the recommendation system model.

The algorithms we describe depend on knowing the time-horizon \(T\), but by standard techniques in the multi-armed bandit literature it is possible to convert these to algorithms achieving the same (up to constant factors) regret without this knowledge. This latter notion of regret, where the algorithm does not know the time-horizon of interest and must achieve good performance across all time-scales, is called *anytime regret* in the literature. In Section B, a method called doubling trick is introduced which converts online algorithms for finite horizons to algorithms achieving the same anytime regret (up to multiplicative constant) [9].

### 2.2 User preferences

We study a latent-variable model for the preferences (‘like’ or ‘dislike’) of the users for the items, based on the idea that there are relatively few *types* of users and/or few *types* of items. Each user \(u \in [N]\) has a user type \(\tau_U(u)\) i.i.d. uniformly distributed on \([q_U]\), where \(q_U\) is the number of user types. Similarly, each item \(i \in N\) has a random item type \(\tau_I(i)\) i.i.d. uniformly distributed on \([q_I]\), where \(q_I\) is the number of item types. The random variables \(\{\tau_U(u)\}_{1 \leq u \leq N}\) and \(\{\tau_I(i)\}_{1 \leq i \leq n}\) are assumed to be jointly independent.

All users of a given type have identical preferences for all the items, and similarly all items of a given type are rated in the same way by any particular user. The entire collection of user preferences \((L_{u,i})_{u,i}\) is therefore encoded into a much smaller *preference matrix* \(\Xi = (\xi_{k,j}) \in \{-1, +1\}^{q_U \times q_I}\), which specifies the preference of each user type for each item type. The preference \(L_{u,i}\) of user \(u \in [N]\) for item \(i \in N\) is the preference \(\xi_{\tau_U(u),\tau_I(i)}\) of the associated user type \(\tau_U(u)\) for the item type \(\tau_I(i)\) in the matrix \(\Xi\), i.e.,

\[
L_{u,i} = \xi_{\tau_U(u),\tau_I(i)}.
\]

We assume that the entries of \(\Xi\) are i.i.d., \(\xi_{k,j} = +1\) w.p. \(1/2\) and \(\xi_{k,j} = -1\) w.p. \(1/2\). Generalizing our results to i.i.d. entries with bias \(p\) is relatively straightforward, but the independence assumption is quite strong and an important future research direction is to obtain results for more realistic preference matrices.

### 2.3 Two regimes of interest

Throughout the paper, we will assume that the number of user types \(q_U\) and the number of item types \(q_I\) are both at least \(18 \log N\). As will become clear via the theorem statements, we can additionally show optimality up to logarithmic factors of the proposed algorithms in certain regimes.
Two specific parameter regimes play a central role in this paper, capturing settings with structure only in user space or only in item space. As described in Section 3, one of user-user or item-item CF is almost optimal in the corresponding regime.

**Definition 2.1** (User structure only \((q_I = 2^{q_U})\)). The user structure model refers to the case that there is no structure in the item space. To simplify matters, we assume that the preference matrix \(\Xi \in \{-1, +1\}^{q_U \times 2^{q_U}}\) is fixed and has columns consisting of all sequences in \(\{-1, +1\}^{q_U}\). Essentially the same preference matrix would arise (with high probability) if the entries are i.i.d. as specified above in Subsection 2.2 if \(q_I\) is much larger than \(2^{q_U}\).

**Definition 2.2** (Item structure only \((q_U = N)\)). The item structure model refers to the case that there is no structure in the user space. This happens when \(q_U\) is much larger than \(N\), since then most user types have no more than one user. For the purpose of proving near-optimality of item-item CF, it suffices to take \(q_U = N\) (and we do so).

## 3 Main results

In this paper, we analyze a version of each of user-user and item-item CF within the general setup described in Section 2. The resulting regret bounds appear in Theorems 4.1 and 6.1. These theorems are complemented by information theoretic-lower bounds, Theorems 5.1 and 7.1, showing that no other algorithm can achieve better regret (up to multiplicative logarithmic factors) in the specific extreme parameter regimes with user-structure only and item-structure only.

**User-user collaborative filtering.** User-user CF exploits structure in the user space: the basic idea is to recommend items to a user that are liked by similar users. We analyze an instance of user-user CF described in detail in Section 4.1, obtaining the regret bound given in Theorem 4.1 below. Essentially, the algorithm clusters users according to type by recommending random items for an initial phase, and then uses this knowledge to efficiently explore the preferences of each user type. The subsequent savings is due to the fact that the cost of exploration can be shared amongst users of the same type.

The random recommendations during the initial phase incur regret with slope \(1/2\), because a random recommendation is disliked with probability half. Afterward, the users are clustered according to type. Recommending an item to \(q_U\) users, one from each type, gives us the preferences of all \(N\) users for the item, and each such recommendation is disliked with probability \(1/2\). This results in a slope of \(q_U/2N\) for regret in the second phase of the algorithm.

**Theorem 4.1** (Regret upper bound in user-user CF, simplified version). **Consider the recommendation system model described in Section 2 with \(N\) users, \(q_U\) user types, and \(q_I > 16 \log N\) item types. There exists a numerical constant \(c\) so that Algorithm 1 achieves regret**

\[
\text{regret}(T) \leq \begin{cases} 
\frac{c}{2} \log N, & \text{if } T \leq c \log N \\
\frac{q_U}{2N} T, & \text{if } T > c \log N.
\end{cases}
\]

If there is no structure in the item space and the number of user types is \(q_U = N^\alpha\) for fixed \(0 < \alpha < 1\), then the user-user CF algorithm is optimal up to a multiplicative constant.

**Theorem 5.1** (Regret lower bound with user structure only, simplified version). There exist numerical constants \(c\) such that in the user structure model (Defn 2.1) with \(q_U = N^\alpha\) user types and
For $N > N_0(\alpha)$ users, any recommendation algorithm must incur regret
\[
\text{regret}(T) \geq \begin{cases} 
0.49T - 4, & \text{if } T \leq c \log q_U \\
\frac{q_I}{c^2} T, & \text{if } T > c \log q_U.
\end{cases}
\]

The reasoning for the first part of the lower bound is as follows. If a user has been recommended fewer than $\log q_U$ items, then its similarity with respect to other users cannot be determined. This implies that any recommendation made to this user has uncertain outcome. The second part of the lower bound is obtained by showing that when an item is recommended for the first time to a user from a given user type the outcome of that recommendation is uncertain, and lower bounding the number of such recommendations. This is where we use the condition that each item is recommended at most once to each user.

The lower bound shows that the poor initial performance of user-user CF, as bad as simply recommending random items, is unavoidable in the setting with only user structure. In the recommendation systems literature the notion of cold start describes the difficulty of providing useful recommendations when insufficient information is available about user preferences. In [6] a formal definition of cold start time was provided, and it was shown that a version of item-item CF obtains much smaller cold start time than user-user CF in a model with item structure only. Our results on item-item CF, described next, corroborate this.

**Item-item collaborative filtering.** Item-item CF exploits structure in the item space: users are recommended items similar to those they have liked. We analyze an instance of item-item CF in Section 6.1, obtaining the regret bound given in Theorem 6.1 below. Crucially, this version of item-item CF has the feature that only a subset of the item space is explored. To the best of our knowledge, the benefit of limiting the scope of item exploration has not been made explicit before; this only became evident to us in seeking to match the lower bound.

Initially, the proposed algorithm recommends a few random items to all users. This is referred to as self-exploration of the item space by the users, because every user learns its preference for the types of the explored items, which we will call representatives. The algorithm then compares a set of items to the representatives by recommending each to a subset of random users. We refer to this phase as joint exploration of the item space, because the effort of classifying the items is shared among all the users. This yields a set of item clusters, one for each explored type. If the number of representatives is smaller than $q_I$, the number of item types, then the portion of items not matching any representative will be cast aside. Subsequently, users are recommended items from the clusters corresponding to liked types.

The number of items to be partitioned is chosen so that there will be enough items to recommend to each user for the entire time-horizon of length $T$. There is a trade-off between the cost of partitioning (the number of random recommendations made to do the partitioning as part of joint exploration) and the cost of learning the preference of users for item types (self-exploration): if more item types are learned, then fewer items per type are needed and hence fewer items in total need to be partitioned. How much of the item space to be explored by all users is chosen at the optimal point in the trade-off to minimize regret as a function of the horizon $T$, the number of item types $q_I$ and the total number of users $N$.

**Theorem 6.1 (Regret upper bound in item-item CF, simplified version).** Consider the recommendation system model described in Section 2 with $N$ users, $q_U$ user types and $q_U > 19 \log N$ item types. There are numerical constants $c_1, c_2, c_3, c_4$ such that Algorithm 2 obtains regret per user
at time $T$ upper bounded as

$$\text{regret}(T) \leq c_1 + c_2 \begin{cases} 
\log T, & \text{if } T \leq c_3 \frac{N}{q_I} (\log T)^2 \\
\sqrt{\frac{q_I \log N}{N}} T, & \text{if } c_3 \frac{N}{q_I} (\log T)^2 < T \leq c_4 \frac{Nq_I}{\log N} \\
\frac{\log N}{N} T, & \text{if } T > c_4 \frac{Nq_I}{\log N}. 
\end{cases}$$

If there is no structure in the user space and the number of item types is $q_U = N^\beta$ for fixed $0 < \beta$, then the item-item CF algorithm is optimal up to a logarithmic multiplicative constant.

**Theorem 7.1** (Regret lower bound in item structure only, simplified version). *In the item-structure model with $q_I = N^\beta$ item types and $N > N_0(\beta)$ users, there exist numerical constants $c_1, c_2, c_3,$ and $c_4$ such that any recommendation algorithm must incur regret

$$\text{regret}(T) > \begin{cases} 
 c_1 \sqrt{\frac{q_I}{N}} T, & \text{if } T \leq c_4 \frac{Nq_I}{(\log q_I)^2} \\
 c_2 \log \frac{q_I}{N} T, & \text{if } T > c_4 \frac{Nq_I}{(\log q_I)^2}. 
\end{cases}$$

The proof of the lower bound is based on two main observations. First, if an item has been recommended to fewer than $\log q_I$ users, then its similarity with respect to other items cannot be determined; this implies that recommending this item to any user has uncertain outcome. Second, when a user is recommended an item from a given item type for the first time, the outcome of that recommendation is uncertain. Lower bounding the number of such uncertain recommendations gives the lower bound for regret.

Algorithms using structure in the item space do not suffer from a long cold-start time like those using only user structure: even for a very short time horizon, they can guarantee nontrivial bounds on regret. In particular, the near-optimal algorithm proposed here suffers from a constant value of regret for an initial period. Note that as $N$ increases, the regret upper bound (given in Theorem 6.1) in the initial phase (constant $c_1$) does not change, but the length of the initial phase increases. This implies that larger $N$ helps making meaningful recommendations easier. The same observation is true for any time horizon: the upper bound of regret at any time $T$ is a decreasing function of $N$.

## 4 User-user algorithm and analysis

In this section, we propose and analyze a version of user-user CF within the latent variable model introduced in Section 2.

### 4.1 Algorithm

We now describe algorithm USER-USER (see Algorithm 1). In Step 1, random items are recommended to all of the users. The ratings of these items are used to construct a partition over the users that recovers the user types correctly with high probability. In Step 2, users are recommended new random items until there is an item that is liked. If the partition agrees with the user types, then this item is also liked by everyone in the partition. The liked item is then recommended to all other users in the same partition (*exploitation*). Step 2 (find and recommend items) is repeated indefinitely.

We make a few remarks regarding the algorithm:
Algorithm 1 User-User($T, q_U, N$)

**Step 1: partition users**
1. $\epsilon \leftarrow 1/N$, $r \leftarrow \lceil 2\log(q_U^2/\epsilon) \rceil$
2. for $t = 1, \ldots, r$ do
3. Pick random item $i$
4. $a_{u,t} \leftarrow i$, for all $u \in [N]$
5. Partition users into fewest possible groups such that each group agrees on all items. Let $\tilde{\tau}_U(u) \in [q_U]$ be the label of user $u$’s partition.
6. $P_k = \{u \in [N] : \tilde{\tau}_U(u) = k\}$, for all $k \in [q_U]$

**Step 2: find and recommend items**
7. $S_k \leftarrow \emptyset$, for all $k \in [q_U]$
8. for $t = r + 1, \ldots, T$ do
9. for $u \in [N]$ do
10. if $S_{\tilde{\tau}_U(u)} \setminus \{a_{u,1}, \ldots, a_{u,t-1}\} \neq \emptyset$ (i.e., $u$ has not rated all of $S_{\tilde{\tau}_U(u)}$) then
11. $a_{u,t} \leftarrow$ an unrated item in $S_{\tilde{\tau}_U(u)}$ (exploit)
12. else
13. $a_{u,t} \leftarrow$ random item not rated by any user in $P_{\tilde{\tau}_U(u)}$ (explore)
14. if $L_{u,a_{u,t}} = +1$ then
15. $S_{\tilde{\tau}_U(u)} \leftarrow S_{\tilde{\tau}_U(u)} \cup \{a_{u,t}\}$

- Our model assumes that users with the same type rate all items similarly. Hence, users of the same type are always in the same group after the partitioning. In contrast, due to random sampling of the items in exploration, users from two (or more) different types might have identical ratings for the items in Step 1 of the algorithm, which results in an erroneous partition with distinct user types being grouped together.

- The above remark implies that the total number of groups in the user partitioning is no more than $q_U$.

- The labeling of user groups in the partitioning step is arbitrary (and may be different from the similarly arbitrary labeling of user types).

- In Step 2 of the algorithm, the sets of items $\{S_k\}$ at each time contain the items exploitable by users in the $k$th group of users in the partition. The algorithm predicts that all users in the $k$th group like items in $S_k$.

**Theorem 4.1.** Consider the model introduced in Section 2 with $N$ users, $q_U$ user types and $q_I$ item types. Let $r = \lceil 2\log(q_U^2/\epsilon) \rceil$ with $\epsilon = 1/N$. If $q_I > 4r$, then Algorithm 1 achieves regret

\[
\text{regret}(T) \leq \begin{cases} 
\frac{1}{2}T, & \text{if } T \leq r \\
\frac{1}{2}r + \frac{2(q_I+1)}{N}T + 2, & \text{if } T > r.
\end{cases}
\]

The above theorem states that up until time $r$, the algorithm is making meaningless recommendations. After that, the algorithm achieves the asymptotic slope performance for which on average, $q_U$ recommendations out of $N$ are random.

The simplified version of this theorem in Section 3 is obtained using $q_U \leq N$ which gives $\log N < r < 3\log N + 1 < 4\log N$. 

8
4.2 Proof of Theorem 4.1

We first bound the probability that the partition created by the algorithm is correct.

Lemma 4.2. Let $B_{uv} = \{ \mathbb{1}_{\{\hat{\tau}_U(u) = \hat{\tau}_U(v)\}} = \mathbb{1}_{\{\tau_U(u) = \tau_U(v)\}} \}$ be the event that users $u$ and $v$ are partitioned correctly with respect to each other in Step 1 of Algorithm User-User. If $q_T > 4r$, then $\mathbb{P}[B_{uv}^c] \leq 2\epsilon / q_T^2$. It follows that if $B = \bigcap B_{uv}$ is the event that all users are partitioned correctly, then $\mathbb{P}[B] \leq \epsilon$.

Proof. As observed earlier, users from the same partition rate items identically and therefore the only way an error in partitioning occurs is in grouping together users of different types. In Step 1, the first $r$ items recommended to all users are chosen uniformly at random independent of feedback.

Hence, the types of these items are uniformly distributed on $[q_T]$. Let $s$ be the number of items with distinct item types among the $r$ items rated by all users in Step 1. This is a balls and bins scenario with $r$ balls into $q_T$ bins, and Lemma A.3 states that if $q_T > 4r$, then $\mathbb{P}[s < r/2] \leq \exp(-r/2) \leq \epsilon / q_T^2$. By symmetry, the types of the $s$ items with distinct types are uniformly distributed on $[q_T]$.

Since all users rate the same items and users of the same type have identical preferences, as far as the lemma is concerned we may think of the user types themselves rating items in Step 1. Two user types $k \neq k'$ rate $s$ independently chosen items of distinct types in the same way with probability $2^{-s}$. On the event $s \geq r/2$, we have $2^{-s} < 2^{-r/2} \leq \epsilon / q_T^2$.

The above two statements show that for users $u$ and $v$ with $\tau_U(u) = k$ and $\tau_U(v) = k'$,

$$\mathbb{P}[(B_{uv})^c] \leq \mathbb{P}[s < r/2] + \mathbb{P}[(B_{uv})^c | s \geq r/2] \leq 2\epsilon / q_T^2.$$  

The second statement in the lemma follows by the union bound over $(q_T^2 / 2) \leq q_T^2 / 2$ pairs of user types. \hfill \Box

Proof of Theorem 4.1. For $t \leq r$, the algorithm recommends random items chosen independently of feedback to all users. So at these times $\mathbb{P}[L_{u,a,u,t} = -1] = 1/2$ for all users $u \in [N]$. Hence for time horizon $T \leq r$,

$$\mathbb{E} \left[ \sum_{t=1}^{T} \frac{1}{N} \sum_{u=1}^{N} \mathbb{1}_{L_{u,a,u,t} = -1} \right] = \sum_{t=1}^{T} \frac{1}{N} \sum_{u=1}^{N} \mathbb{P}[L_{u,a,u,t} = -1] = \frac{T}{2}. \tag{2}$$

Now suppose $T > r$. At $t = r$, by Lemma 4.2, the partitioning step recovers the user types correctly with probability at least $1 - \epsilon$, i.e., $\mathbb{P}[B] \geq 1 - \epsilon$. On event $B$ all users in a partition have the same type. For that reason, items in $S_{\tau_U(u)}$ are liked by at least one user of the same type as $u$ and therefore also by $u$, and

$$\mathbb{E} \left[ \sum_{t=r+1}^{T} \sum_{u \in [N]} \mathbb{1}_{L_{u,a,u,t} = -1, a_{u,t} \in S_{\tau_U(u)}} \big| B \right] = 0.$$  

Because there are $TN$ terms in the sum and $\mathbb{P}[B^c] \leq \epsilon$, it follows by law of total probability that

$$\mathbb{E} \left[ \sum_{t=r+1}^{T} \sum_{u \in [N]} \mathbb{1}_{L_{u,a,u,t} = -1, a_{u,t} \in S_{\tau_U(u)}} \right] \leq TN \epsilon. \tag{3}$$

Now, we need to find an upper bound for the expected number of disliked exploration recommendations in Step 2 of the algorithm, i.e., $\mathbb{E} \left[ \sum_{t=r+1}^{T} \sum_{u \in [N]} \mathbb{1}_{L_{u,a,u,t} = -1, a_{u,t} \notin S_{\tau_U(u)}} \right]$.  

9
It will be useful to relate the expected number of liked and disliked explorations, and to this end we consider the event that each user type likes at least 1/3 of the item types: let

$$C = \left\{ \sum_{j \in [q_t]} \mathbb{1}[\xi_{k,j} = +1] > q_t/3 \text{ for all } k \in [q_t] \right\}.$$  

Partitioning according to $C$, we get

$$
\mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1} \left[ L_{u,a,u,t} = -1, a_{u,t} \notin S_{U}(u) \right] \right] 
\leq \mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1} \left[ L_{u,a,u,t} = -1, a_{u,t} \notin S_{U}(u), C \right] \right] + NTP(C^c). \tag{4}
$$

The second term is controlled by an application of the Chernoff bound (Lemma A.1), which gives $P[C^c] \leq q_t \exp(-q_t/36)$.

To obtain an upper bound for the first term, in Claim 4.3 below we will upper bound the expected number of exploration recommendations that were liked, and on event $C$ this will provide also an upper bound for the expected number of exploration recommendations which were disliked. The number of liked explorations is easier to deal with, because these result in items added to sets $\{S_k\}$ for exploitation, and exploration only happens when there are not enough items to be exploited.

We now relate the number of liked and disliked explorations. At $t > r$ if $a_{u,t} \notin S_{U}(u)$, then it means the item is an exploratory recommendation and thus $a_{u,t}$ is an independent new random item with uniformly random type $\tau_I(a_{u,t})$ on $[q_t]$. Hence,

$$P \left[ L_{u,a,u,t} = +1 | a_{u,t} \notin S_{U}(u), C \right] \geq 1/3,$$

and

$$P \left[ L_{u,a,u,t} = -1 | a_{u,t} \notin S_{U}(u), C \right] \leq 2P \left[ L_{u,a,u,t} = +1 | a_{u,t} \notin S_{U}(u), C \right].$$

This means that to bound the first term in (4) it suffices to bound the contribution from the sum with $L_{u,a,u,t} = +1$, as achieved in the following claim.

**Claim 4.3.** On event $C$, the number of liked ‘explore’ recommendations (line 13 of Algorithm 1) by time $T$ can be bounded as

$$\sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a,u,t} = +1, a_{u,t} \notin S_{U}(u), C] \leq Tq_t + N.$$

**Proof.** For user partition $k$ and time $t$, define $S_k^T$ to be the set of items denoted by $S_k$ in the algorithm at time $t$, after making the time-step $t$ recommendations. Item $a_{u,t}$ is added to $S_k$ precisely on the event $\{t > r, \hat{\tau}_U(u) = k, a_{u,t} \notin S_k, L_{u,a,u,t} = +1\}$. Therefore,

$$\sum_{t=r+1}^T \sum_{u: \hat{\tau}_U(u) = k} \mathbb{1}[L_{u,a,u,t} = +1, a_{u,t} \notin S_k, C] \leq \sum_{t=r+1}^T \sum_{u: \hat{\tau}_U(u) = k} \mathbb{1}[L_{u,a,u,t} = +1, a_{u,t} \notin S_k] = |S_k^T| \tag{5}.$$
The rest of the proof entails bounding $|S^T_k|$. The number of items added to $S_k$ at time $t$ is

$$|S^t_k| - |S^{t-1}_k| = \sum_{u: \tau_U(u) = k} 1[L_{u,a_{u,t}} = +1, a_{u,t} \notin S_k] \leq \sum_{u: \tau_U(u) = k} 1[a_{u,t} \notin S_k].$$  

(6)

If $|S^{t-1}_k| \geq t$, then $S^{t-1}_k \setminus \{a_{u,1}, \ldots, a_{u,t-1}\} \neq \emptyset$. Meanwhile, the exploration event (i.e., recommending $a_{u,t} \notin S_{\tau_U(u)}$) in Step 13 happens only if there are no items left in $S_{\tau_U(u)}$ for user $u$ to exploit, i.e., $S^{t-1}_{\tau_U(u)} \setminus \{a_{u,1}, \ldots, a_{u,t-1}\} = \emptyset$. Consequently,

$$\sum_{u: \tau_U(u) = k} 1[a_{u,t} \notin S_{\tau_U(u)}] \begin{cases} = 0, & \text{if } |S^{t-1}_k| \geq t \\ \leq |P_k|, & \text{otherwise.} \end{cases}$$

(7)

The bound $|P_k|$ is due to the sum having $|P_k|$ terms, each upper bounded by 1.

Let $t^* = \max\{t : r \leq t < T, |S^t_k| < t\}$. Note that the set over which we take the maximum is nonempty if $T > r$ since $|S^r_k| = 0$. By definition, $|S^{t^*_r}_k| < t^* < T$. Since for $t > t^*$ we have $|S^t_k| \geq t$, by (6) and (7), for $t^* < t < T$ we have $|S^{t+1}_k| - |S^t_k| = 0$. It follows that

$$|S^T_k| = |S^t_k| + \sum_{t=t^*}^{T-1} (|S^{t+1}_k| - |S^t_k|) \leq |S^t_k| + (|S^{t^*_r+1}_k| - |S^{t^*_r}_k|) \leq T + |P_k|.$$

Using (5) and summing the last displayed inequality over the (at most) $q_U$ partition indices proves the claim.

We can now complete the proof of Theorem 4.1. By the preceding claim, (3), and (4), we get the bound

$$\mathbb{E} \left[ \sum_{t=r+1}^{T} \sum_{u \in [N]} 1[L_{u,a_{u,t}} = -1] \right] = \mathbb{E} \left[ \sum_{t=r+1}^{T} \sum_{u \in [N]} 1[L_{u,a_{u,t}} = -1, a_{u,t} \in S_{\tau_U(u)}] \right]$$

$$+ \mathbb{E} \left[ \sum_{t=r+1}^{T} \sum_{u \in [N]} 1[L_{u,a_{u,t}} = -1, a_{u,t} \notin S_{\tau_U(u)}] \right]$$

$$\leq TN_\epsilon + 2(q_U T + N) + TNq_U \exp\left(-\frac{q_U}{36}\right)$$

$$\leq 2(q_U T + N) + 2T,$$

where in the last step we used $q_U \geq 18r \geq 36 \log(q_U, N)$ and the choice $\epsilon = 1/N$. For $T > r$, we can now bound the regret by combining (2) with the previous display:

$$\text{regret}(T) = \mathbb{E} \left[ \sum_{t=1}^{r} \frac{1}{N} \sum_{u=1}^{N} 1[L_{u,t} = -1] \right] + \mathbb{E} \left[ \sum_{t=r+1}^{T} \frac{1}{N} \sum_{u=1}^{N} 1[L_{u,a_{u,t}} = -1] \right]$$

$$\leq \frac{1}{2} r^{1/2} + \frac{2(q_U + 1)}{N} T + 2.$$

5 User structure only: lower bound

In this section we prove a lower bound on the regret of any online recommendation system in the situation with user structure only described in Definition 2.1.
Theorem 5.1. Let $\delta > 0$ and define $r = \lfloor \log q_U - \log \left( 6 \log q_U \log \frac{2N \log q_U}{\delta} \right) \rfloor$. In the user structure model with $N$ users and $q_U$ user types, any recommendation algorithm must incur regret

$$\text{regret}(T) \geq \begin{cases} 
(\frac{1}{2} - \delta)T - 4, & \text{for } T \leq r \\
[1 - \exp(-N/q_U)] \frac{q_U}{2N} T, & \text{for all } T.
\end{cases}$$

The simplified version of this theorem in Section 3 is derived using $N > q_U$ and $\lfloor \log q_U - 3 \log \log N - 5 \rfloor < r < \lfloor \log q_U - \log \log N \rfloor$ for $\delta = 1/100$.

5.1 Proof strategy

At a high level, the lower bound is based on two observations:

- **A good estimate of user types is necessary in order to make meaningful recommendations and estimating similarity between users requires approximately $\log q_U$ items rated in common.**

  Suppose that the preference matrix is known and suppose that we have obtained feedback from some user $u$ for $t - 1$ items. Relative to the total number of types, user $u$ must belong to a restricted set of user types consistent with this feedback. If $t - 1$ is small, the set of possible types is large (for instance, if a user has rated only one item, there are roughly $q_U/2$ candidate user types for this user). Now, user $u$ likes some item $i$ with probability proportional to the number of consistent types liking the item. Control of this count amounts to a property of the matrix we call $(t, \epsilon)$-column regularity in Definition 5.1, which holds with high probability.

- **Even if we know the user types (i.e., clustering of users), the first time a given item is recommended to a user from a given type, the outcome is uniformly random.**

  Learning the preference of a user type for an item is achieved by recommending the item to one user from the user type. One way to observe this is that the random variable $\xi_{\tau_U(u), \tau_I(i)}$ in the preference matrix is independent of all previous history in the situation described.

5.2 Proof of Theorem 5.1

We separately prove the two lower bounds in the maximum, starting with the first. The following regularity property in submatrices of the preference matrix allows us to control the posterior probability for an item being liked.

**Definition 5.1** $(t, \epsilon)$-column regularity. Let $A \in \{-1, +1\}^{m \times n}$. For ordered tuple of distinct (column) indices $w = (i_1, \ldots, i_t) \in [n]^t$, let $M = (A_{i_j})_{i \in w} \in \{-1, +1\}^{m \times r}$ be the matrix formed from the columns of $A$ indexed by $w$. For given row vector $b \in \{-1, +1\}^r$, let $K_{b, w}(A) \subseteq [m]$ be the set of rows in $M = (A_{i_j})_{i \in w}$ which are identical to the row $b$. The cardinality of $K_{b, w}(A)$ is denoted by $k_{b, w}(A)$. $A$ is said to be $(r, \epsilon)$-column regular if

$$\max_{w, b} \left| k_{b, w}(A) - \frac{m}{2^r} \right| \leq \epsilon \frac{m}{2^r},$$

where the maximum is over tuples $w$ of $r$ columns and $\pm 1$ vectors $b$ of size $r$.

We define $\Omega_{r, \epsilon}$ to be the set of $(r, \epsilon)$-column regular matrices.

**Claim 5.2**. If a matrix $A \in \{-1, +1\}^{m \times n}$ is $(r, \epsilon)$-column regular, then it is also $(s, \epsilon)$-column regular for all $s < r$. 

12
Proof. Suppose that $A$ is $(r, \epsilon)$-column regular. By induction it suffices to show that $A$ is $(r - 1, \epsilon)$-column regular. We will check that $(1 - \epsilon) \frac{m}{2^r} \leq k_{b,w}(A) \leq (1 + \epsilon) \frac{m}{2^r}$ for all size $r - 1$ tuples $w$ and vectors $b$. For any given $w \in [n]^r - 1$, let $b^+ = [b 1] \in \{-1, +1\}^r$ be obtained from $b$ by appending $+1$. Similarly $b^-$ is obtained from $b$ by appending $-1$. If $w' = (w, i) \in [n]^r$ for any $i \notin w$, then $K_{b,w} = K_{b^+,-w'} \cup K_{b^-,-w'}$ and $K_{b^+,-w'} \cap K_{b^-,-w'} = \emptyset$, so $k_{b,w} = k_{b^+,-w'} + k_{b^-,-w'}$. Since $A$ is $(r, \epsilon)$-column regular, $(1 - \epsilon) \frac{m}{2^r} \leq k_{b^+,-w'} + k_{b^-,-w'} \leq (1 + \epsilon) \frac{m}{2^r}$, hence $(1 - \epsilon) \frac{m}{2^r} \leq k_{b,w} \leq (1 + \epsilon) \frac{m}{2^r}$. \hfill $\square$

Lemma 5.3. Let matrix $A \in \{-1, +1\}^{m \times n}$ have i.i.d. Bern($1/2$) entries. Then for $\epsilon < 1$, $A$ is $(r, \epsilon)$-column regular with probability at least

$$1 - 2(2n)^r \exp \left( -\frac{\epsilon^2 m}{3 \cdot 2^r} \right).$$

Proof. Note that for given column tuple $w$ and row vector $b$, the expected number of times the row vector $b$ appears is $\frac{m}{2^r}$. Using a Chernoff bound (Lemma A.1) with $\epsilon < 1$,

$$\mathbb{P}\left[ k_{b,w}(A) - \frac{m}{2^r} \geq \frac{m}{2^r} \right] \leq 2 \exp \left( -\frac{\epsilon^2 m}{3 \cdot 2^r} \right).$$

There are no more than $n^r$ possible choices of column tuple $w$, and $2^r$ possible choices of row vector $b$; the union bound yields the proof. \hfill $\square$

Proposition 5.4. Let $\delta > 0$ and define $r = \lceil \log q_U - \log(6 \log q_U \log(2N \log q_U)) \rceil$. For any $T \leq r$, the regret is lower bounded by

$$\text{regret}(T) \geq (1/2 - \delta) T - 4.$$ 

Proof. We will show that for preference matrices satisfying column regularity, at any time $t \leq r$, most users have probability roughly half of liking any particular item given the feedback obtained thus far, even if the preference matrix is known. (Recall that the preference matrix contains the preference of each user type for each item type; there is still uncertainty in the actual type of each user or item.)

At time $t$, suppose that $n$ items in total have been sampled by the algorithm ($n \leq Nt$ since each of the $N$ users can rate one item per time-step). We label the set of items to be $[n] = \{1, \ldots, n\}$ since a priori the items are identical. Let $A$ be the $q_U \times n$ matrix indicating the preference of each user type for these $n$ items. Each item $i$ has type $\tau_U(i) \sim \text{Unif}([2^q])$ and because the set of columns of the preference matrix $\Xi$ is precisely $\{-1, +1\}^q$, the columns of $A$ are independent and uniformly distributed in $\{-1, +1\}^q$.

We now focus on a particular user $u$. Let $w = \{a_u,s\}_{s \in [t-1]}$ be the items recommended to user $u$ up to time $t - 1$, and let $b = (L_{u,a_u,s})_{s \in [t]} \in \{-1, +1\}^{t-l}$ be the vector of feedback from user $u$ for these items. We claim that conditional on the matrix $A$, vectors $b$ and $w$, the type $\tau_U(u)$ of user $u$ at the end of time instant $t - 1$ is uniformly distributed over the set of user types $K_{b,w}(A)$ consistent with this data ($K_{b,w}(A)$ is defined in Definition 5.1).

Let $b^+ = [b 1] \in \{-1, +1\}^t$ be obtained from $b$ by appending $+1$. $L_{u,a_u,t} = +1$ precisely when $\tau_U(u) \in K_{b^+,(w,a_u,t)}(A)$, which in words reads “user $u$ is among those types that are consistent with the first $t - 1$ ratings of $u$ and have preference vector with ‘+1’ for the item recommended to $u$ at time $t$”. It follows that for any matrix $A$ corresponding to items $[n]$,

$$\mathbb{P}[L_{u,a_u,t} = +1|H_{t-1}, A] = \mathbb{P}[\tau_U(u) \in K_{b^+,(w,a_u,t)}(A)|H_{t-1}, A] = \frac{k_{b^+,(w,a_u,t)}(A)}{k_{b,w}(A)}.$$ 

13
The second equality is due to: i) $w$, $b$, and $a_{u,t}$ are functions of $\mathcal{H}_{t-1}$; ii) for fixed $w$ and $b$ the set $K_{b,w}(A)$ is determined by $A$; iii) $\tau_U(u)$ is uniformly distributed on $K_{b,w}(A)$ conditional on $A, b, w$.

Recall that we define $\Omega_{t, \epsilon}$ to be the set of $(t, \epsilon)$-column regular matrices. It now follows by the tower property of conditional expectation that

$$
P[L_{u,a_{u,t}} = +1 | A \in \Omega_{t, \epsilon}] = E[P[L_{u,a_{u,t}} = +1 | A, \mathcal{H}_{t-1}] | A \in \Omega_{t, \epsilon}]$$

$$= E\left[\frac{k_{b^+, \langle w, i \rangle}(A)}{k_{b, w}(A)} | A \in \Omega_{t, \epsilon}\right]$$

$$= \frac{1 + \epsilon}{2 - 1 - \epsilon} \leq \frac{1}{2} (1 + 4\epsilon).$$

The last two equalities are justified as follows: if $A \in \Omega_{t, \epsilon}$ then by Claim 5.2, $A \in \Omega_{t-1, \epsilon}$. By Definition 5.1, this means that $k_{b, w}(A) \geq (1 - \epsilon)m/2^t - 1$ and $k_{b^+, \langle w, i \rangle}(A) \leq (1 + \epsilon)m/2^t$. We pick $\epsilon < 1/2$ to get the last inequality.

Fix $\delta > 0$ and

$$\epsilon = \sqrt{\frac{2t}{q_U} \log \frac{(2N \log q_U)^t}{\delta}}.$$ 

Lemma 5.3 shows that at time $t \leq r$, $A \in \Omega_{t, \epsilon}$ for this choice of $\epsilon$, with probability at least $1 - \delta$. We get the bound

$$P[L_{u,a_{u,t}} = +1] \leq P[L_{u,a_{u,t}} = +1, A \in \Omega_{t, \epsilon}] + P[A \notin \Omega_{t, \epsilon}] \leq \frac{1}{2} (1 + 4\epsilon) + \delta.$$

From the last display it follows that for $T \leq r$,

$$\text{regret}(T) = \frac{1}{N} \sum_{t \in [T], u \in [N]} P[L_{u,a_{u,t}} = -1]$$

$$\geq (\frac{1}{2} - \delta)T - \sum_{t \in [T]} 2^{t} \sqrt{\frac{2t}{q_U} \log \frac{(2N \log q_U)^t}{\delta}}$$

$$\geq (\frac{1}{2} - \delta)T - 4 \sqrt{3 \frac{2T+1}{q_U} \log \frac{(2N \log q_U)^t}{\delta}}$$

$$\geq (\frac{1}{2} - \delta)T - 4. \quad \Box$$

**Proposition 5.5.** For any $T$, $\text{regret}(T) \geq \left[1 - \exp(-N/q_U)\right] \frac{q_U}{2N} T$.

The main ingredient in the proof of the proposition is definition of an event that implies the associated recommendation is uniformly random. Let $\mathcal{B}^{t}_{\tau_U(u), i}$ be the event that some user of same type $\tau_U(u)$ as $u$ has rated item $i$ by time $t - 1$:

$$\mathcal{B}^{t}_{\tau_U(u), i} = \{ \exists v \in [N] \setminus \{ u \}, s \in [t - 1] : \tau_U(v) = \tau_U(u), a_{v,s} = i \}.$$  

Note that $\mathcal{B}^{t}_{\tau_U(u), i}$ is a function of $\mathcal{H}_{t-1}$ and the set of user types $(\tau_U(u))_{u \in [N]}$.

**Claim 5.6.** If no user with the same type as $u$ has rated item $i$ by time $t - 1$, the probability that user $u$ likes item $i$ conditional on any history consistent with this is $P[L_{u,i} = -1 | (\mathcal{B}^{t}_{\tau_U(u), i})^c, \mathcal{H}_{t-1}] = \frac{1}{2}$, or equivalently,

$$P[L_{u,a_{u,t}} = +1, (\mathcal{B}^{t}_{\tau_U(u), a_{u,t}})^c | \mathcal{H}_{t-1}] = \frac{1}{2} \cdot P[(\mathcal{B}^{t}_{\tau_U(u), a_{u,t}})^c | \mathcal{H}_{t-1}].$$
Proof. Let $\tau_U(\cdot) = (\tau_U(u))_{u \in [N]}$ be the vector of all of the user types. Let $A$ be the matrix of size $q_U \times n$ of the ratings of all user types for all the $n$ items observed by time $t$ (note that $n \leq Nt$). Denote by $A_{\{k,i\}}$ the matrix obtained from $A$ by erasing the entry corresponding to the preference of user type $k$ for item $i$.

To ease notation slightly, write $B = B_{\tau_U(u),a,u,t}$. By the tower property,

$$
P[L_{u,a,u,t} = +1, B^c | \mathcal{H}_{t-1}] = \mathbb{E}[P[L_{u,a,u,t} = +1, B^c | \tau_U(\cdot), A_{\{\tau_U(u),a,u,t\}}, \mathcal{H}_{t-1}] | \mathcal{H}_{t-1}].
$$

It suffices to prove that $L_{u,a,u,t}$ is $\pm 1$ w.p. $1/2$ each, conditional on any values $a_{u,t} = i, \tau_U(\cdot), A_{\{k,i\}},$ and $\mathcal{H}_{t-1}$ for which $B^c$ holds.

We will use the lack of structure in the item space, which manifests in all possible columns $\{-1,+1\}^{q_U}$ appearing (once each) in the preference matrix $\Xi$. To start, by definition of the $A$ matrix, $L_{u,i} = A_{\tau_U(u),i}$. Because all possible columns appear in the preference matrix $\Xi$, we may identify the item type indices $\tau_I(\cdot)$ with the corresponding binary vector in the column. A priori $\tau_I(i)$ is uniform on $[q_I]$, equivalently $\{-1,+1\}^{q_U}$, and conditional on $A_{\{\tau_U(u),i\}}$ the type $\tau_I(i)$ is uniform over the two columns consistent with $A_{\{\tau_U(u),i\}}$ having entries $\pm 1$ in $A_{\tau_U(u),i}$. \hfill \Box

The final ingredient in the proof of Proposition 5.5 is a lower bound on the number of items recommended for which Claim 5.6 applies.

Claim 5.7. The total number of times a new item is recommended to a user type by time $T$ is lower bounded as

$$
\mathbb{E}\left[\sum_{t \in [T], u \in [N]} \mathbbm{1}[(B_{\tau_U(u),a,u,t}^t)^c]\right] \geq q_U T \left[1 - \left(1 - \frac{1}{q_U}\right)^N\right].
$$

Proof. At the end of time-step $T$ each user has been recommended $T$ items, hence each user type has been recommended at least $T$ items. Let $\widetilde{q}_U$ be the number of user types in which there is at least one user. The total number of times an item is recommended to a user type for the first time is at least $\widetilde{q}_U T$:

$$
\sum_{t \in [T], u \in [N]} \mathbbm{1}[(B_{\tau_U(u),a,u,t}^t)^c] \geq \widetilde{q}_U T.
$$

Lemma A.4 shows that $\mathbb{E}[\widetilde{q}_U] \geq q_U (1 - (1 - 1/q_U)^N).$ \hfill \Box

We now complete the proof of Proposition 5.5.

Proof of Prop 5.5. Partitioning recommendations according to $B_{\tau_U(u),a,u,t}^t$ gives

$$
N[T - \text{regret}(T)] = \mathbb{E}\left[\sum_{t \in [T], u \in [N]} \mathbbm{1}[L_{u,a,u,t} = +1]\right]
$$

$$
= \mathbb{E}\left[\sum_{t \in [T], u \in [N]} P[L_{u,a,u,t} = +1, B_{\tau_U(u),a,u,t}^t | \mathcal{H}_{t-1}]\right]
$$

$$
+ \mathbb{E}\left[\sum_{t \in [T], u \in [N]} P[L_{u,a,u,t} = +1, (B_{\tau_U(u),a,u,t}^t)^c | \mathcal{H}_{t-1}]\right]
$$

$$
= \mathbb{E}\left[\sum_{t \in [T], u \in [N]} P[L_{u,a,u,t} = +1, B_{\tau_U(u),a,u,t}^t | \mathcal{H}_{t-1}]\right]
$$
By Claim 5.6 we may bound this as
\[
\begin{align*}
&\leq \mathbb{E}\left[ \sum_{t \in [T], u \in [N]} \left( \mathbb{P}[(B^t_{\tau_U(u),a_{u,t}} | \mathcal{H}_{t-1}) + \frac{1}{2} \mathbb{P}[(B^t_{\tau_U(u),a_{u,t}})^c | \mathcal{H}_{t-1})] \right) \\
&= NT - \frac{1}{2} \mathbb{E}\left[ \sum_{t \in [T], u \in [N]} \mathbb{P}[(B^t_{\tau_U(u),a_{u,t}})^c | \mathcal{H}_{t-1})] \right] \\
&= NT - \frac{1}{2} \mathbb{E}\left[ \sum_{t \in [T], u \in [N]} \mathbf{1}[(B^t_{\tau_U(u),a_{u,t}})^c] \right] .
\end{align*}
\]

We now apply Claim 5.7 to the second term, to get the upper bound
\[
\begin{align*}
&\leq NT - \frac{T}{2} q_U \left[ 1 - (1 - \frac{1}{q_U})^N \right] \leq NT - \frac{T}{2} q_U \left[ 1 - \exp\left(\frac{-N}{q_U}\right) \right] .
\end{align*}
\]

Rearranging shows that \( \text{regret}(T) \geq \left[ 1 - \exp\left(\frac{-N}{q_U}\right) \right] \frac{3q_U}{2N} T \) for all \( T \). \( \square \)

### 6 Item-item algorithm and analysis

This section describes a version of item-item CF tailored to the model described in Section 2 and analyzes its performance.

#### 6.1 Algorithm

Algorithm Item-Item performs the following steps (see Algorithm 2). First, items are partitioned according to type; next, each user’s preference for each item type is determined; finally, items from liked partitions are recommended. These steps are now described in a bit more detail.

Two sets \( \mathcal{M}_1 \) and \( \mathcal{M}_2 \), each containing \( M \) random items, are selected. In the exploration step, each item is recommended to \( r = \lceil 2 \log(q_U/\epsilon) \rceil \) random users. The feedback from these recommendations later helps to partition the items according to type. The parameter \( r \) is chosen large enough to guarantee small probability of error in partitioning (\( \leq \epsilon \) for each item). The use of two sets \( \mathcal{M}_1 \) and \( \mathcal{M}_2 \), as opposed to just one, is to simplify the analysis; as described next, item type representatives are selected from \( \mathcal{M}_1 \), which are used to represent clusters of items from \( \mathcal{M}_2 \).

An item from each of \( \ell \) explored types is recommended to all users, where \( \ell \) is a parameter determined by the algorithm. It turns out that it is often beneficial (depending on system parameters) to learn user preferences for only a subset of the types, in which case \( \ell \) is strictly less than \( q_U \). Each of the \( \ell \) items chosen from \( \mathcal{M}_1 \) is thought of as a representative of its type. For each representative item \( i_j, j = 1, \ldots, \ell \), all items in \( \mathcal{M}_1 \) that appear to be of the same type as \( i_j \) are stored in a set \( \mathcal{S}_j \) and removed from \( \mathcal{M}_1 \). This guarantees that at each time \( \mathcal{M}_1 \) does not contain items with the same type as any of the previous selected representative items. The same is done for \( \mathcal{M}_2 \), with sets of items appearing similar to \( i_j \), denoted by \( \mathcal{S}_j \).

For each user \( u \), items in groups \( \mathcal{S}_j \) with liked representative \( i_j \) are added to a set of exploitable items \( \mathcal{R}_u \). Finally, in the exploitation phase, each user is recommended items from \( \mathcal{R}_u \). The number \( M \) of items in each of \( \mathcal{M}_1 \) and \( \mathcal{M}_2 \), is chosen as a function of \( \ell \) to make sure that there are enough exploitable items in \( \mathcal{R}_u \) for all users \( u \) for the entire length-\( T \) time-horizon, and \( \ell \) is then chosen to minimize regret.

The algorithm description uses some notation: For an item \( i \) and time \( t > 0 \),
\[
\text{rated}_t(i) = \{ u \in [N] : a_{u,s} = i \text{ for some } s < t \}
\]
is the set of users that have rated item $i$ before time $t$. The time $t$ is implicit in the algorithm description, with $\text{rated}(i)$ used to represent $\text{rated}_t(i)$ at the time of its appearance.

**Theorem 6.1.** Consider the model introduced in Subsection 2 with $q_I, q_U > 18 \log(3N)$. Let $r = \left[2 \log(2N q_I^2)\right]$. Algorithm 2 obtains regret per user at time $T$ upper bounded as

$$\text{regret}(T) < 4 + 57 \max\left\{ \log T, \sqrt{\frac{3q_I}{N} T}, 53 \frac{r}{N T} \right\}.$$  

The simplified version of this theorem in Section 3 is derived by using $N > q_U$ which gives $2 \log N < r < 7 \log N$ and $N > 3$.

**Algorithm 2** Item-Item($T, q_I, N$) (fixed time horizon)

1: $\ell = \max\left\{4, \min\left\{ \left\lceil \log T, \sqrt{330q_I T} \right\rceil, q_I \right\} \right\}$
2: $\epsilon = \frac{1}{2q_I N}; r = \left[2 \log \frac{q_I}{\epsilon}\right]; M = \left\lceil \frac{64q_I r}{T} \right\rceil$
3: $\mathcal{M}_1 = M$ random items; $\mathcal{M}_2 = M$ new random items;
4: $\mathcal{R}_u \leftarrow \emptyset, \ u \in [N]$ (items exploitable by user $u$)
5: $\{\mathcal{R}_u\}_{u \in [N]} \leftarrow \text{ITEMEXPLORE}(\mathcal{M}_1, \mathcal{M}_2, \ell)$
6: $\text{ITEMEXPLOIT}(\{\mathcal{R}_u\}_{u \in [N]})$

**Algorithm 3** ItemExplore($\mathcal{M}_1, \mathcal{M}_2, \ell$)

1: $\hat{S}_j \leftarrow \emptyset, \ j \in [q_I]$ (initialize sets of items of type $j$ in $\mathcal{M}_2$)
2: $\hat{S}_j \leftarrow \emptyset, \ j \in [q_I]$ (initialize sets of items of type $j$ in $\mathcal{M}_1$)
3: Recommend each item in $\mathcal{M}_1$ and $\mathcal{M}_2$ to $r$ users from $[N]$. (This is done over $\left\lceil (|\mathcal{M}_1| + |\mathcal{M}_2|)r/N \right\rceil \leq 2Mr/N + 1$ time-steps, with extra recommendations being random new items.)
4: for $j = 1, \ldots, \ell$ if $\mathcal{M}_1 \neq \emptyset$ do
5: $i_j \leftarrow$ a random item in $\mathcal{M}_1$ (representative item)
6: $a_{u,t} \leftarrow i_j$ if $u \notin \text{rated}(i_j)$, otherwise a random item not from $\mathcal{M}_1$.
7: $\hat{S}_j \leftarrow \{i \in \mathcal{M}_1 : U = \text{rated}(i) \cap \text{rated}(i_j), L_{U,i} = L_{U,i_j}\}$ (users agree on $i$ vs. $i_j$)
8: $\mathcal{M}_1 \leftarrow \mathcal{M}_1 \setminus \hat{S}_j$
9: $\mathcal{S}_j \leftarrow \{i \in \mathcal{M}_2 : U = \text{rated}(i) \cap \text{rated}(i_j), L_{U,i} = L_{U,i_j}\}$
10: $\mathcal{M}_2 \leftarrow \mathcal{M}_2 \setminus \mathcal{S}_j$
11: $\mathcal{R}_u = \bigcup_{j \in [\ell]: L_{u,i_j} = +1} \mathcal{S}_j$ for each $u \in [N]$
return $\{\mathcal{R}_u\}_{u \in [N]}$

### 6.2 Proof of Theorem 6.1

We now prove Theorem 6.1, deferring several lemmas and claims to the next subsection.
Algorithm 4 ItemExploit(\{R_u\}_{u \in [N]})

1: for remaining \( t \leq T \) do
2:   for \( u \in [N] \) do
3:     if there is an item \( i \in R_u \) such that \( u \notin rated_t(i) \). then
4:       \( a_{u,t} \leftarrow i \)
5:   else \( a_{u,t} \leftarrow \) a random item not yet rated by \( u \).

The basic error event is misclassification of an item. In Algorithm 3, \( S_j \) is the set of items that the algorithm posits are of the same type as the \( j \)th representative \( i_j \). Let \( \mathcal{E}_i \) be the event that item \( i \) was mis-classified,

\[
\mathcal{E}_i = \{ \exists j : i \in S_j, \tau_I(i) \neq \tau_I(i_j) \}.
\]

The number of time-steps spent making recommendations in Algorithm ItemExplore is denoted by \( T_0 \). We partition the recommendations made by Algorithm Item-Item to decompose the regret as follows:

\[
N_{\text{regret}}(T) = \mathbb{E} \left[ \sum_{u=1}^{N} \sum_{t=1}^{T_0} \mathbb{1}[L_{u,a_{u,t}} = -1] \right]
+ \mathbb{E} \left[ \sum_{u=1}^{N} \sum_{t=T_0+1}^{T} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \notin R_u] \right]
+ \mathbb{E} \left[ \sum_{u=1}^{N} \sum_{t=T_0+1}^{T} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in R_u, \mathcal{E}_{a_{u,t}}] \right]
+ \mathbb{E} \left[ \sum_{u=1}^{N} \sum_{t=T_0+1}^{T} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in R_u, (\mathcal{E}_{a_{u,t}})^c] \right]
=: A_1 + A_2 + A_3 + A_4 .
\]

\( A_1 \) is the regret from early time-steps up to \( T_0 \). \( A_2 \) is the regret due to not having enough items available for the exploitation phase, which is proved to be small with high probability for a good choice of \( M \). \( A_3 \) is the regret due to exploiting the misclassified items. It is small since few items are misclassified with the proper choice of \( \epsilon \) and \( r \). \( A_4 \) is the regret due to exploiting the correctly classified items. We will see that \( A_4 = 0 \).

Bounding \( A_1 \). Line 3 of Algorithm ItemExplore takes at most \( \lceil \frac{2Mr}{N} \rceil \) units of time to rate each item in \( M_1 \) and \( M_2 \) by \( r \) users each, since \( N \) users provide feedback at each time-step. After this, \( \ell \) representative items are rated by every user in the for loop (Lines 4 through 10 of ItemExplore), which takes \( \ell \) time-steps. This gives

\[
T_0 \leq \lceil \frac{2Mr}{N} \rceil + \ell \leq \frac{2Mr}{N} + \ell + 1 .
\]

For \( t \leq T_0 \), random items are recommended irrespective of feedback to all users. Since each item \( i \) has a uniformly distributed type \( \tau_I(i) \) and any given user likes type \( j \) with probability half,

\[
A_1 = \mathbb{E} \left[ \sum_{u=1}^{N} \sum_{t=1}^{T_0} \mathbb{1}[L_{u,a_{u,t}} = -1] \right] = \frac{1}{2} \sum_{u=1}^{N} T_0 \leq Mr + \frac{1}{2}(\ell + 1)N .
\]
Bounding A2. Time-steps \( t > T_0 \) are devoted to exploitation as described in ItemExploit. An item \( a_{u,t} \notin \mathcal{R}_u \) is recommended to user \( u \) only when all items in \( \mathcal{R}_u \) have already been recommended to \( u \). So, the total number of times an item \( a_{u,t} \notin \mathcal{R}_u \) is recommended at time interval \( T_0 \leq t \leq T \), is at most \((T - |\mathcal{R}_u|)_+\). Hence,

\[
A2 = E \left[ \sum_{u \in [N]} \sum_{t=T_0+1}^{T} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \notin \mathcal{R}_u] \right] \leq E \left[ \sum_{u \in [N]} \sum_{t=T_0+1}^{T} \mathbb{1}[a_{u,t} \notin \mathcal{R}_u] \right] \leq \sum_{u \in [N]} E \left[ (T - |\mathcal{R}_u|)_+ \right] \leq 3TN \exp(-\ell/18) + 148T ,
\]

where the last inequality is from Lemma 6.4.

Bounding A3. Term A3 in (9) is the expected number of mistakes made by the algorithm as a result of misclassification. Claim 6.2 upper bounds the expected number of “potential misclassifications” (defined in Equation (13)) in the algorithm to provide an upper bound for this value:

\[
A3 = E \left[ \sum_{u=1}^{N} \sum_{t=T_0+1}^{T} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in \mathcal{R}_u, \mathcal{E}_{a_{u,t}}] \right] \leq E \left[ \sum_{u=1}^{N} \sum_{t=T_0+1}^{T} \sum_{i \in [N]} \mathbb{1}[a_{u,t} \in \mathcal{R}_u, \mathcal{E}_{a_{u,t}}] \right] \leq 2NM\epsilon \leq \frac{64T}{\ell}.
\]

Bounding A4. By definition of event \( \mathcal{E}_i \) given in Equation (8), if an item \( i \in S_j \) is correctly classified, then \( \tau_I(i) = \tau_I(i_j) \). According to the model introduced in Section 2, user preferences for an item depend only on the type of the item (since \( L_{u,i} = \xi \tau_U(u), \tau_I(i) \)), so all users rate \( i \) the same as \( i_j \). For an item \( i \in \mathcal{R}_u \), there is some \( j \in [q_I] \) such that \( i \in S_j \) and \( u \) likes item \( i_j \). Hence,

\[
P[L_{u,i} = -1 | i \in \mathcal{R}_u, (\mathcal{E}_i)^c] = P[L_{u,i} = -1 | \exists j: L_{u,i_j} = +1, i \in S_j, \tau_I(i) = \tau_I(i_j)] = P[L_{u,i} = -1 | \xi \tau_U(u), \tau_I(i) = +1] = 0.
\]

It follows that \( A4 = 0 \).

Combining all the bounds. Plugging in Equations (10), (11), (12) and \( A4 = 0 \) into Equation (9) gives

\[
N_{\text{regret}}(T) \leq Mr + \frac{1}{2}(\ell + 1)N + 3TN \exp\left(-\frac{\ell}{18}\right) + 148T + \frac{64T}{\ell}.
\]

Setting \( M = \frac{64Tq_I}{\ell} \) gives

\[
\text{regret}(T) \leq (a) \frac{1}{2} + 165Tq_I r + \ell + 3T \exp\left(-\frac{\ell}{18}\right)
\leq (b) 4 + 3T\exp\left(-\frac{q_I}{18}\right) + 3\max\left\{18\log T, \sqrt{330q_I T}, \frac{330 r}{N} T\right\}.
\]
where (a) holds for $\ell \geq 4$ (which holds via definition in the algorithm) (b) is obtained by choosing the parameter $\ell$ to be

$$\ell = \max \left\{ 4, \min \left\{ \max \left\{ 18 \log T, \sqrt{\frac{330q_f^2}{N} T} \right\}, q_f \right\} \right\}.$$ 

Note that $r = \log(Nq_f^2) \leq 3 \log N$. By assumption $q_f > 18 \log(3N)$, so $3 \exp(-q_f/18) < 1/N$ and

$$\text{regret}(T) < 4 + 57 \max \left\{ \log T, \sqrt{\frac{3q_f^2}{N} T}, 53 \frac{L}{N} T \right\}. \quad \Box$$

6.3 Lemmas used in the proof

In the remainder of this section we state and prove the various lemmas used in the analysis above.

Claim 6.2. Suppose that $q_U > 4r$. The expected total number of times a misclassified item is recommended in the exploitation step is upper bounded as

$$\mathbb{E} \left[ \sum_{u=1}^{N} \sum_{t=T_0+1}^{T} \sum_{i \in \mathbb{N}} 1[a_{u,t} = i, i \in \mathcal{R}_u, \mathcal{E}_i] \right] \leq N M \epsilon.$$

Proof. Event $\mathcal{E}_i$, misclassifying item $i$ by the algorithm, occurs if the preferences of random users classifying item $i$ is the same as for a previous representative item with a different type. Given matrix $\Xi$, this event is a function of the order of choosing the representative items and the choice of random users. Instead of directly analyzing $\mathcal{E}_i$, we will define an event called "potential error event" $\tilde{\mathcal{E}}_{i,u}$, which will be shown to satisfy $\mathcal{E}_i \subseteq \tilde{\mathcal{E}}_{i,u}$; an upper bound for $\mathbb{P}[\tilde{\mathcal{E}}_{i,u}]$ is given in Claim 6.3.

For item $i \in \mathcal{M}_2$ and subset of users $U \subseteq [N]$, define

$$\tilde{\mathcal{E}}_{i,U} = \{ \exists j \neq \tau_U(i) : L_{u,i} = \xi_{\tau_U(u),j}, \text{ for all } u \in U \} \quad (13)$$

to be the event that the ratings of users in $U$ for item $i$ agree with some other type. For item $i \in \mathcal{S}_j$, if $t$ is the time $i$ is added to $\mathcal{S}_j$ in the exploration phase, let $U_i = \text{rated}_t(i) \cap \text{rated}_t(i_j)$ be the set of users whose ratings were used to conclude that $i$ and $i_j$ are of the same type. Item $i$ is added to $\mathcal{S}_j$ only if all users in $U_i$ agree on $i$ vs. $i_j$, so misclassification $\mathcal{E}_i$ (defined in Equation (8)) implies $\tilde{\mathcal{E}}_{i,U_i}$. We can now deduce the inequalities, justified below:

$$\mathbb{E} \left[ \sum_{u=1}^{N} \sum_{t=T_0+1}^{T} \sum_{i \in \mathbb{N}} 1[a_{u,t} = i, i \in \mathcal{R}_u, \mathcal{E}_i] \right] \leq \mathbb{E} \left[ \sum_{u=1}^{N} \sum_{t=T_0+1}^{T} \sum_{i \in \mathcal{M}_2} 1[a_{u,t} = i, i \in \mathcal{R}_u, \tilde{\mathcal{E}}_{i,U_i}] \right] \leq N M \epsilon. \quad \Box$$

Claim 6.3. Suppose that $q_U > 4r$. Consider the “potential error” event $\tilde{\mathcal{E}}_{i,U_i}$ defined in (13) with the set of users $U_i$ defined immediately after. Then $\mathbb{P}[\tilde{\mathcal{E}}_{i,U_i}] \leq 2\epsilon$ for all $i \in \mathcal{M}_2$.  

20
Proof. Each representative item $i_j$ chosen in ItemExplore is rated by all of the users and other items in $M_1$ and $M_2$ are rated by at least $r$ users. By Lemma A.3, if $q_T > 4r$, then with probability at least $1 - \exp(-r/2) \geq 1 - \epsilon/q_T$ there are $r/2$ users with distinct user types in a specific $U_i$. It follows that the $r/2$ users with distinct types (chosen independently of the feedback) that rate item $i$ of type $\tau_l(i)$ also have the same ratings for type $j \neq \tau_l(i)$ with probability at most $2^{-r/2}$. (Any two item types $j \neq j'$ have jointly independent user preferences $(\xi_{u,j})_{u \in [N]}, (\xi_{u,j'})_{u \in [N]}$.) The choice $r \geq 2 \log(q_T/\epsilon)$ and a union bound over item types $j$ completes the proof.

\begin{lemma}
For user $u \in [N]$, let $R_u$ be defined as in Line 10 of algorithm ItemExplore. Then

\[ \Pr[|R_u| \leq T] \leq 3 \exp\left(-\frac{\ell}{18}\right) + \frac{148}{N} \]

and

\[ \mathbb{E}[(T - |R_u|)_+ + 3T \exp\left(-\frac{\ell}{18}\right) + \frac{148}{N^2}T. \]

\end{lemma}

Proof. We begin with a sketch. Any user $u$ likes roughly half of the $\ell$ item types that have representatives $\{i_j\}_{j=1}^{\ell}$ of each type in $M_1$ from each of the item types in $M_2$. Adding over the $\ell$ types, the set $R_u$ of items in $M_2$ that user $u$ likes will typically have size at least $T$.

Making this argument rigorous requires some care for the following reasons: 1) $R_u$ is the union of the $S_j$'s with $L_{u,i_j} = +1$, but $S_j$ can be missing items due to misclassification since there may be items of the same type as $i_j$ that have been classified as belonging to the same type as some $i_j'$ for which $L_{u,i_j'} = -1$. 2) The type of each item representative depends on the number of remaining items of each type in $M_1$ when the representative is chosen. Again, due to misclassification, this can be different from the actual number of items of each type. Moreover, because misclassification of an item depends on the ratings of users for the item, the choice of next item type to be represented is therefore dependent on ratings of users for other types. The effect of this dependence is addressed in Claim 6.5 below.

We now proceed with the proof, controlling $R_u$ by introducing a different set. For $u \in [N]$ let

\[ \tilde{R}_u = \{i \in M_2 : \tau_l(i) = \tau_l(i_j), \text{ for some } j \in [\ell] \text{ such that } L_{u,i_j} = +1\} \]

be the items in $M_2$ whose types are similar to one of the representatives $i_j$ that are liked by $u$. Note that if an item $i \in \tilde{R}_u$ is correctly classified by the algorithm (i.e. $E_i^c$ occurs, where $E_i$ is defined in Equation (8)), then $i \in R_u$. Hence $i \in \tilde{R}_u \setminus R_u$ implies $E_i$ and since $E_i \subseteq E_{i,L}$, we have

\[ |R_u| \geq |\tilde{R}_u| - \sum_{i \in M_2} 1[|E_i|] \geq |\tilde{R}_u| - \sum_{i \in M_2} 1[E_{i,L}] \].

(15)

It follows that

\[ \Pr[|R_u| \leq T] \leq \Pr[|\tilde{R}_u| - \sum_{i \in M_2} 1[E_{i,L}] \leq T] \leq \Pr[|\tilde{R}_u| \leq 3T/2] + \Pr[\sum_{i \in M_2} 1[E_{i,L}] \geq T/2] .

(16)

To bound the second term we use Claim 6.3 above, which gives

\[ \mathbb{E}\left[ \sum_{i \in M_2} 1[E_{i,L}] \right] \leq 2|M_2|\epsilon = \frac{64}{N}T, \]

(17)
and by Markov’s Inequality
\[ P \left[ \sum_{i \in M_2} \mathbb{I}[\bar{E}_{i,\bar{U}_i}] \geq \frac{T}{2} \right] \leq \frac{128}{\ell N}. \] (18)

We now bound the first term on the right-hand side of (16). To this end, let \( \widetilde{R}(j) \) be the set of items in \( M_2 \) whose types are the same as \( i_j \):
\[ \widetilde{R}(j) = \{ i \in M_2 : \tau_I(i) = \tau_I(i_j) \}. \] (19)
The sets \( \widetilde{R}(j) \) are disjoint because the representatives necessarily have different types. Let \( \mathcal{L}_u = \{ j \in [\ell] : L_{u,i_j} = +1 \} \) (20) be the set of item representatives that are liked by user \( u \). By definition of \( \widetilde{R}_u \) in (14),
\[ \widetilde{R}_u = \bigcup_{j \in \mathcal{L}_u(\ell)} \widetilde{R}(j), \]
and hence
\[ |\widetilde{R}_u| = \sum_{i \in M_2} \mathbb{I}[\tau_I(i) \in \mathcal{L}_u]. \]

Now, we claim that the set \( \mathcal{L}_u \) is independent of all items \( i \in M_2 \) and their types. To see this, note that \( \mathcal{L}_u \) is determined by row \( u \) of the type matrix \( \Xi \) and the items in \( M_1 \), their types, and some randomness in the algorithm, which determines the choice of \( i_j \) and \( \tau_I(i_j) \); these together determine \( L_{u,i_j} = \xi_{\tau_I(u),\tau_I(i_j)} \). So, conditioning on \( \mathcal{L}_u \) having cardinality \( \ell_u \), \( |\widetilde{R}_u| \) is the sum of \( |M_2| = 64\frac{3T}{\ell} \) i.i.d. Bernoulli variables with parameter \( \frac{\ell_u}{q_I} \) and hence \( |\widetilde{R}_u| \sim \text{Binom}(64\frac{3T}{\ell}, \frac{\ell_u}{q_I}) \). Conditioning on \( \ell_u \geq \ell/30 \), by a Chernoff bound (Lemma A.1) and stochastic domination of binomials with increasing number of trials we obtain
\[ P \left[ |\widetilde{R}_u| < 3T/2 \middle| \ell_u \geq \ell/30 \right] \leq \exp(-T/11). \]

In Lemma 6.5 below, we will lower bound the probability that \( \ell_u \geq \ell/30 \). Combining the last displayed inequality with Lemma 6.5, we get
\[ P[|\widetilde{R}_u| < 3T/2] \leq \exp(-T/11) + \exp(-\ell/5) + \exp(-q_I/18) + 20/N. \]

Plugging this and (18) into (16) gives
\[ P[|\mathcal{R}_u| < T] \leq \exp(-T/11) + \exp(-\ell/5) + \exp(-q_I/18) + 20/N + 128/\ell N, \]
and since \( \ell \leq T \) and \( \ell < q_I \),
\[ P[|\mathcal{R}_u| < T] \leq 3 \exp(-\ell/18) + 148/N. \]
The bound on \( \mathbb{E}[(T - |\mathcal{R}_u|)^+] \) follows immediately.

**Lemma 6.5.** Fix a \( u \in [N] \) and suppose that \( q_I > 8r \). With \( \mathcal{L}_u \) defined in (20) and \( \ell_u = |\mathcal{L}_u| \),
\[ P[\ell_u < \ell/30] \leq \exp(-\ell/5) + \exp(-q_I/18) + 20/N. \]
Proof. For a given user \( u \), \( \ell_u \) is the number of item type representatives liked by user \( u \). Let \( \tilde{L}_u \) be the item types in \([q]\) that are liked by user \( u \):

\[
\tilde{L}_u = \{ j \in [q] : \xi_{\tau_l(u),j} = +1 \}.
\] (21)

The variables \( \{\xi_{u,j}\}_{j \in [q]} \) are i.i.d. Bern(1/2), so a Chernoff bound (Lemma A.1) gives

\[
P[|\tilde{L}_u| < q_l/3] < \exp(-q_l/18).
\] (22)

We will show that

\[
P[\ell_u \leq \ell/30 \mid |\tilde{L}_u| \geq q_l/3] \leq \exp(-\ell/5) + 20/N,
\] (23)

which will prove the lemma by combining with (22).

We now work towards defining an error event in Equation (26) below. Let the sequence of random variables \( X_1 = \tau_l(i_1), X_2 = \tau_l(i_2), \ldots, X_\ell = \tau_l(i_\ell) \) denote the types of the item representatives chosen by the algorithm, so that \( \ell_u = |L_u| = \sum_{j \in [q]} 1[X_j \in \tilde{L}_u] \) (with \( L_u \) defined in (20)). Let \( \tilde{R}(j) \) be the set of items in \( M_1 \) of type \( j \)

\[
\tilde{R}(j) = \{ i \in M_1 : \tau_l(i) = j \}.
\] (24)

Later we will use the notation \( \tilde{R}(\cdot) = \{\tilde{R}(j)\}_{j=1}^{q} \) for the collection of these sets. Now let the event \( \tilde{E}_i \) denote misclassification of an item \( i \in M_1 \) (similar to event \( \mathcal{E}_i \) for \( i \in M_2 \) in (8)),

\[
\tilde{E}_i = \{ \exists j : i \in \tilde{S}_j, \tau_l(i) \neq \tau_l(i_j) \}.
\] (25)

Let \( \text{Err} \) be the event that for some item type \( j \), more than a fraction 1/10 of the items in \( M_1 \) of type \( j \) are misclassified:

\[
\text{Err} = \left\{ \sum_{i \in \tilde{R}(j)} 1[\tilde{E}_i] > \frac{|\tilde{R}(j)|}{10}, \text{for some } j \in [q] \right\}.
\] (26)

Conditioning on \( \text{Err} \) in the left-hand side of (23) gives

\[
P\left[ \sum_{i=1}^{\ell} 1[X_i \in \tilde{L}_u] < \ell/30 \mid |\tilde{L}_u| \geq q_l/3 \right]
\leq P\left[ \sum_{i=1}^{\ell} 1[X_i \in \tilde{L}_u] < \ell/30 \mid \text{Err}, |\tilde{L}_u| \geq q_l/3 \right] + P[\text{Err} \mid |\tilde{L}_u| \geq q_l/3].
\] (27)

**Bound on second term of** (27). We will use Markov inequality on \( P[\tilde{E}_i \mid \tilde{R}(\cdot), \tilde{L}_u] \) to bound

\[
P[\text{Err} \mid |\tilde{L}_u| \geq q_l/3] = E[P[\text{Err} \mid \tilde{R}(\cdot), \tilde{L}_u] \mid |\tilde{L}_u| \geq q_l/3].
\]

We need to consider the effect of matrix \( \Xi \) (and particularly its \( \tau_l(u) \)th row delineated by \( \tilde{L}_u \)) on the probability of error in categorizing items. Notably, if users rate two item types similarly, the probability of misclassifying these types is higher. As a result, \( \tilde{L}_u \) contains some information about discriminability of distinct item types.

Claim 6.2 shows that for \( i \in M_1 \), the probability of \( \mathcal{E}_i \) (the event that item \( i \) is misclassified) is \( P[\mathcal{E}_i] \leq 2\epsilon \). The same proof gives \( P[\tilde{E}_i \mid \tilde{L}_u] \leq 2\epsilon \) (with \( \tilde{E}_i \) defined in (25) for \( i \in M_2 \)). To show that, let
the potential error event \( \tilde{E}_{i,U_i} \), be the event that there exists \( j \neq \tau_I(i) \) such that for all the users in \( U_i \), we have \( L_{u,i} = X_{\tau_I(i),j} \). For \( i \in \tilde{S}_j \), let \( U_i = \text{rated}(i) \cap \text{rated}(i_j) \) at the time the algorithm added \( i \) to \( \tilde{S}_j \). Then the containment \( \tilde{E}_i \subseteq \tilde{E}_{i,U_i} \), holds. We use \( \mathbb{P}[	ilde{E}_i|R(\cdot),\tilde{L}_u] \geq \mathbb{P}[	ilde{E}_{i,U_i}|R(\cdot),\tilde{L}_u] = \mathbb{P}[	ilde{E}_{i,U_i}|\tilde{L}_u] \) since the event \( \tilde{E}_{i,U_i} \) is independent of \( R(\cdot) \).

As specified in Algorithm ItemExplore, \( |U_i| \geq r \); the set \( U_i \) of at least \( r \) users, chosen independently of feedback, rate item \( i \). By Lemma A.3, if \( q_U > 4r \), then there are users of at least \( r/2 \) distinct user types in \( U_i \) with probability at least \( 1 - \exp(-r/2) \geq 1 - \epsilon/q_U \).

To bound \( \mathbb{P}[	ilde{E}_{i,U_i}|\tilde{L}_u] \), we focus on the set of users \( U_i \setminus \{u\} \). Conditional on \( U_i \) having at least \( r/2 \) distinct user types, there are at least \( r/2 - 1 \) users of distinct types in \( U_i \setminus \{u\} \) whose preferences for item \( i \) are independent of \( \tilde{L}_u \). Any two item types \( j \neq \bar{j} \) have jointly independent user preferences by \( r/2 - 1 \) users with distinct types in \( U_i \setminus \{u\} \); they are rated in the same way by these users with probability at most \( 2^{-(r/2-1)} \). Hence, a union bound over item types \( j \neq \tau_I(i) \) gives \( \mathbb{P}[\tilde{E}_{i,U_i}|\tilde{L}_u] \leq q_U2^{-(r/2-1)} \). The choice \( r > 2\log(q_U/\epsilon) \) and \( \epsilon = \frac{1}{2qUN} \) gives \( \mathbb{P}[\tilde{E}_{i,U_i}|\tilde{L}_u] \leq \frac{1}{qU^N} \). So, \( \mathbb{P}[\tilde{E}_i|R(\cdot),\tilde{L}_u] \leq \frac{1}{qU^N} \) for all \( i \in M_1 \). Markov’s inequality gives

\[
\mathbb{P} \left[ \sum_{i \in R(j)} 1[\tilde{E}_i] > \frac{|\tilde{R}(j)|}{10} \left| R(\cdot), \tilde{L}_u \right] \right] \leq \frac{20}{qU^N}.
\]

Union bounding over \( j \in [q_U] \) and tower property gives the desired bound.

**Bound on first term of (27).** We start by showing that conditioned on the event \( \text{Err}^c \), variables \( R(\cdot), \tilde{L}_u \) such that \( |\tilde{L}_u| \geq q_U/3 \) and conditional on \( X_1, \cdots, X_{m-1} \), the type of the \( m \)th representative item, \( X_m \), is almost uniform over all the item types not learned yet, \( [q_U] \setminus \{X_n\}_{n=1}^{m-1} \). We will find upper and lower bounds for any \( j \in [q_U] \setminus \{X_n\}_{n=1}^{m-1} \) on

\[
\mathbb{P}[X_m = j] \{X_n\}_{n=1}^{m-1}, R(\cdot), \tilde{L}_u, \text{Err}^c \}.
\]

**Lower bound for (28).** For \( m \leq \ell \), let \( t \) be the time the \( m \)th item type representative is chosen by the algorithm. For \( j \in [q_U] \setminus \{X_n\}_{n=1}^{m-1} \) the probability of choosing representative of type \( X_m = j \) is equal to the proportion of items of type \( j \) in the remaining items in \( M_1 \) at time \( t - 1 \). The number of items of type \( j \) in \( M_1 \) at time \( t - 1 \) is at least \( |\tilde{R}(j)| - \sum_{i \in R(j)} 1[\tilde{E}_i] \).

The number of items removed from \( M_1 \) by time \( t \) is at least \( \sum_{t=1}^{m-1} |\tilde{R}(X_n)| \) (there could be some items with types not in \( \{X_n\}_{n=1}^{m-1} \) that were removed from \( M_1 \) due to misclassification). Hence, the total number of remaining items in \( M_1 \) by time \( t \) is at most \( M - \sum_{n=1}^{m-1} |\tilde{R}(X_n)| \). Let \( Z_{\tilde{E}} \) be the collection of indicator variables \( Z_{\tilde{E}} = \{1[\tilde{E}_i]\}_{i \in M_1} \). Thus, for any \( j \in [q_U] \setminus \{X_n\}_{n=1}^{m-1} \):

\[
\mathbb{P}[X_m = j | \{X_n\}_{n=1}^{m-1}, R(\cdot), Z_{\tilde{E}}, \tilde{L}_u] \geq \frac{|\tilde{R}(j)| - \sum_{i \in R(j)} 1[\tilde{E}_i]}{M - \sum_{n=1}^{m-1} |\tilde{R}(X_n)|}.
\]

On the event \( \text{Err}^c \) (defined in (26)), \( \sum_{i \in R(j)} 1[\tilde{E}_i] \leq |\tilde{R}(j)|/10 \) for any \( j \in [q_U] \setminus \{X_n\}_{n=1}^{m-1} \). Therefore, using tower property of expectation,

\[
\mathbb{P}[X_m = j | \{X_n\}_{n=1}^{m-1}, R(\cdot), \tilde{L}_u, \text{Err}^c] \geq \frac{|\tilde{R}(j)|}{M - \sum_{n=1}^{m-1} |\tilde{R}(X_n)|} \frac{9}{10}.
\]
some items from them. This gives

$$\Pr[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \mathcal{R}(\cdot), \tilde{E}, \tilde{L}_u] \leq \frac{\|\mathcal{R}(j)\|}{M - \sum_{n=1}^{m-1} |\mathcal{R}(X_n)| - \sum_{i \in M_1}^{m-1} \mathcal{R}(X_n) 1[i]}.$$  

By definition of ERR (in (26)), we have

$$\Pr[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \mathcal{R}(\cdot), \tilde{E}, \tilde{L}_u, \text{ERR}^c] \leq \frac{10}{9} \frac{|\mathcal{R}(j)|}{M - \sum_{n=1}^{m-1} |\mathcal{R}(X_n)|}. \quad (30)$$  

**Combining lower and upper bounds.** The expected value of $|\mathcal{R}(j)|$ conditional on variables $\{X_n\}_{n=1}^{m-1}$, $\{\mathcal{R}(X_n)\}_{n=1}^{m-1}$, $\tilde{L}_u$ is invariant to choice of $j \in [q_I] \setminus \{X_n\}_{n=1}^{m-1}$. So, the conditional expectation

$$C = \mathbb{E} \left[ \frac{|\mathcal{R}(j)|}{M - \sum_{n=1}^{m-1} |\mathcal{R}(X_n)|} \{ \{X_n\}_{n=1}^{m-1}, \{\mathcal{R}(X_n)\}_{n=1}^{m-1}, \tilde{L}_u, \text{ERR}^c \} \right],$$

is independent of $j$ for $j \in [q_I] \setminus \{X_n\}_{n=1}^{m-1}$. Using tower property of expectation on Equation (29) and (30) to remove the conditioning on $\{\mathcal{R}(j)\}_{j \notin \{X_n\}_{n=1}^{m-1}}$ along with the definition of $C$ above gives

$$\frac{9}{10} C < \Pr[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \{\mathcal{R}(X_n)\}_{n=1}^{m-1}, \tilde{L}_u, \text{ERR}^c] < \frac{10}{9} C.$$  

Since there are $q_I - (m - 1)$ types $j \in [q_I] \setminus \{X_n\}_{n=1}^{m-1}$, summing over $j$ in the second inequality of the last display gives $C \geq \frac{9}{10} q_I - \frac{1}{m - 1}$. Plugging this into the first inequality of the last display gives for all $j \in [q_I] \setminus \{X_n\}_{n=1}^{m-1}$,

$$\Pr[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \{\mathcal{R}(X_n)\}_{n=1}^{m-1}, \tilde{L}_u, \text{ERR}^c] \geq \frac{2}{5} \frac{q_I - 1}{q_I - m + 1}.$$  

Conditional on $\tilde{L}_u$ such that $|\tilde{L}_u| \geq q_I/3$, for $m \leq \ell \leq q_I$ and $s < \min\{m, \ell/30\}$ we have

$$\Pr[X_m \in \tilde{L}_u \mid \{X_n\}_{n=1}^{m-1}, \sum_{n' = 1}^{m-1} 1[X_{n'} \in \tilde{L}_u] < s, \tilde{L}_u, |\tilde{L}_u| \geq q_I/3, \text{ERR}^c] \geq \frac{2}{5} \frac{q_I - s}{q_I - m + 1} \geq \frac{1}{9}.$$  

So, the random variable $\sum_{n' = 1}^{\ell} 1[X_{n'} \in \tilde{L}_u]$ conditional on $\tilde{L}_u$ and events $|\tilde{L}_u| \geq q_I/3$ and ERR$^c$, stochastically dominates a Binomial random variable with mean $\ell/9$. Hence, by a Chernoff bound,

$$\Pr \left[ \sum_{n' = 1}^{\ell} 1[X_{n'} \in \tilde{L}_u] < \frac{\ell}{30} \mid \tilde{L}_u, |\tilde{L}_u| \geq \frac{q_I}{3}, \text{ERR}^c \right] \leq \exp(-\ell/5).$$

### 7 Item structure only: lower bound

In this section we prove the following lower bound.

**Theorem 7.1.** Consider the item-structure model. Let $r = \lfloor \log q_I - 6 \log \log N \rfloor$ and $\eta = 1/\log N$. Any recommendation algorithm must incur regret

$$\text{regret}(T) \geq \frac{1 - 4\eta}{2} \max \left\{ \min \left\{ \sqrt{\frac{T q_I}{16 N}}, \frac{T}{16 \log q_I} \right\}, \frac{r}{2N} T, 1 \right\} - \frac{1}{N} T.$$  

The simplified version of this theorem in Section 3 is derived by using $N > 16q_I \log q_I$ which gives $\eta < 1/2$ and $\sqrt{\frac{T q_I}{16 N}} < \frac{T}{16 \log q_I}$. 

25
7.1 Proof strategy

We call a recommendation $a_{u,t}$ to user $u$ at time $t$ a good recommendation when the probability of $L_{u,a_{u,t}} = +1$ given the history is large (close to 1). Conversely, recommendations for which the probability of $L_{u,a_{u,t}} = +1$ is close to (or smaller than) $1/2$ are considered bad (or uncertain) recommendations. Good and bad refers only to the confidence that the recommendation is liked: a good recommendation is not always liked and a bad recommendation is not always disliked.

We identify two scenarios in which recommendations are necessarily bad: (i) Similar to the lower bound for the model with user structure only in Section 5, we use the concept of $(r, \eta)$-row regularity (Definition 7.1). Lemma 7.3 shows that for a $(r, \eta)$-row regular preference matrix, items with fewer than $r$ ratings have probability roughly half of being liked by any user, even if the preference matrix is known. (ii) Lemma 7.4 shows that if a user $u$ has not rated any item with the same type as item $i$ before, the probability that user $u$ likes item $i$ is $1/2$.

Next, in Lemma 7.6 we upper bound the number of good recommendations, which entails a counting argument.

The statement of the theorem follows immediately from Lemmas 7.5 and 7.6 below.

7.2 Proof of Theorem 7.1

Definition 7.1 (Row-regularity). The matrix $A \in \{-1, +1\}^{n \times m}$ is said to be $(r, \eta)$-row regular if its transpose is $(r, \eta)$-column regular. We write $A^T \in \Omega_{r,\eta}$, where $\Omega_{r,\eta}$ is the set of $(r, \eta)$-column regular matrices.

The following lemma is an immediate corollary of Lemma 5.3.

Lemma 7.2. Let matrix $A \in \{-1, +1\}^{n \times m}$ have i.i.d. Bern(1/2) entries. If $\eta < 1$, then the matrix $A$ is $(r, \eta)$-row regular with probability at least

$$1 - 2(2n)^r \exp \left( -\frac{\eta^2 m}{3 2^r} \right).$$

Throughout this section we will fix

$$r = \lfloor \log q_I - 6 \log \log N \rfloor \quad \text{and} \quad \eta = \frac{1}{\log N}. \quad (31)$$

Applying Lemma 7.2 with $r$ and $\eta$ defined above we obtain that the preference matrix $\Xi$ is $(r, \eta)$ row-regular (i.e., $\Xi^T \in \Omega_{r,\eta}$) with probability

$$\Pr[\Xi^T \in \Omega_{r,\eta}] \geq 1 - \frac{1}{N}, \quad (32)$$

so long as $N > 20$.

Let $c_{i}^t$ be the number of times item $i$ has been rated by any user up until time $t - 1$,

$$c_{i}^t := \sum_{u=1}^{N} \sum_{s=1}^{t-1} \mathbb{1}[a_{u,s} = i].$$

The following lemma shows that if the number $c_{i}^t$ of times an item is recommended is small and the preference matrix is row-regular, then the outcome of recommending the item is uncertain.
Lemma 7.3. Let $r$ and $\eta$ be as in Equation (31). For any user $u \in [N]$ and item $i \in \mathbb{N}$,

$$
P[L_{u,i} = +1 \mid a_{u,t} = i, c^t_i < r, \Xi^\top \in \Omega_{r,\eta}] \leq (1 + 4\eta)/2.
$$

The next lemma shows that if a user $u$ has not rated any item with the same type as item $i$ before, the probability that $u$ likes item $i$ is $1/2$.

Let $B^t_{u,\tau_f(i)}$ denote the event that user $u$ has rated an item of the same type as item $i$ by time $t - 1$, i.e.,

$$
B^t_{u,\tau_f(i)} = \{\exists i' \in \mathbb{N} : a_{u,s} = i' \text{ for some } s < t \text{ and } \tau_f(i) = \tau_f(i')\}.
$$

Lemma 7.4. We have that $P[L_{u,i} = -1 \mid a_{u,t} = i, (B^t_{u,\tau_f(i)})^c, c^t_i \geq r] = 1/2$.

Lemmas 7.3 and 7.4 identify scenarios in which recommendations are bad. In the complementary scenario recommendations are not necessarily bad (and may be good), and we denote their number by

$$
good(T) = \sum_{t \in [T]} \sum_{u \in [N]} 1[c^t_{a_{u,t}} \geq r, B^t_{u,\tau_f(a_{u,t})}].
$$

The following lemma uses Lemmas 7.3 and 7.4 to provide a lower bound for regret in terms the number $\text{good}_T$.

Lemma 7.5. For $\eta$ and $r$ defined in Equation (31),

$$
N \text{regret}(T) \geq \frac{1 - 4\eta}{2} (T \cdot N - \mathbb{E} [\text{good}(T)]) - T.
$$

Proof. We partition the liked recommendations based on $c^t_{a_{u,t}}, B^t_{u,\tau_f(a_{u,t})}$, and row regularity of $\Xi$:

$$
N(T - \text{regret}(T)) = \sum_{t \in [T]} \sum_{u \in [N]} P[L_{u,a_{u,t}} = +1]
$$

$$
= \sum_{t \in [T]} \sum_{u \in [N]} P[L_{u,a_{u,t}} = +1, c^t_{a_{u,t}} < r, \Xi^\top \notin \Omega_{r,\eta}]
$$

$$
+ \sum_{t \in [T]} \sum_{u \in [N]} P[L_{u,a_{u,t}} = +1, c^t_{a_{u,t}} < r, \Xi^\top \in \Omega_{r,\eta}]
$$

$$
+ \sum_{t \in [T]} \sum_{u \in [N]} P[L_{u,a_{u,t}} = +1, c^t_{a_{u,t}} \geq r, (B^t_{u,\tau_f(a_{u,t})})^c]
$$

$$
+ \sum_{t \in [T]} \sum_{u \in [N]} P[L_{u,a_{u,t}} = +1, c^t_{a_{u,t}} \geq r, B^t_{u,\tau_f(a_{u,t})}]
$$

$$
=: A1 + A2 + A3 + A4.
$$

The proof is obtained by plugging in the four bounds below.

Bounding $A1$. Plugging in the probability of $\Xi$ being row regular from (32) gives

$$
A2 = \sum_{t \in [T]} \sum_{u \in [N]} P[L_{u,a_{u,t}} = +1, c^t_{a_{u,t}} < r, \Xi^\top \notin \Omega_{r,\eta}]
$$

$$
\leq NT \cdot P[\Xi^\top \notin \Omega_{r,\eta}] \leq T.
$$
Bounding A2. Multiplying the statement of Lemma 7.3 by \( \mathbb{P}[a_{u,t} = i, c_i^t < r, \Xi^T \in \Omega_{r,\eta}] \) and summing over \( i \) gives

\[
A1 = \sum_{i \in [T]} \sum_{t \in [T]} \mathbb{P}[L_{u,i} = +1, a_{u,t} = i, c_i^t < r, \Xi^T \in \Omega_{r,\eta}]
\leq \frac{1 + 4\eta}{2} \sum_{i \in [T]} \mathbb{P}[c_i^t < r, \Xi^T \in \Omega_{r,\eta}] \leq \frac{1 + 4\eta}{2} \sum_{i \in [T]} \mathbb{P}[c_i^t < r].
\]

Bounding A3. Lemma 7.4 gives

\[
\mathbb{P}[L_{u,i} = +1, a_{u,t} = i, c_i^t \geq r, (B^t_{u,t}(i))^c] = \frac{1}{2}.
\]

Multiplying this by \( \mathbb{P}[a_{u,t} = i, (B^t_{u,t}(i))^c, c_i^t \geq r] \) and taking the summation over \( i \) gives

\[
A4 = \frac{1}{2} \sum_{u \in [N], t \in [T]} \mathbb{P}[c_{u,t} \geq r, (B^t_{u,t}(i))^c].
\]

Bounding A4. We bound by one the probability that a good recommendation is liked to obtain

\[
A3 = \sum_{u \in [N], t \in [T]} \mathbb{P}[L_{u,a_{u,t}} = +1, c_{a_{u,t}}^t \geq r, B^t_{u,t}(a_{u,t})]
\leq \sum_{u \in [N], t \in [T]} \mathbb{P}[c_{a_{u,t}}^t \geq r, B^t_{u,t}(a_{u,t})] = \mathbb{E}[good(T)]. \tag*{\[\diamondsuit\]}
\]

Next, we find an upper bound on the expected number of “good” recommendations made by the algorithm in terms of parameters of the model.

Lemma 7.6. For any algorithm,

\[
\mathbb{E}[good(T)] \leq T N - \max \left\{ \frac{1}{2} \min \left\{ \sqrt{\frac{NT q_l}{4}}, \frac{NT}{8\log q_l} \right\}, Tr, N \right\}.
\]

As noted above, Theorem 7.1 is an immediate consequence of Lemmas 7.5 and 7.6.

7.3 Proofs of lemmas

Proof of Lemma 7.3. We show that if an item has been rated by fewer than \( r \) users, its type is uncertain, because many item types are consistent with the history. For a row regular preference matrix, uncertainty in the item type makes it impossible to accurately predict user preferences for the item.

We focus on a particular item \( i \) at time \( t \). Let \( \{u \in [N] : a_{u,s} = i, s < t\} \) be the set of users that were recommended item \( i \) up to time \( t - 1 \), and let \( b = \{L_{u,i}\}_{u \in w} \) be the vector of feedback from users in \( w \) about item \( i \). Note that \( c_i^t < r \) implies \( |w| < r \). We re-introduce the notation from Definition 5.1: if \( M \) is the matrix obtained by concatenating the rows of \( \Xi \) indexed by \( w \), then \( K_{h,w}(\Xi) \) is the set of columns of \( M \) (corresponding to the item types) equal to \( b \). This is the set of item types consistent with the ratings \( b \) of users \( w \) for item \( i \).
Conditional on Ξ, w, and b, the type τ_I(i) of item i at the end of time instant t − 1 is uniformly distributed over the set of item types K_{b,w}(Ξ^T). This allows us to relate the posterior probability of i being liked to row regularity of Ξ as follows. Let \( b^+ = [b^+ 1] = \{-1, +1\} \) be obtained from b by appending +1. For a given user u \( \notin w \), we have \( L_{u,i} = +1 \) precisely when \( \tau_I(i) \in K_{b^+,\{w,u\}}(\Xi^T) \), which in words reads “item i is among those types that are consistent with the ratings of i up to time t − 1 and have preference vector with ‘+1’ for user u”. It follows that for any preference matrix Ξ and any user u which has not rated i up to time t − 1,\[
\mathbb{P}[L_{u,i} = +1 | \mathcal{H}_{t-1}, \Xi] = \mathbb{P}[\tau_I(i) \in K_{b^+,\{w,u\}}(\Xi^T) | \mathcal{H}_{t-1}, \Xi] = \frac{k_{b^+,\{w,u\}}(\Xi^T)}{k_{b,w}(\Xi^T)}.
\] (35)
The second equality is due to: (i) w and b are functions of \( \mathcal{H}_{t-1} \) (ii) for fixed w and b, the set \( K_{b,w}(\Xi^T) \) is determined by \( \Xi^T \); (iii) \( \tau_I(i) \) is uniformly distributed on \( K_{b,w}(\Xi^T) \).

Recall that \( \Xi^T \in \Omega_{\tau,\eta} \) if the preference matrix Ξ is \((r, \eta)\)-row regular. We have the inequalities
\[
\mathbb{P}[L_{u,i} = +1, a_{u,t} = i | c_t^i \leq r, \Xi^T \in \Omega_{r,\eta}] = \mathbb{E} \left[ \mathbb{P}[L_{u,i} = +1 | \mathcal{H}_{t-1}, \Xi] | a_{u,t} = i, \Xi^T \in \Omega_{r,\eta} \right]
\]
\[
= \mathbb{E} \left[ \frac{k_{b^+,\{w,u\}}(\Xi^T)}{k_{b,w}(\Xi^T)} \mathbb{1}[u \notin w] \mathbb{P}[\tau_I(i) \in K_{b^+,\{w,u\}}(\Xi^T) | a_{u,t} = i, \Xi^T \in \Omega_{r,\eta}] \right]
\]
\[
\leq \frac{(1 + \eta)}{(1 - \eta)} \frac{q^{|w||t|+1}}{q^{|w|}} \mathbb{P}[a_{u,t} = i | c_t^i \leq r, \Xi^T \in \Omega_{r,\eta}] \leq \frac{1}{2}(1 + 4\eta) \mathbb{P}[a_{u,t} = i | c_t^i \leq r, \Xi^T \in \Omega_{r,\eta}],
\]
and this proves the lemma. It remain to justify the steps above. (a) follows since conditional on \( \mathcal{H}_{t-1} \), the random variable \( a_{u,t} \) is independent of all other random variables. (b) uses (35) and the fact that \( \mathbb{P}[a_{u,t} = i | \mathcal{H}_{t-1}] \) is nonzero only if u has not rated item i, so we may add \( \mathbb{1}[u \notin w] \). (c) is justified as follows: if \( \Xi^T \in \Omega_{r,\eta} \), then by Claim 5.2, \( \Xi^T \in \Omega_{r-1,\eta} \). By Definition 5.1, this means that \( k_{b,w}(\Xi^T) \geq (1 - \eta)q^{|w|}/2^{|w|+1} \) and \( k_{b^+,\{w,u\}}(\Xi^T) \leq (1 + \eta)q^{|w|}/2^{|w|+1} \). (d) If \( N > 8 \) then η given in (31) is less than 1/2 and \((1 + \eta)/(1 - \eta) \leq 1 + 4\eta \). □

**Proof of Lemma 7.4.** We make two observations: (i) if user u has not rated any item with item type \( \tau_I(i) \) before, the feedback in history \( \mathcal{H}_{t-1} \) is independent of the value of \( \xi_{u,\tau_I(i)} \) given all other elements of matrix \( \Xi \) and the item types; (ii) According to the model, the types of items (function \( \tau_I(\cdot) \)) are independent of the matrix \( \Xi \), and the elements \( \Xi \) are i.i.d. Hence, conditional on \( (\mathcal{B}^t_{u,\tau_I(i)})^c \), the posterior distribution at time t of \( L_{u,i} \) is uniform on \{-1, +1\}.

First, define \( \Xi_{\{\{u,j\}\}} \) to be the entries of \( \Xi \) except \( \xi_{u,j} \), i.e.,
\[
\Xi_{\{\{u,j\}\}} = \{\xi_{u',j'} : (u', j') \neq (u, j)\},
\]
which is independent of \( \xi_{u,j} \). Also define \( \tau_I(\cdot) = \{\tau_I(i)\}_{i \in \mathbb{N}} \) to be the sequence of item types. Note that this set of random variables is also independent of the elements of \( \Xi \) and specifically \( \xi_{u,j} \). Also, recall that conditional on \( \mathcal{H}_{t-1} \), \( a_{u,t} \) is independent of \( \Xi \) and item types \( \tau_I(\cdot) \).

Next, we show that conditional on \( \Xi_{\{\{u,j\}\}} \) and \( \tau_I(\cdot) \), the event \( \{a_{u,t} = i, \tau_I(i) = j, (\mathcal{B}^t_{u,\tau_I(i)})^c, c_t^i \geq r, \mathcal{H}_{t-1} = \mathcal{H}\} \) is independent of \( \xi_{u,j} \):
\[
\xi_{u,j} | \{a_{u,t} = i, \tau_I(i) = j, (\mathcal{B}^t_{u,\tau_I(i)})^c, c_t^i \geq r, \mathcal{H}_{t-1} = \mathcal{H}\} \{\Xi_{\{\{u,j\}\}}, \tau_I(\cdot)\}.
\] (36)
The above follows from: (i) $a_{u,t}$ is conditionally independent of $\Xi$ and $\tau_I(\cdot)$ given $\mathcal{H}_{t-1}$; (ii) The event \{\(\tau_I(i) = j\)\} is contained in the specification of $\tau_I(\cdot)$; (iii) $B^t_{u,\tau_I(i)}$ defined in (33) is a deterministic function of $\mathcal{H}_{t-1}$ and $\tau_I(i)$; (iv) $c^t_i$ is a deterministic function of $\mathcal{H}_{t-1}$; (v) The history $\mathcal{H}_{t-1}$ is defined to be the sequence of random variables

\[
\{(a_{u',s}, L_{u',a_{u',s}}), \text{ for all } u' \in [N] \text{ and } s < t\}.
\]

We show inductively that if \((B^t_{u,\tau_I(i)})^c\) and $\tau_I(i) = j$ holds, conditional on \(\{\Xi \setminus \{u,j\}, \tau_I(\cdot)\}\), for $s < t$, \(\mathcal{H}_s\) is independent of $\xi_{u,j}$. Let this be true for $s - 1$. Recall that given $\mathcal{H}_{s-1}$, for all $u' \in [N]$, the recommended items $a_{u',s}$ are independent of $\Xi$ and $\tau_I(\cdot)$ and in particular independent of $\xi_{u,j}$. We also know that $\tau_I(a_{u,s}) \neq j$ as a result of $(B^t_{u,\tau_I(i)})^c$ and $\tau_I(i) = j$. Thus, $L_{u',a_{u',s}}$ for all $u' \in [N]$ is a function of $\Xi \setminus \{u,j\}$ and $\tau_I(\cdot)$ and is independent of $\xi_{u,j}$.

Since i) $\Xi \setminus \{u,j\}$ and $\tau_I(\cdot)$, are independent of $\xi_{u,j}$ and ii) Conditional independence property in Equation (36) holds for every $\mathcal{H}$ and $j \in [q_t]$:

\[
\mathbb{P}[L_{u,i} = +1 | a_{u,t} = i, \tau_I(i) = j, (B^t_{u,\tau_I(i)})^c, c^t_i \geq r, \mathcal{H}_{t-1} = \mathcal{H}, \Xi \setminus \{u,j\}, \tau_I(\cdot)]
\]

\[
= \mathbb{P}[\xi_{u,j} = +1 | a_{u,t} = i, \tau_I(i) = j, (B^t_{u,\tau_I(i)})^c, c^t_i \geq r, \mathcal{H}_{t-1} = \mathcal{H}, \Xi \setminus \{u,j\}, \tau_I(\cdot)]
\]

\[
= \mathbb{P}[\xi_{u,j} = +1 | \Xi \setminus \{u,j\}, \tau_I(\cdot)] = \frac{1}{2}.
\]

The lemma now follows by the tower property.

To prepare for proving Lemma 7.6, we introduce the following quantities. Let $\mathcal{F}_t$ be the set of items that have been rated by at least $r$ users by time $T$ and $f_t := |\mathcal{F}_t|$ their number,

\[
f_t = \sum_{i \in [N]} 1[c^t_i \geq r].
\]

Let $\mathcal{G}_t$ be the set of items that have been rated by at least one and fewer than $r$ users by time $T$ and $g_t := |\mathcal{G}_t|$ their number,

\[
g_t = \sum_{i \in [N]} 1[0 < c^t_i < r].
\]

In the following claim, we bound the number of good recommendations in terms of $f_T$ and $g_T$.

**Claim 7.7.** Let $\text{good}(T)$ be defined as Equation (34). Then $\text{good}(T) \leq TN - g_T - f_Tr$.

**Proof.** Any $i \in \mathcal{F}_T$ is recommended to at least $r$ users by the end of time $T$. So, for any $i \in \mathcal{F}_T$, there are at least $r$ recommendations in which $i$ has been rated fewer than $r$ previous times:

\[
\sum_{\substack{t \in [T] \atop u \in [N]}} 1[c^t_i < r, a_{u,t} = i] \geq r.
\]

Any $i \in \mathcal{G}_T$ has been recommended at least once, so for these items

\[
\sum_{\substack{t \in [T] \atop u \in [N]}} 1[c^t_i < r, a_{u,t} = i] \geq 1.
\]
All recommended items are either in $F_T$ or $G_T$. Hence, the total number of recommendations satisfies

$$NT = \sum_{t \in [T]} \sum_{u \in [N]} 1[a_{u,t} = i] = \sum_{t \in [T]} \sum_{u \in [N]} \sum_{i \geq 1} 1[c^T_i \geq r, a_{u,t} = i] + \sum_{t=1}^{T} \sum_{u=1}^{N} \sum_{i \geq 1} 1[c^T_i < r, a_{u,t} = i]$$

$$\geq \text{good}(T) + \sum_{t \in [T]} \sum_{u \in [N]} \sum_{i \geq 1} 1[c^T_i < r, a_{u,t} = i] + \sum_{t \in [T]} \sum_{u \in [N]} \sum_{i \geq 1} 1[c^T_i < r, a_{u,t} = i]$$

$$\geq \text{good}(T) + g_T + f_T r.$$

where we used $\text{good}(T) \leq \sum_{t \in [T]} \sum_{u \in [N]} \sum_{i \geq 1} 1[c^T_i \geq r, a_{u,t} = i].$ 

The rest of the section contains the proof of Lemma 7.6.

**Proof of Lemma 7.6.** The lemma consists of three bounds on $\text{good}(T)$. The bound $\text{good}(T) \leq TN - Tr$ is the easiest so we start with this. Because each item is recommended at most once to each user, each item (and specifically items in $F_T$) are recommended at most $N$ times, while items in $G_T$ are recommended at most $r$ times. The total number of recommendations made by time $T$ is therefore upper bounded as

$$TN \leq f_T N + g_T r,$$

and rearranging yields

$$g_T \geq (T - f_T) \frac{N}{r}.$$  \hspace{1cm} (37)

Plugging this into the statement of Claim 7.7 gives

$$\text{good}(T) \leq TN - g_T (1 - \frac{r^2}{N}) - Tr \leq TN - Tr,$$  \hspace{1cm} (38)

where the second inequality is due to the assumptions $q_i < N$ and $N > 20$.

We now prove the other two bounds on $\text{good}(T)$, this time in terms of the number of item types each user has rated by time $T$.

Let $\Gamma^T_u$ be the set of item types that are recommended to user $u$ up to time $T$,

$$\Gamma^T_u = \left\{ j \in [q_f] : \sum_{t \in [T]} 1[i = j] : i = t_{ij} \right\},$$  \hspace{1cm} (39)

and let $\gamma^T_u = |\Gamma^T_u|$ and $\gamma^T_* := \min_u \gamma^T_u$.

**Claim 7.8.** $\text{good}(T) \leq N(T - \gamma^T_*)$.

**Proof.** For user $u$, the number of times an item type is rated for the first time is equal to the number of item types rated by user $u$. Hence, $\sum_{t \in [T]} 1[B^T_{u,t_{ij}(a_{u,t})}] = \gamma^T_u$ and $\sum_{t \in [T]} 1[B^T_{u,t_{ij}(a_{u,t})}] = T - \gamma^T_u$. Summing over $u$ and using $\text{good}(T) \leq \sum_{u \in [N]} 1[B^T_{u,t_{ij}(a_{u,t})}]$ proves the claim. \hfill \square

**Proof of inequality** $E[\text{good}(T)] \leq NT - N$. This follows from Claim 7.8 since $\gamma^T_* \geq 1$. 

31
Proof of inequality $\mathbb{E}[\text{good}(T)] \leq NT - \frac{1}{2} \min \left\{ \sqrt{\frac{NTq_t}{4}}, \frac{NT}{8\log q_t} \right\}$. Let $r_j^t$ be the number of items with type $j$ that have been recommended (to any user) by time $t$:

$$r_j^t = |\{i : \tau_t(i) = j, a_{u,s} = i \text{ for some } u \in [N] \text{ and } s \in [t]\}|.$$

$T$ items with types in $\Gamma^T_u$ (defined in (39)) are recommended to user $u$ by time $T$. Hence, $T \leq \sum_{j \in \Gamma^T_u} r_j^T \leq \gamma_u T \max_j r_j^T$. This implies that

$$\gamma_u^T \geq \frac{T}{\max_j r_j^T}. \quad (40)$$

We get a lower bound on $\gamma_u^T$ via an upper bound on $\max_j r_j^T$, which is in turn obtained via martingale concentration for each $r_j^T$. This is essentially just a question of bounding the fullest bin in a balls and bins scenario, with the added complication that the time-steps in which balls are thrown is a martingale. Thus the number of balls (i.e., $f_T + g_T$) is random, and the decision to throw a ball at a given time may depend on the configuration of balls in bins.

Let $r^t = (r_1^t, \ldots, r_q^t)$. One may check that the sequence $r_j^t - (f_t + g_t)/q_t$ is a martingale with respect to filtration $\mathcal{F}_t = \sigma(r^0, r^1, \ldots, r^t)$. It turns out to be easier to work with a different martingale that considers each user separately, so that the item counts are incremented by at most one at each step. Consider the lexicographical ordering on pairs $(t, u)$, where $(s, v) \leq (t, u)$ if either $s < t$ or $s = t$ and $v \leq u$ (the recommendation to user $v$ at time $s$ occurred before that of user $u$ at time $t$). Let $r^{t,u} = (r_1^{t,u}, \ldots, r_q^{t,u})$, where for $j \in [q]$

$$r_j^{t,u} = |\{i : \tau_t(i) = j, a_{v,s} = i \text{ for some } (s, v) \leq (t, u)\}|,$$

and $\rho^{t,u} = \sum_j r_j^{t,u}$. We now define a sequence of stopping times $Z_k \in \mathbb{N} \times [N]$, with $Z_0 = (0, N)$ and

$$Z_k = \min \\{ (t, u) > Z_{k-1} : \rho^{t,u} > \rho^{t,u-1} \},$$

where $(t, 0)$ is interpreted as $(t-1, N)$. The $Z_k$ is the first $(t, u)$ such that a new item is recommended by the algorithm for the $k$th time, therefore $\rho^{Z_k} = k$. The $Z_k$ are stopping times with respect to $(\rho^{t,u})$, and observe that for $k^* = \max\{k : Z_k \leq (T, N)\} = \rho^{T,N} = f_T + g_T$ since $f_T + g_T$ is the total number of items recommended by the algorithm by the end of time $T$. Also, we have $\rho^T = f_T + g_T$ and $r_j^{T,N} = r_j^{(T,N)}$ for all $j \in [q]$. The sequence $M_{t,u} = r_j^{t,u} - \rho^{t,u}/q_t$ is a martingale w.r.t. $\mathcal{F}_{t,u} = \sigma(\rho^{t+1}, \ldots, \rho^{t,u})$, and it follows that so is $\widetilde{M}_k = M_{(T,N) \Delta Z_k}$, this time with respect to $\mathcal{F}_{k} := \mathcal{F}_{(T,N) \Delta Z_k}$. Since $Z_{k^*} \leq (T, N)$ (by using the definition of $k^*$), we have $\widetilde{M}_{k^*} = M_{k^*}$. Also, note that $r_j^{T,N} = r_j^{T,N}$ and $\rho^{k^*} = \rho^{T,N} = f_T + g_T$.

We’d like to apply martingale concentration (Lemma A.2) to $\widetilde{M}_k$, and to this end observe that $\text{Var}(\widetilde{M}_{k} | \mathcal{F}_{k}) \leq 1/q_t$ and $|\widetilde{M}_k - \widetilde{M}_{k-1}| \leq 1$ almost surely.

For any $k \geq k_0 := 2q_t \log q_t$, Lemma A.2 gives

$$\mathbb{P}\left[ \widetilde{M}_k \geq \frac{3k}{q_t} \right] \leq \exp\left( \frac{-9k^2/q_t^2}{2k/q_t + k/q_t} \right) = \exp\left( -\frac{3k}{q_t} \right).$$

By a union bound,

$$\mathbb{P}\left[ \exists k \geq k_0 \text{ s.t. } \widetilde{M}_k \geq \frac{3k}{q_t} \right] \leq \sum_{k \geq k_0} \exp\left( -\frac{3k}{q_t} \right) = \frac{\exp\left( -\frac{2k_0}{q_t} \right)}{1 - \exp\left( -\frac{3}{q_t} \right)} \leq \frac{1}{q_t}. \quad (41)$$

32
where the last inequality uses $1 - \exp(-3/q_r) > q_r^{-2}$ (which is derived from $e^{-a} \leq 1 - a + a^2/2$ and $q_r > 1$).

For $k < k_0$ we get

$$P[\tilde{M}_k \geq 6 \log q_r] \leq \exp\left(\frac{-36 \log^2 q_r}{2k_0 + 2 \log q_r}\right) \leq \frac{1}{q_r^2}.$$  

So

$$P[\exists k < k_0 \text{ s.t. } \tilde{M}_k \geq 6 \log q_r] \leq \sum_{k < k_0} \frac{1}{q_r^6} = \frac{k_0}{q_r^6} = \frac{1}{q_r^2}. \quad (42)$$

For any $j \in [q_r]$,

$$P[r_j^T \geq \max\{8 \log q_r, \frac{4(f_T + g_T)}{q_r}\}]$$

$$\leq P[\tilde{M}_{k^*} \geq \max\{8 \log q_r, \frac{4k^*}{q_r}\} - \frac{k^*}{q_r}]$$

$$= P[\tilde{M}_{k^*} \geq \max\{8 \log q_r - \frac{k^*}{q_r}, \frac{3k^*}{q_r}\}, k^* \geq k_0] + P[\tilde{M}_{k^*} \geq \max\{8 \log q_r - \frac{k^*}{q_r}, \frac{3k^*}{q_r}\}, k^* < k_0]$$

$$\leq P[\tilde{M}_{k^*} \geq \frac{3k^*}{q_r}, k^* \geq k_0] + P[\tilde{M}_{k^*} \geq 8 \log q_r - \frac{k^*}{q_r}, k^* < k_0]$$

$$\leq P[\exists k \geq k_0 \text{ s.t. } \tilde{M}_k \geq \frac{3k}{q_r}] + P[\exists k < k_0 \text{ s.t. } \tilde{M}_k \geq 6 \log q_r, k^* < k_0]$$

$$\leq P[\exists k \geq k_0 \text{ s.t. } \tilde{M}_k \geq \frac{3k}{q_r}] + P[\exists k < k_0 \text{ s.t. } \tilde{M}_k \geq 6 \log q_r] \leq \frac{2}{q_r^2}.$$  

Here (a) follows by unpacking the definitions of $\tilde{M}_k$ and $k^*$, (b) and (c) use the definition $k_0 = 2q_r \log q_r$; (d) follows by (41) and (42).

Equation (40) and a union bound for $j \in [q_r]$ for $q_r > 1$ now imply that

$$P[\gamma_j^T > \min \left\{ \frac{T}{8 \log q_r}, \frac{T q_r}{4(f_T + g_T)} \right\}] \leq P\left[ \max_{j \in [q_r]} r_j^T \geq \max \left\{ 8 \log q_r, \frac{4(f_T + g_T)}{q_r} \right\} \right] \leq \frac{2}{q_r^2} \leq \frac{1}{2}.$$  

Plugging this into Claim 7.8, with probability at least $1/2$ we have that

$$\text{good}(T) \leq NT - N \min \left\{ \frac{T}{8 \log q_r}, \frac{T q_r}{4(f_T + g_T)} \right\}$$

and combining with the statement of Claim 7.7 gives that with probability at least $1/2$,

$$\text{good}(T) \leq NT - \max \left\{ f_T r + g_T, N \min \left\{ \frac{T}{8 \log q_r}, \frac{T q_r}{4(f_T + g_T)} \right\} \right\}$$

$$\leq NT - \max \left\{ f_T + g_T, N \min \left\{ \frac{T}{8 \log q_r}, \frac{T q_r}{4(f_T + g_T)} \right\} \right\}$$

$$\leq NT - \min \left\{ \max \left\{ f_T + g_T, \frac{T N q_r}{4(f_T + g_T)} \right\}, \frac{NT}{8 \log q_r} \right\}$$

$$\leq NT - \min \left\{ \frac{1}{2} \sqrt{NT q_r}, \frac{NT}{8 \log q_r} \right\}.$$
(a) and (b) holds for any $f_T, g_T$. Also, since $\gamma^T \geq 1$, from Claim 7.7 we have $\text{good}(T) \leq NT - N$ with probability 1. Hence,
\[
\mathbb{E}[\text{good}(T)] \leq NT - \frac{1}{2} \min\left\{\frac{1}{2} \sqrt{N T q_I}, \frac{N T}{8 \log q_I}\right\}.
\]
The above two statements and Equation (38) give the result in the Lemma.

8 Discussion

In this paper, we analyzed the performance of online collaborative filtering based on a latent variable model of the preference of users for items. We proposed a variation of user-user and a variation of item-item algorithm tailored to the latent variable model. We also provided lower bounds for regret in the extreme regimes of parameters corresponding to user-structure only (with no structure in item space) and item-structure only (with no structure in user space). The lower bounds showed that the proposed algorithms are almost information theoretically optimal in these extreme parameter regimes.

Several extensions of the problem of are interest: a hybrid algorithm exploiting structure in both user space and item space appears in a subsequent paper. Modification of algorithms to guarantee robustness to noise in observations with independent identically distributed noise is rather easily derived from our proposed algorithms.

Studying the model of preference matrix where the preference of various user types for item types are not statistically independent is also of importance from the practical point of view.

Acknowledgment

We thank Devavrat Shah for many helpful discussions and Alexander Rakhlin for suggesting several references. This work was supported in part by grants NSF CCF-1565516, ONR N00014-17-1-2147, and DARPA W911NF-16-1-0551.

A Concentration Lemmas

The following lemma is derived by application of Chernoff bound to Binomial variables [10].

Lemma A.1 (Chernoff bound). Let $X_1, \ldots, X_n \in [0, 1]$ be independent random variables. Let $X = \sum_{i=1}^{n} X_i$ and $\bar{X} = \sum_{i=1}^{n} \mathbb{E}X_i$. Then, for any $\epsilon > 0$,
\[
\mathbb{P}[X \geq (1 + \epsilon) \bar{X}] \leq \exp\left(-\frac{\epsilon^2}{2 + \epsilon} \bar{X}\right) \leq \max\left\{\exp\left(-\frac{\epsilon^2}{3} \bar{X}\right), \exp\left(-\frac{\epsilon}{2} \bar{X}\right)\right\}
\]
\[
\mathbb{P}[X \leq (1 - \epsilon) \bar{X}] \leq \exp\left(-\frac{\epsilon^2}{2} \bar{X}\right)
\]
\[
\mathbb{P}[|X - \bar{X}| \geq \epsilon \bar{X}] \leq 2 \max\left\{\exp\left(-\frac{\epsilon^2}{3} \bar{X}\right), \exp\left(-\frac{\epsilon}{2} \bar{X}\right)\right\}
\]

Lemma A.2 (McDiarmid [18]). Let $X_1, \ldots, X_n$ be a martingale adapted to filtration $(\mathcal{F}_n)$ satisfying
(i) $\text{Var}(X_i|\mathcal{F}_{i-1}) \leq \sigma^2_i$, for $1 \leq i \leq n$, and
(ii) $|X_i - X_{i-1}| \leq M$, for $1 \leq i \leq n$. 

\[ Let \ X = \sum_{i=1}^{n} X_i. \ Then \]
\[ \mathbb{P}[X - \mathbb{E}X \geq r] \leq \exp \left( -\frac{r^2}{2(\sum_{i=1}^{n} \sigma_i^2 + Mr/3)} \right). \]

**Lemma A.3** (Balls and bins: tail bound for number of nonempty bins). *Suppose \( m \leq n/4 \). If \( m \) balls are placed into \( n \) bins each independently and uniformly at random, then with probability at least \( 1 - \exp(-m/2) \) at least \( m/2 \) bins are nonempty.*

**Proof.** Any configuration with at most \( m/2 \) nonempty bins has at least \( n - m/2 \) empty bins. Thus we may bound the probability of having some set of \( n - m/2 \) bins be empty. There are \( \binom{n}{n-m/2} = \binom{n}{m/2} \) possible choices for these empty bins, and each ball has to land outside of these, which has probability \( [(m/2)/n]^m \). Thus, the probability of at most \( m/2 \) nonempty bins is bounded by
\[ \binom{n}{m/2} \left( \frac{m/2}{n} \right)^m \leq \left( \frac{n \cdot e}{m/2} \right)^{m/2} \left( \frac{m/2}{n} \right)^m \leq \left( \frac{m e}{2n} \right)^{m/2} \leq \exp(-m/2), \]
where we used \( m \leq n/4 \).

The following lemma records a simple consequence of linearity of expectation.

**Lemma A.4** (Balls and bins: bound for the expected number of nonempty bins). *If we throw \( m \) balls into \( n \) bins independently uniformly at random, then, the expected number of nonempty bins is \( n(1 - (1 - 1/n)^m) \).*

### B  Converting to anytime regret

The *doubling trick* converts an online algorithm designed for a finite known time horizon to an algorithm which does not require knowledge of time horizon and yet provides the same performance (up to multiplicative constant) for any time \([9]\). This is the so called *anytime regret*.

The trick is to divide time into intervals and restart algorithm at the beginning of each interval. Let \( A(\theta,T) \) be an online algorithm with known time horizon. It takes the model parameters \( \theta \) and the time horizon \( T \) as the inputs and achieves regret \( R(T) \) at time \( T \). For example if \( R(T) = O(T^n) \), then to achieve anytime regret, the doubling trick suggests using the time intervals of length \( 2, 2^2, 2^3, \ldots, 2^m \). This achieves regret of at most \( R(T)/(1 - 2^n) \) at time \( T \) for any \( T \). Alternatively, if \( R(T) = O(\log T) \), then using the intervals of length \( 2^2, 2^2, \ldots, 2^m \) achieves regret of at most \( 4R(T) \) at time \( T \) for any \( T \).

If the algorithm achieves regret curve which is \( O(\log T) \) if \( T < T_1 \) and \( O(\sqrt{T}) \) if \( T \geq T_1 \), then intervals of different sizes should be used before and after \( T_1 \).

### References

[1] S. Aditya, O. Dabeer, and B. K. Dey. A channel coding perspective of collaborative filtering. *IEEE Transactions on Information Theory*, 57(4):2327–2341, 2011.

[2] K. Barman and O. Dabeer. Analysis of a collaborative filter based on popularity amongst neighbors. *IEEE Transactions on Information Theory*, 58(12):7110–7134, 2012.

[3] A. Bellogin and J. Parapar. Using graph partitioning techniques for neighbour selection in user-based collaborative filtering. In *Proceedings of the sixth ACM Conference on Recommender Systems*, pages 213–216. ACM, 2012.
[4] G. Biau, B. Cadre, and L. Rouviere. Statistical analysis of k-nearest neighbor collaborative recommendation. *The Annals of Statistics*, 38(3):1568–1592, 2010.

[5] G. Bresler, G. H. Chen, and D. Shah. A latent source model for online collaborative filtering. In *Advances in Neural Information Processing Systems*, pages 3347–3355, 2014.

[6] G. Bresler, D. Shah, and L. F. Voloch. Collaborative filtering with low regret. In *SIGMETRICS Performance Evaluation Review*, volume 44, pages 207–220, New York, NY, USA, June 2016. ACM.

[7] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

[8] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717, 2009.

[9] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

[10] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics*, 3(1):79–127, 2006.

[11] O. Dabeer. Adaptive collaborating filtering: The low noise regime. In *International Symposium on Information Theory Proceedings (ISIT)*, pages 1197–1201. IEEE, 2013.

[12] A. S. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*, pages 271–280. ACM, 2007.

[13] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM Symposium on Theory of Computing*, pages 665–674. ACM, 2013.

[14] R. H. Keshavan, A. Montanari, and S. Oh. Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998, 2010.

[15] B.-H. Kim, A. Yedla, and H. D. Pfister. Imp: A message-passing algorithm for matrix completion. In *Turbo Codes and Iterative Information Processing (ISTC), 2010 6th International Symposium on*, pages 462–466. IEEE, 2010.

[16] T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[17] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80, 2003.

[18] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, pages 195–248. Springer, 1998.

[19] S. Negahban and M. J. Wainwright. Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *Journal of Machine Learning Research*, 13(May):1665–1697, 2012.
[20] A. Rohde and A. B. Tsybakov. Estimation of high-dimensional low-rank matrices. *The Annals of Statistics*, 39(2):887–930, 2011.

[21] D. Russo and B. Van Roy. Learning to optimize via information-directed sampling. In *Advances in Neural Information Processing Systems*, pages 1583–1591, 2014.

[22] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[23] B.-H. Shen, S. Ji, and J. Ye. Mining discrete patterns via binary matrix factorization. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 757–766. ACM, 2009.

[24] D. Song, C. E. Lee, Y. Li, and D. Shah. Blind regression: Nonparametric regression for latent variable models via collaborative filtering. In *Advances in Neural Information Processing Systems*, pages 2155–2163, 2016.

[25] N. Srebro, N. Alon, and T. S. Jaakkola. Generalization error bounds for collaborative prediction with low-rank matrices. In *Advances In Neural Information Processing Systems*, pages 1321–1328, 2005.

[26] J. Wang, A. P. De Vries, and M. J. Reinders. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 501–508. ACM, 2006.

[27] J. Xu, R. Wu, K. Zhu, B. Hajek, R. Srikant, and L. Ying. Jointly clustering rows and columns of binary matrices: Algorithms and trade-offs. In *ACM SIGMETRICS Performance Evaluation Review*, volume 42, pages 29–41. ACM, 2014.