

## MIT Open Access Articles

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

The MIT Faculty has made this article openly available. **Please share**  
how this access benefits you. Your story matters.

**Citation:** Bresler, Guy and Karzand, Mina. 2021. "Regret Bounds and Regimes of Optimality for User-User and Item-Item Collaborative Filtering." IEEE Transactions on Information Theory, 67 (6).

**As Published:** 10.1109/TIT.2021.3071345

**Publisher:** Institute of Electrical and Electronics Engineers (IEEE)

**Persistent URL:** <https://hdl.handle.net/1721.1/142891>

**Version:** Original manuscript: author's manuscript prior to formal peer review

**Terms of use:** Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International



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

Guy Bresler\* and Mina Karzand†

May 8, 2019

## 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. Each user may be recommended a given item at most once. 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. We assume that the matrix encoding the preferences of each user type for each item type is randomly generated; in this way, the model captures structure in both the item and user spaces, the amount of structure depending on the number of each of the types. The measure of performance of the recommendation system is the expected number of disliked recommendations per user, defined as expected regret. We propose 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, actors, director of movies). In contrast, collaborative filtering is based on observed user

---

\*Guy Bresler is with the Department of Electrical Engineering and Computer Science at MIT and a member of LIDS and IDSS. [guy@mit.edu](mailto:guy@mit.edu)

†Mina Karzand is with the Wisconsin Institute of Discovery at University of Wisconsin, Madison. The work was done when the author was in the Department of Electrical Engineering and Computer Science at MIT and a member of LIDS and IDSS. [karzand@wisc.edu](mailto:karzand@wisc.edu)

‡The authors are ordered alphabetically.

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, 2, 3, 4, 5, 6, 7, 8]. In this paper, since our model has no item and user features, all algorithms must do collaborative filtering.

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 [9, 10, 11]. In the multi-armed bandit setup, optimal 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 [12, 13]).

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.

With the aforementioned issues in mind, we work within a mathematical framework for evaluating the performance of various recommendation system algorithms, related to the models studied in [12, 13]. The framework is detailed in Section 2, but in brief, at each time-step each user in the system is given a recommendation and then provides binary feedback in the form of 'like' or 'dislike'. The user preferences are described by 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 have identical preferences for all items and items belonging to the same type have identical ratings from all users.<sup>1</sup> The basic measure of performance is *expected regret*, defined as the expected number of bad recommendations made per user over a time horizon of interest. A second performance criterion is the *cold start time*, the first time at which recommendations become nontrivial in quality. Our goal is to understand the dependence of these quantities on system parameters and we will therefore seek bounds accurate only to within constant or logarithmic factors.

In the literature there are two categories of collaborative filtering (CF) algorithms. *User-user* algorithms [16, 12, 17] 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 [13, 18, 19] 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

---

<sup>1</sup>A similar model of data to ours, in which there is an underlying clustering of rows and columns, has been studied in other settings [14, 15].

users that have rated both  $i$  and  $i'$ . In Sections 4 and 6 we develop versions of user-user and item-item CF algorithms tailored to our online recommendation system model and prove performance guarantees. In order to achieve good performance, these algorithms must carefully explore the *a priori* unknown relationships between users and items. One of the unexpected insights that emerge from the analysis is that the item-item algorithm must limit the exploration to only a subset of the items types, where the size of this subset depends on the system parameters and time-horizon. The straightforward approach to Item-Item CF algorithms is to learn the whole preference matrix, and as described in Section 3 this results in a qualitatively suboptimal cold-start time that can be arbitrarily worse than the one obtained by our algorithm.

In order to focus on the information structure of the recommendation problem, and the associated exploration-exploitation tradeoff, the majority of the paper assumes that user feedback is noiseless. We generalize our user-user algorithm to handle noisy feedback and also describe how one would similarly accommodate noisy feedback in the item-item algorithm. In essence, estimation of similarity between users (or items) requires some redundancy in the information collected in order to average out the noise.

We prove nearly tight lower bounds on regret for two parameter regimes of interest, identifying settings in which the proposed algorithms cannot be significantly improved. In the *user structure only* scenario, the model parameters are such that there is no structure in the item space. Analogously, in the *item structure only* scenario, the parameters are such that there is no structure in the user space. We prove information-theoretic lower bounds for the performance of any algorithm in the user-structure only and item-structure only models, which match to within a logarithmic factor the performance obtained by our proposed user-user and item-item CF algorithms. These results are outlined in Section 3.

One of this paper’s main contributions is the development of techniques for proving lower bounds on the performance of online recommendation algorithms. Our lower bounds depend crucially on the inability to repeatedly recommend the same item to a given user, and for this reason are completely different from lower bounds for multi-armed bandit problems [10, 9]. At a high level, however, the basic challenge is the same as when proving lower bounds for bandits: one must connect the information obtained by the algorithm to the regret incurred. This allows to reason that subsequent recommendations will have low regret only if prior recommendations yielded significant information, which in turn necessitated exploratory recommendations with correspondingly substantial regret. Thus, regret is a conserved quantity and cannot be avoided by employing complicated adaptive algorithms.

The methods used for the lower bounds are elementary in nature. For example, in the user structure only model, the arguments in Section 7 are based on two observations. First, one 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, and in agreement, 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. Second, the outcome of recommending item  $i$  to user  $u$  at time  $t$  is also uncertain if none of the users that actually are 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 reasoning leads to lower bounds in Section 5 for the model with only item-structure.

A few papers including [12, 13, 20] have theoretical analyses for online collaborative filtering. The paper [12] analyzes a user-user CF algorithm in a similar setting to ours and [13] 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 algorithms and analysis. The model studied by Dabeer and coauthors [20, 1, 22] is also quite similar to our setup, but their objective is different: they seek an algorithm that *exploits* in a provably optimal fashion asymptotically in time, but their approach does not reveal how to explore. In a different direction, Kerenidis and Prakash [21] seek to achieve low computational complexity for recommendation in a similar setup as ours. What they show is that reconstructing the preference matrix only partially, which is what our item-item CF algorithm does, is useful also with regards to computation.

Hybrid algorithms exploiting both structure in user space and item space have been studied before in [23, 24, 25]. Both Song et al. [23] and Borgs et al. [26] study 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 and includes numerical simulations to complement the theoretical analyses. 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. 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. Appendix A contains a few basic probabilistic lemmas, and Appendix B relates so-called anytime regret (unknown time horizon) to known time horizon.

## 1.2 Notation

For an integer  $a$  we write  $[a] = \{1, \dots, a\}$  and for real-valued  $x$  let  $(x)_+ = \max\{x, 0\}$ . All logarithms are to the base of 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  $x$  and  $\lceil x \rceil$  denotes the smallest integer greater than or equal to  $x$ . Numerical constants ( $c, c_1, c_2$  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  $\{1, \dots, N\}$ . At each time  $t = 1, 2, 3, \dots$  the algorithm recommends an item  $a_{u,t} \in \mathbb{N}$  to each user  $u$  and receives feedback  $L_{u,a_{u,t}} \in \{+1, -1\}$  (‘like’ or ‘dislike’). For the reasons stated 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  $\mathcal{H}_t = \{a_{u,s}, L_{u,a_{u,s}}, \text{ for } u \in [N], s \in [t]\}$  is the collection of actions and feedback up to time  $t$ . We are interested in online learning algorithms, in which the action  $a_{u,t}$  is a (possibly random) function of the history up through the end of the previous time-step  $\mathcal{H}_{t-1}$ . This additional randomness is encoded in a random variable  $\zeta_{u,t}$ , assumed to be independent of all other variables. In this way,  $a_{u,t} = f_{u,t}(\mathcal{H}_{t-1}, \zeta_{u,t})$ , 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:

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

Here the expectation is with respect to the randomness in both the model and the algorithm. The algorithms we describe depend on knowing the time-horizon  $T$ , but by a standard doubling trick (explained in Appendix B) it is possible to convert these to algorithms achieving the same (up to constant factors) regret without this knowledge (see, e.g., [27]). 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.

The time at which point recommendations become nontrivial in quality is another important performance criterion, because until that point users invest effort but get little in return. 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. We define the *cold start time* to be the first time at which the slope of regret as a function of  $T$  is bounded by some value  $\gamma$ :

$$\text{coldstart}(\gamma) = \min \left\{ T : \frac{\text{regret}(T)}{T} \leq \gamma \right\}.$$

This is similar to (but somewhat simpler than) the definition in [13].

### 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. uniform on  $[\mathbf{q}_U]$ , where  $\mathbf{q}_U$  is the number of user types.

$N$	Number of users
$T$	Time horizon
$\mathbf{q}_U$	Number of user types
$\mathbf{q}_I$	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.

We assume that  $\mathbf{q}_U \leq N$ , because if  $\mathbf{q}_U > N$  then most users have their own type and all of the results remain unchanged upon replacing  $\mathbf{q}_U$  by  $N$ . Similarly, each item  $i \in \mathbb{N}$  has a random item type  $\tau_I(i)$  i.i.d. uniform on  $[\mathbf{q}_I]$ , where  $\mathbf{q}_I$  is the number of item types<sup>2</sup>. 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*  $\Xi = (\xi_{k,j}) \in \{-1, +1\}^{\mathbf{q}_U \times \mathbf{q}_I}$ , which specifies the preference of each user type for each item type. The preference  $L_{u,i}$  of user  $u \in [N]$  for item  $i \in \mathbb{N}$  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 straightforward. However, the independence assumption is quite strong and an important future research direction is to obtain results for more realistic preference matrices. We also consider a noisy model with  $L_{u,i} = \xi_{\tau_U(u), \tau_I(i)} \cdot z_{u,i}$  where  $z_{u,i}$  are i.i.d. random variables with  $\mathbb{P}[z_{u,i} = +1] = 1 - \gamma$  and  $\mathbb{P}[z_{u,i} = -1] = \gamma$ .

### 2.3 Two regimes of interest

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, each of user-user or item-item CF is almost optimal in the corresponding regime.

**Definition 2.1** (User structure only ( $\mathbf{q}_I = 2^{\mathbf{q}_U}$ )). 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  $\Xi \in \{-1, +1\}^{\mathbf{q}_U \times 2^{\mathbf{q}_U}}$  is deterministic and has columns consisting of all sequences in  $\{-1, +1\}^{\mathbf{q}_U}$ . Essentially the same preference matrix would arise (with high probability) if  $\mathbf{q}_I$  is much larger than

---

<sup>2</sup>Throughout the paper, we will assume that the number of user types  $\mathbf{q}_U$  and the number of item types  $\mathbf{q}_I$  are both  $\Omega(\log N)$ , since the whole problem becomes easy if either of these parameters are too small.

$2^{q_U}$  (when the entries are i.i.d. as specified above in Subsection 2.2).

**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

We will 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 much better regret (up to multiplicative logarithmic factors) in the specific extreme parameter regimes with user-structure only and item-structure only. The simplified versions of these theorems appear in this section. Towards the end of this section we present simulation results supporting the theorems.

#### 3.1 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 (as opposed to each user individually). 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 made 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  $q_I > 126 \log N$  item types. There exists numerical constants  $c, C$  so that Algorithm 1 achieves regret*

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

The cold-start time, the time until the slope of the regret drops below  $\gamma$ , is evidently  $\Theta(\log N)$  for any  $\gamma \in (\frac{Cq_U}{N}, \frac{1}{2})$ . It follows from the next theorem that 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 achieves both regret and cold start time that are optimal up to multiplicative constants.

**Theorem 5.1** (Regret lower bound with user structure only, simplified version). *There exist a numerical constant  $c$  such that in the user structure model (Defn 2.1) with  $q_U > (\log N)^{1.1}$  user types and  $N > N_0$  users, any recommendation algorithm must incur regret*

$$\text{regret}(T) \geq \begin{cases} 0.49T - 4, & \text{if } T \leq c \log q_U \\ 0.2 \frac{q_U}{N} T, & \text{if } T > c \log q_U. \end{cases}$$

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 and that its duration depends on the number of user types. In [13] 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.

### 3.2 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. The algorithm creates several clusters of items, as well as a set of unclustered items. Similarity of two items is estimated by having random users rate both items. Users then explore a single item from each cluster and liked clusters are subsequently recommended. The effort of clustering is shared amongst all the users, and the savings is due to liked explorations yielding an entire cluster of items to recommend.

Crucially, this version of item-item CF has the feature that only a *subset* of the item space is explored (*i.e.*, only a subset of the item types are clustered, with the others cast aside)<sup>3</sup>. 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. The total number of

---

<sup>3</sup>In comparison, learning the entire preference matrix requires  $q_I$  explorations per user, resulting in a cold start time of at least  $q_I$  versus our  $q_I/N$ . The more fundamental savings in clustering only a portion of the space is in reducing the shared cost, and thereby again reducing the cold start time. A naive clustering of  $M$  items from the entire space requires  $q_I(\log q_I)M$  comparisons (each item is compared  $\log q_I$  times to a representative from each of  $q_I$  types), and for the clusters to be larger than constant size on average,  $M$  must be significantly larger than  $q_I$ . The resulting cold start time is thus at least  $q_I^2(\log q_I)/N$ , which is much larger than our cold start time on the order of  $q_I/N$ .

items compared and the number of clusters are chosen depending on the system parameters to give the best regret bound.

**Theorem 6.1** (Regret upper bound in item-item CF, simplified version). *Consider the recommendation system model described in Section 2 with  $N > 5$  users,  $q_I > 13 \log N$  item types, and  $q_U > 16 \log(N q_I)$  user types. There are numerical constants  $C, c_1, c_2$  and  $c_3$  such that Algorithm 3 obtains regret per user at time  $T$  upper bounded as*

$$\text{regret}(T) \leq C \begin{cases} T, & \text{if } T \leq \max\{c_1, c_2 \frac{q_I \log(N q_I)}{N}\} \\ \log T + \sqrt{\frac{q_I \log(N q_I)}{N}} T, & \text{if } \max\{c_1, c_2 \frac{q_I \log(N q_I)}{N}\} < T \leq c_3 \frac{N q_I}{\log(N q_I)} \\ \frac{\log(N q_I)}{N} T, & \text{if } c_3 \frac{N q_I}{\log(N q_I)} \leq T. \end{cases}$$

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

**Theorem 7.1** (Regret lower bound for item structure only, simplified version). *In the item-structure model (Defn. 2.2) with  $q_I > 25 (\log N)^5$  item types and  $N > 32$  users, there exist numerical constants  $C, c_1, c_2, c_3$ , and  $c_4$  such that any recommendation algorithm must incur regret*

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

It follows that the cold start time  $\text{coldstart}(\gamma)$  with  $\gamma = C$  in the item-structure only regime is lower bounded as  $\tilde{\Omega}(\sqrt{q_I}/N)$ , while the upper bound based on our proposed algorithm is  $\tilde{O}(q_I/N)$ . Note that the cold start time with  $\gamma = 1/\log q_I$  is  $\tilde{\Theta}(q_I/N)$ . The gap in the upper and lower bounds on cold start time is a consequence of the existence of the second regime in the lower bound given above, and appears to be an artifact of our proof.

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 since this reveals a new variable in the preference matrix. Lower bounding the number of such uncertain recommendations gives the lower bound for regret.

If there is structure in the item space it is possible to avoid the long cold-start time of algorithms 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. Thus increasing  $N$  makes it easier to make meaningful recommendations. The same phenomenon is true more generally: the upper bound on regret at any time  $T$  is a decreasing function of  $N$ .

### 3.3 Numerical Simulations

We simulated our versions of USER-USER and ITEM-ITEM Algorithms (As described in Sections 4 and 6). In Figure 2, we plot the regret as a function of time for the USER-USER Algorithm (Alg. 1 in Section 4). We observe that the slope of regret in the asymptotic regime increases by increasing  $q_U$  for fixed  $N$ . We also observe that increasing  $N$  decreases the asymptotic slope but does not decrease the cold start time of the algorithm.

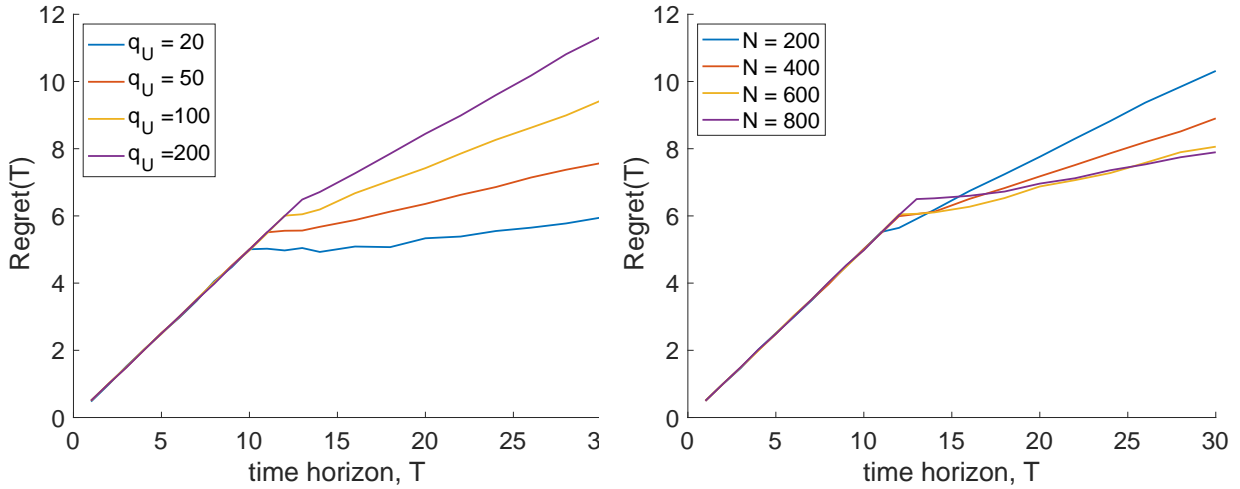


Figure 2: Simulated performance for Algorithm USER-USER. System parameters are (a)  $N = 400$  and  $q_I = 100$  and (b)  $q_U = 80$  and  $q_I = 100$ .

In Figure 3, we plot the regret as a function of time for the ITEM-ITEM Algorithm (Alg. 3 in Section 6). We observe that with fixed  $N$ , increasing  $q_I$  increases the cold-start time. But with fixed  $q_I$ , the cold-start time shrinks linearly in  $N$ . We also observe that the slope of regret after the cold start time increases with increasing  $q_I$  and decreasing  $N$ , consistent with the statement of Theorem 6.1.

## 4 User-user algorithm and analysis

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

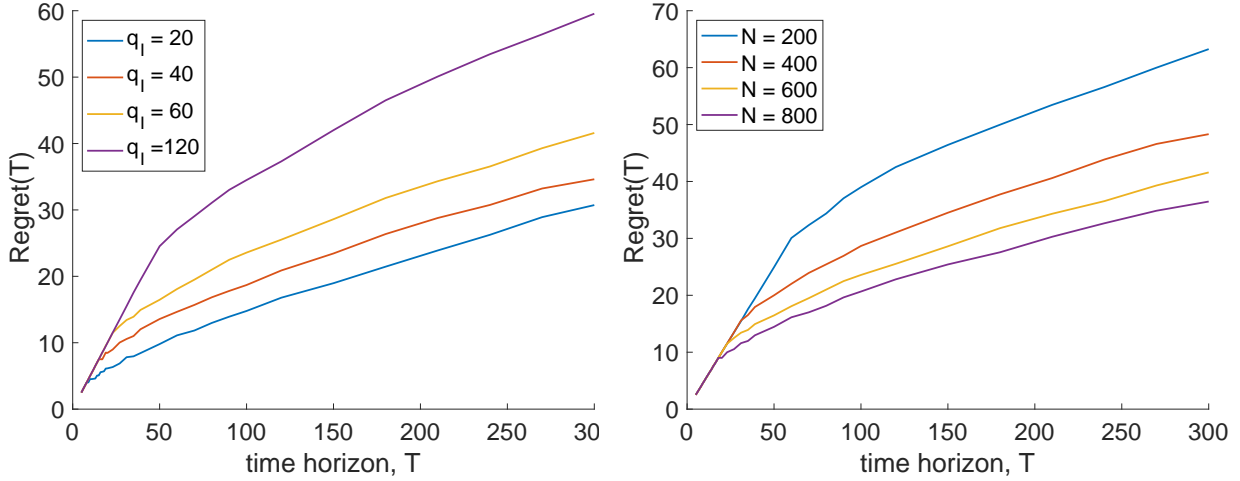


Figure 3: Simulated performance for Algorithm ITEM-ITEM. System parameters are (a)  $N = 600$  and  $q_U = 100$  and (b)  $q_I = 60$  and  $q_U = 100$ .

#### 4.1 Algorithm

Pseudocode for algorithm USER-USER appears as 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  $\{\mathcal{P}_k\}_k$  of users that recovers the user types correctly with high probability. In Step 2, users are recommended new random items (*exploration*) until an item is liked. If the user is in group  $\mathcal{P}_k$  of the partition, the item is added to a set  $\mathcal{S}_k$  of items to be recommended to all other users in the same partition (*exploitation*). Step 2 (find and recommend items) is repeated indefinitely.

**Remark 4.1.** *Our model assumes that users of the same type have identical ratings. Hence, users of the same type are always in the same group after partitioning. However, due to random sampling of the items in exploration, users from different types can have identical ratings for the items recommended in Step 1, in which case they will end up in the same partition. It follows that the total number of groups in the user partition is at most  $q_U$ .*

We make a few additional remarks regarding the algorithm:

- 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, the sets of items  $\{\mathcal{S}_k\}$  at each time contain the items exploitable by users in the  $k$ -th group in the partition. The algorithm predicts that all users in the  $k$ -th group like items in  $\mathcal{S}_k$ .
- The algorithm takes  $T$ ,  $q_U$ , and  $N$  as input. As mentioned in Section 2, a doubling trick described in Appendix B converts the algorithm to one oblivious to  $T$ . It is also fairly straightforward to modify the algorithm to be adaptive to  $q_U$ . The adaptive algorithm initializes with a trivial partition placing all users in one group. The algorithm subsequently refines

---

**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**  $t = 1, \dots, r$  **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  $\hat{\tau}_U(u) \in [q_U]$  be the label of user  $u$ 's partition.
- 6:  $\mathcal{P}_k = \{u \in [N] : \hat{\tau}_U(u) = k\}$ , for all  $k \in [q_U]$

*Step 2: find and recommend items*

- 7:  $\mathcal{S}_k \leftarrow \emptyset$ , for all  $k \in [q_U]$
  - 8: **for**  $t = r + 1, \dots, T$  **do**
  - 9:     **for**  $u \in [N]$  **do**
  - 10:         **if**  $\mathcal{S}_{\hat{\tau}_U(u)} \setminus \{a_{u,1}, \dots, a_{u,t-1}\} \neq \emptyset$  (i.e.,  $u$  has not rated all items in  $\mathcal{S}_{\hat{\tau}_U(u)}$ ) **then**
  - 11:              $a_{u,t} \leftarrow$  an unrated item in  $\mathcal{S}_{\hat{\tau}_U(u)}$  (*exploit*)
  - 12:         **else**
  - 13:              $a_{u,t} \leftarrow$  random item not rated by any user (*explore*)
  - 14:             **if**  $L_{u,a_{u,t}} = +1$  **then**
  - 15:                  $\mathcal{S}_{\hat{\tau}_U(u)} \leftarrow \mathcal{S}_{\hat{\tau}_U(u)} \cup \{a_{u,t}\}$
- 

the partition whenever a user's feedback indicates that they have been grouped incorrectly. We chose not to do so since it complicates the analysis.

**Theorem 4.1.** *Consider the model introduced in Section 2 with  $N$  users,  $q_U$  user types and  $q_I$  item types. Let  $r = \lceil 2\log(Nq_U^2) \rceil$ . If  $q_I > 18r$ , then USER-USER achieves regret*

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

The assumption  $q_I > 18r$  ensures that with probability  $1 - o(\frac{1}{N})$  for each user type, there is at least one item type that is liked. This assumption also ensures that with probability  $1 - o(\frac{1}{N})$ , for any pair of user types, there is at least one item type which is rated differently by them. If there is no such item type, then the two user types rate everything similarly and are indistinguishable.

The theorem indicates that up until time  $r$ , the algorithm is making meaningless (randomly chosen independent of feedback) recommendations. Random recommendations have probability half of being liked, hence incur regret with slope  $1/2$ . After that, the algorithm achieves the asymptotic slope indicating that on average  $q_U$  recommendations out of  $N$  are random. The simplified version of this theorem in Section 3 is obtained using  $2\log N < r < 7\log N$  (since  $q_U \leq N$ ). We also pick the constant  $C$  large enough so that  $\frac{r}{2} + \frac{2q_U+2}{N}T + 2 \leq C(\log N + \frac{q_U}{N}T)$  for  $T > r$ .

## 4.2 Proof of Theorem 4.1

We first bound the probability that the partition created by the algorithm is correct in Lemma 4.2. Next, to prove the theorem we will show that conditioned on the partition being correct, the number of exploratory recommendations (and hence the regret) is upper bounded.

**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 USER-USER. Let  $\epsilon$  and  $r$  be as defined there. If  $q_I > 4r$ , then  $\mathbb{P}[B_{uv}^c] \leq \frac{2\epsilon}{q_U^2}$ . It follows that if  $B = \bigcap B_{uv}$  is the event that all users are partitioned correctly, then  $\mathbb{P}[B] > 1 - \epsilon$ .*

*Proof.* As observed in Remark 4.1, users from the same partition rate items identically. Therefore the only way an error in partitioning occurs is if users of different types are grouped together. This happens when two users rate all exploratory items identically in Step 1. In Step 1, the first  $r$  items recommended to all users are chosen uniformly at random independent of feedback, so the types of these items are uniformly distributed on  $[q_I]$ . Let  $s$  be the number of items with distinct item types among the  $r$  exploratory items from Step 1. This is a balls and bins scenario with  $r$  balls into  $q_I$  bins, and Lemma A.3 states that if  $q_I > 4r$ , then  $\mathbb{P}[s < r/2] \leq \exp(-r/2) \leq \epsilon/q_U^2$ . By symmetry, each of the types of the  $s$  items with distinct types is uniformly distributed on  $[q_I]$ .

Since all users rate the same items and users of the same type have identical preferences, as far as the lemma is concerned we only consider how the user types themselves rate 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  $2^{-s} < 2^{-r/2} \leq \epsilon/q_U^2$ .

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

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

The second statement in the lemma follows by union bounding over  $\binom{q_U}{2} \leq q_U^2/2$  pairs of user types.  $\square$

*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  $u \in [N]$ . It follows that, for  $T \leq r$ ,

$$\mathbb{E} \left[ \sum_{t=1}^T \frac{1}{N} \sum_{u=1}^N \mathbb{1}[L_{u,a_{u,t}} = -1] \right] = \sum_{t=1}^T \frac{1}{N} \sum_{u=1}^N \mathbb{P}[L_{u,a_{u,t}} = -1] = \frac{T}{2}. \quad (2)$$

Now consider the case  $T > r$ . At  $t = r$ , by Lemma 4.2, the partitioning step recovers the user types correctly with probability at least  $1 - \epsilon$ , i.e.,  $\mathbb{P}[B] > 1 - \epsilon$ . On event  $B$  all users in a partition have the same type, so by construction of the sets  $\mathcal{S}_{\hat{\tau}_U(u)}$  in Line 15 of USER-USER, items in  $\mathcal{S}_{\hat{\tau}_U(u)}$  are

liked by at least one user of the same type as  $u$  and therefore also by  $u$ , and

$$\mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in \mathcal{S}_{\hat{\tau}_U(u)}] \middle| B \right] = 0. \quad (3)$$

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

$$\mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in \mathcal{S}_{\hat{\tau}_U(u)}] \right] \leq TN\epsilon = T. \quad (4)$$

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

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

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

A Chernoff bound (Lemma A.1) applied to the i.i.d.  $\xi$  variables gives  $\mathbb{P}[C^c] \leq \mathbf{q}_U \exp(-\mathbf{q}_I/36) \leq 1/N$ , where the last inequality due to  $\mathbf{q}_I > 18r$ . Conditioning on event  $C$ , we get

$$\begin{aligned} & \mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}] \right] \\ & \leq \mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}, C] \right] + NT\mathbb{P}(C^c). \end{aligned} \quad (5)$$

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 that were disliked. The number of liked explorations is easier to deal with, because of a self-limiting effect: these result in items added to sets  $\{\mathcal{S}_k\}$  for exploitation, and exploration only happens when there are not enough items to be exploited.

We now relate the expected number of liked and disliked explorations. At  $t > r$  if  $a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}$ , then it means the item is an exploratory recommendation and thus  $a_{u,t}$  is an independent new random item with uniformly random type  $\tau_I(a_{u,t}) \in [\mathbf{q}_I]$ . Hence, using the definition of event  $C$ ,

$$p := \mathbb{P}[L_{u,a_{u,t}} = +1 | a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}, C] \geq 1/3 \quad (6)$$

and  $1 - p \leq 2p$ . It follows that

$$\mathbb{P}[L_{u,a_{u,t}} = -1, a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}, C] \leq 2\mathbb{P}[L_{u,a_{u,t}} = +1, a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}, C]. \quad (7)$$

This means that to bound the first term in (5) it suffices to bound the contribution from the sum with  $L_{u,a_{u,t}} = +1$ , as derived 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*

$$\sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a_{u,t}} = +1, a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}, C] \leq Tq_U + N.$$

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

$$\sum_{t=r+1}^T \sum_{u: \hat{\tau}_U(u)=k} \mathbb{1}[L_{u,a_{u,t}} = +1, a_{u,t} \notin \mathcal{S}_k^{t-1}, C] \leq \sum_{t=r+1}^T \sum_{u: \hat{\tau}_U(u)=k} \mathbb{1}[L_{u,a_{u,t}} = +1, a_{u,t} \notin \mathcal{S}_k^{t-1}] = |\mathcal{S}_k^T|. \quad (8)$$

The rest of the proof entails bounding  $|\mathcal{S}_k^T|$ . The number of items added to  $\mathcal{S}_k$  at time  $t$  is

$$|\mathcal{S}_k^t| - |\mathcal{S}_k^{t-1}| = \sum_{u: \hat{\tau}_U(u)=k} \mathbb{1}[L_{u,a_{u,t}} = +1, a_{u,t} \notin \mathcal{S}_k^{t-1}] \leq \sum_{u: \hat{\tau}_U(u)=k} \mathbb{1}[a_{u,t} \notin \mathcal{S}_k^{t-1}]. \quad (9)$$

If  $|\mathcal{S}_k^{t-1}| \geq t$ , then  $\mathcal{S}_k^{t-1} \setminus \{a_{u,1}, \dots, a_{u,t-1}\} \neq \emptyset$ . Meanwhile, at time  $t$ , the exploration event (recommending  $a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}^{t-1}$  in line 13) happens only if there are no items left in  $\mathcal{S}_{\hat{\tau}_U(u)}^{t-1}$  for user  $u$  to exploit, i.e.,  $\mathcal{S}_{\hat{\tau}_U(u)}^{t-1} \setminus \{a_{u,1}, \dots, a_{u,t-1}\} = \emptyset$ . In this way,  $|\mathcal{S}_k^{t-1}| \geq t$  guarantees that there is an exploitable item at time  $t$  for each user in  $\mathcal{P}_k$ . Consequently,

$$\sum_{u: \hat{\tau}_U(u)=k} \mathbb{1}[a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}^{t-1}] \begin{cases} = 0, & \text{if } |\mathcal{S}_k^{t-1}| \geq t \\ \leq |\mathcal{P}_k|, & \text{otherwise.} \end{cases} \quad (10)$$

The bound  $|\mathcal{P}_k|$  is due to the sum having  $|\mathcal{P}_k|$  terms, each upper bounded by 1.

Let  $t^* = \max\{t : r \leq t < T, |\mathcal{S}_k^{t-1}| < t\}$  be the last time for which we are *not* guaranteed (based on the reasoning before the last displayed eqn.) to have an exploitable item. Note that the set over which we take the maximum is nonempty if  $T > r$  since  $|\mathcal{S}_k^r| = 0$ . It follows that

$$|\mathcal{S}_k^T| = |\mathcal{S}_k^{t^*-1}| + \sum_{t=t^*}^{T-1} (|\mathcal{S}_k^{t+1}| - |\mathcal{S}_k^t|) \stackrel{(a)}{\leq} |\mathcal{S}_k^{t^*-1}| + (|\mathcal{S}_k^{t^*}| - |\mathcal{S}_k^{t^*-1}|) \stackrel{(b)}{\leq} T + |\mathcal{P}_k|.$$

Since for  $t^* < t < T$  we have  $|\mathcal{S}_k^{t-1}| \geq t$ , by (9) and (10) for these times we have  $|\mathcal{S}_k^{t+1}| - |\mathcal{S}_k^t| = 0$ . This gives (a) in the above display. By definition,  $|\mathcal{S}_k^{t^*-1}| < t^* < T$ . Inequality (b) uses (9) and (10) to bound  $|\mathcal{S}_k^{t^*}| - |\mathcal{S}_k^{t^*-1}|$ .

Note that  $\sum_{k \in [q_U]} |\mathcal{P}_k| = N$ . Using (8) and summing the last displayed inequality over the (at most)  $q_U$  partition indices proves the claim.  $\square$



We can now complete the proof of Theorem 4.1. By the preceding claim and Equations (4), (5), and (7) we get

$$\begin{aligned} \mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a_{u,t}} = -1] \right] &= \mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in \mathcal{S}_{\hat{\tau}_U(u)}] \right] \\ &\quad + \mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}] \right] \\ &\leq 2(q_U T + N) + 2T. \end{aligned}$$

For  $T > r$ , we can now bound the regret by combining Equation (2) with the previous display:

$$\begin{aligned} \text{regret}(T) &= \mathbb{E} \left[ \sum_{t=1}^r \frac{1}{N} \sum_{u=1}^N \mathbb{1}[L_{u,t} = -1] \right] + \mathbb{E} \left[ \sum_{t=r+1}^T \frac{1}{N} \sum_{u=1}^N \mathbb{1}[L_{u,a_{u,t}} = -1] \right] \\ &\leq \frac{1}{2}r + \frac{2(q_U + 1)}{N}T + 2. \end{aligned} \quad \square$$

### 4.3 User-User Algorithm with Noisy Preferences

We generalize the result to the scenario in which the feedback to the recommendation system is noisy. In this case, the preference of user  $u$  for item  $i$  is

$$L_{u,i} = \xi_{\tau_U(u), \tau_I(i)} \cdot z_{u,i}, \quad (11)$$

where  $z_{u,i}$  are i.i.d. random variables with  $\mathbb{P}[z_{u,i} = +1] = 1 - \gamma$  and  $\mathbb{P}[z_{u,i} = -1] = \gamma$  (we assume  $0 < \gamma < 1/2$ ). With probability  $\gamma$ , the preference of user  $u$  for item  $i$  is flipped relative to the preference of user type  $\tau_U(u)$  for item type  $\tau_I(i)$  in the preference matrix  $\Xi$ .

To accommodate the noisy feedback, we modify the partitioning subroutine in Step 1 of USER-USER algorithm with NOISYUSERPARTITION given in Algorithm 2. The main modification is that in Lines 5 and 6 of USER-USER algorithm, users are placed in the same partition if they rate all of the first  $r$  items similarly. Instead, users are now placed in the same partition if they rate the majority of the first  $r$  items similarly. The parameters  $\lambda$  and  $r$  are chosen to guarantee that the partitioning over users is consistent with their type with probability greater than  $1 - \epsilon$ .

**Remark 4.2.** In Line 8, the algorithm checks whether there is a partitioning over the users consistent with variables  $g_{u,v}$ . This is true precisely when the graph with edge set  $g_{u,v}$  is a disjoint union of cliques.

**Remark 4.3.** The noisy feedback decreases performance in two ways: partitioning users correctly requires more exploration recommendations, resulting in a larger cold-start time. Additionally, in Step 2, even good exploitation recommendations can be disliked due to noise. The next theorem quantifies these observations.

---

**Algorithm 2** NOISYUSERPARTITION( $T, \mathbf{q}_U, N, \gamma$ )

---

```
1:  $\epsilon \leftarrow 1/N, \lambda \leftarrow \frac{2}{3}(1 - 2\gamma)^2, r \leftarrow \lceil \frac{4}{(1-2\gamma)^2} \log(N^2/2\epsilon) \rceil,$ 
2: for  $t = 1, \dots, r$  do
3:   Pick random item  $i$ 
4:    $a_{u,t} \leftarrow i$ , for all  $u \in [N]$ 
5: for  $u = 1, \dots, N$  do
6:   for  $v = u + 1, \dots, N$  do
7:      $g_{u,v} \leftarrow \mathbb{1}\{\sum_{s=1}^r L_{ua_{us}} L_{va_{vs}} \geq \lambda r\}$ 
8: if There exists a partitioning over users consistent with the variables  $g_{u,v}$  then
9:   Let  $\hat{\tau}_U(u) \in [\mathbf{q}_U]$  be the label of user  $u$ 's partition.
10: else
11:    $\hat{\tau}_U(u) = 1$ , for all  $u \in [N]$  (all users are in one partition)
12:  $\mathcal{P}_k = \{u \in [N] : \hat{\tau}_U(u) = k\}$ , for all  $k \in [\mathbf{q}_U]$ 
```

---

**Theorem 4.4.** Consider the model introduced in Section 2 with  $N$  users,  $\mathbf{q}_U$  user types and  $\mathbf{q}_I$  item types. Let  $r = \lceil \frac{12}{(1-2\gamma)^2} \log N \rceil$ . If  $\mathbf{q}_I > 432 \log N$ , then USER-USER achieves regret

$$\text{regret}(T) \leq \begin{cases} \frac{1}{2}T, & \text{if } T \leq r \\ \frac{1}{2}r + \lceil \frac{5\mathbf{q}_U + 2}{N} + \gamma \rceil T + 5, & \text{if } T > r. \end{cases}$$

*Proof.* The proof of this theorem is very similar to the proof of Theorem 4.1. Lemma 4.5 replaces Lemma 4.2 to show that with the given choice of parameters in Algorithm 2, the partitioning  $\mathcal{P}_k$  is the same as partitioning over the users by their types with probability greater than  $1 - 1/N$ . Additionally, Equation (3) in the proof of Theorem 4.1 changes as follows to be consistent as a result of noisy feedback modeled in (11):

$$\mathbb{E} \left[ \sum_{t=r+1}^T \sum_{u \in [N]} \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in \mathcal{S}_{\hat{\tau}_U(u)}] \middle| B \right] \leq TN\gamma.$$

Equation (6) is replaced with

$$p := \mathbb{P}[L_{u,a_{u,t}} = +1 | a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}, C] \geq \frac{1-\gamma}{3}$$

and since  $\gamma < 1/2$ , then  $1 - p \leq 5p$ . It follows that

$$\mathbb{P}[L_{u,a_{u,t}} = -1, a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}, C] \leq 5\mathbb{P}[L_{u,a_{u,t}} = +1, a_{u,t} \notin \mathcal{S}_{\hat{\tau}_U(u)}, C]. \quad (12)$$

Claim 4.3 bounds the right-hand side. Plugging in these equations in the subsequent part of the proof of Theorem 4.1 gives the statement of the theorem.  $\square$

**Lemma 4.5.** Consider the user similarities computed in Step 7 of NOISYUSERPARTITION (Algo-

rithm 2). Define the event  $B_{uv} = \{g_{u,v} = \mathbb{1}_{\{\tau_U(u)=\tau_U(v)\}}\}$  that these similarities coincide with the underlying user types. If  $q_I > 144 \log(N^2/\epsilon)$ , then  $\mathbb{P}[B_{uv}^c] \leq 2\epsilon/N^2$ . It follows that if  $B = \bigcap B_{uv}$  is the event that all users are partitioned correctly, then  $\mathbb{P}[B^c] > 1 - \epsilon$ .

The proof of this lemma is similar to the proof of Lemma 4.2 and is deferred to Appendix C.

## 5 User structure only: lower bound

In this section we prove a lower bound on the regret of any online recommendation system in the regime with user structure only where  $q_I = 2^{q_U}$  as described in Definition 2.1.

**Theorem 5.1.** *Let  $\delta > 0$  and  $r = \lfloor \log q_U - \log(16(\log q_U) \log \frac{N}{\delta}) \rfloor$ . In the user structure model with  $N$  users and  $q_U$  user types, any recommendation algorithm must incur regret*

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

**Remark 5.1.** *The lower bound depends on a parameter  $\delta$  that has two effects: (1) the slope of the regret curve during the cold start grows (approaching 1/2) as the chosen parameter  $\delta$  shrinks to 0; (2) the cold start time  $r$  is upper bounded as  $r < \log q_U - \log \log(1/\delta)$ .*

*Additionally, if the number of user types satisfies  $q_U/N \rightarrow 0$ , the slope of regret after the cold start time (the asymptotic rate of regret) approaches  $\frac{q_U}{2N}$ . This is expected since each item can be recommended at most once to each user. Hence, even if the structure in the user space is known, the algorithm should explore new items. On average,  $q_U$  explorations (about half of which are disliked) are necessary for every  $N$  recommendations.*

*The simplified version of this theorem in Section 3 is obtained using  $N \geq q_U \geq (\log N)^{1.1}$ ,  $\delta = 1/100$  and  $N > N_0$  for a constant  $N_0$ .*

### 5.1 Proof strategy

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

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

Suppose that the preference matrix  $\Xi$  (with elements  $\xi_{k,j}$ , the preference of user types for item types) is known (which is the case in user-structure only model). Also, suppose that we have obtained feedback from some user  $u$  for  $t$  items. Relative to the total number of types, user  $u$  must belong to a restricted set of user types consistent with this feedback. If  $t$  is small, the set of consistent 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). At this point, 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.*

Since there is no structure in item space ( $\mathbf{q}_I = 2^{\mathbf{q}_U}$ ), learning the preference of a user type for an item is only achieved by recommending the item to one user from the user type. This is for the reason 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 statement of the theorem, 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** ( $(r, \epsilon)$ -column regularity). Let  $A \in \{-1, +1\}^{m \times n}$ . For ordered tuple of distinct (column) indices  $w = (i_1, \dots, i_r) \in [n]^r$ , let  $M = (A_{\cdot i})_{i \in w} \in \{-1, +1\}^{m \times r}$  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  $M$  that are identical to  $b$  and denote its cardinality by  $k_{b,w}(A)$ . The matrix  $A$  is said to be  $(r, \epsilon)$ -column regular if

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

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

*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  $(1-\epsilon)\frac{m}{2^{r-1}} \leq k_{b,w}(A) \leq (1+\epsilon)\frac{m}{2^{r-1}}$  for all size  $r-1$  tuples  $w$  and vectors  $b$ . For any given  $w \in [n]^{r-1}$  and  $b \in \{-1, +1\}^{r-1}$ , 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  $(1-\epsilon)\frac{m}{2^{r-1}} \leq k_{b,w} \leq (1+\epsilon)\frac{m}{2^{r-1}}$ .  $\square$

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

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

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

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

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.  $\square$

**Proposition 5.4.** *Consider the user structure only model in Definition 2.1. Let  $\delta > 0$  and  $r = \lfloor \log q_U - \log(16(\log q_U) \log \frac{N}{\delta}) \rfloor$ . For any  $T \leq r$ , the regret is lower bounded by*

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

*Proof.* We will show that for preference matrices  $\Xi$  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 recommended by the algorithm ( $n \leq Nt$  since each of the  $N$  users rates one item per time-step). We label the set of items by  $[n] = \{1, \dots, n\}$ . 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  $\tau_I(i) \sim \text{Unif}([2^{q_U}])$  and because the set of columns of the preference matrix  $\Xi$  is precisely  $\{-1, +1\}^{q_U}$ , the columns of  $A$  are independent and uniformly distributed in  $\{-1, +1\}^{q_U}$  according to this model.

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  $b = (L_{u,a_{u,s}})_{s \in [t-1]} \in \{-1, +1\}^{t-1}$  be the vector of feedback 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  $+1$ . Then  $L_{u,a_{u,t}} = +1$  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  $[n]$ ,

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

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  $\Omega_{t,\epsilon}$  was defined as the set of  $(t, \epsilon)$ -column regular matrices. It now follows by the

tower property of conditional expectation that

$$\begin{aligned}\mathbb{P}[L_{u,a_{u,t}} = +1 | A \in \Omega_{t,\epsilon}] &= \mathbb{E}[\mathbb{P}[L_{u,a_{u,t}} = +1 | A, \mathcal{H}_{t-1}] | A \in \Omega_{t,\epsilon}] \\ &= \mathbb{E}\left[\frac{k_{b^+, \{w,i\}}(A)}{k_{b,w}(A)} \middle| A \in \Omega_{t,\epsilon}\right] = \frac{1}{2} \frac{1+\epsilon}{1-\epsilon} \leq \frac{1}{2}(1+4\epsilon).\end{aligned}$$

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  $k_{b^+, \{w,i\}}(A) \leq (1+\epsilon)m/2^t$ . We pick  $\epsilon < 1/2$  to get the last inequality.

Fix  $\delta > 0$  and define

$$\epsilon_t = \sqrt{3 \frac{2^t}{q_U} \log \frac{(2N \log q_U)^t}{\delta}}.$$

Lemma 5.3 shows that at time  $t \leq r < \log q_U$ , for this choice of  $\epsilon_t$ , we have  $A \in \Omega_{t,\epsilon_t}$  with probability  $1 - \delta$ . We get the bound

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

It follows from the above display and the definition of  $\epsilon_t$  that for  $T \leq r$ ,

$$\begin{aligned}\text{regret}(T) &= \frac{1}{N} \sum_{t \in [T], u \in [N]} \mathbb{P}[L_{u,a_{u,t}} = -1] \stackrel{(a)}{\geq} \left(\frac{1}{2} - \delta\right)T - \sum_{t \in [T]} 2 \sqrt{3 \frac{2^t}{q_U} \log \frac{(2N \log q_U)^t}{\delta}} \\ &\stackrel{(b)}{\geq} \left(\frac{1}{2} - \delta\right)T - 4 \sqrt{3 \frac{2^{T+1}}{q_U} \log \frac{(2N \log q_U)^{\log q_U}}{\delta}} \stackrel{(c)}{\geq} \left(\frac{1}{2} - \delta\right)T - 4,\end{aligned}$$

where (a) uses the definition of  $\epsilon_t$ . (b) uses the summation of a Geometric series and  $t \leq T < r < \log q_U$ . (c) uses the definition of  $r$  and  $T \leq r < \log q_U - \log(12 \log q_U) - \log \log \frac{2N \log q_U}{\delta}$ .  $\square$

We now proceed to the proof of the second lower bound in Theorem 5.1.

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

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

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

Note that  $\mathcal{B}_{\tau_U(u),i}^t$  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  $\mathbb{P}[L_{u,i} = -1 | (\mathcal{B}_{\tau_U(u),i}^t)^c, \mathcal{H}_{t-1}] = \frac{1}{2}$ .*

*Proof.* According to Definition 2.1, in the user-structure only model, the matrix  $\Xi$  is deterministic and has columns consisting of all sequences in  $\{-1, +1\}^{q_U}$ . We will show that  $\mathbb{P}[L_{u,i} =$

$-1|(\mathcal{B}_{\tau_U(u),i}^t)^c, \mathcal{H}_{t-1}, \tau_U(\cdot)] = \frac{1}{2}$  where  $\tau_U(\cdot)$  is the sequence of all user types.

A priori  $\tau_I(i)$  is uniform on  $[q_I]$ . Given the sequence  $\tau_U(\cdot)$ , the matrix  $\Xi$  and the feedback  $\mathcal{H}_{t-1}$  up to time  $t-1$ , the posterior distribution of  $\tau_I(i)$  is uniform over the set of all item types  $j$  which are consistent with the outcome of recommending  $i$  to users of various types. We call this set  $\mathcal{S}_t(i)$ .

Since on the event  $(\mathcal{B}_{\tau_U(u),i}^t)^c$ , no user with the same type as  $u$  has rated  $i$  by time  $t-1$ , and since the matrix  $\Xi$  has columns consisting of all sequences in  $\{-1, +1\}^{\mathbf{q}_U}$ , half of the item types in set  $\mathcal{S}_t(i)$  are liked by user  $u$  and half of them are disliked.  $\square$

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 expected number of times a new item is recommended to a user type by time  $T$  is lower bounded as*

$$\mathbb{E} \left[ \sum_{t \in [T], u \in [N]} \mathbb{1}[(\mathcal{B}_{\tau_U(u),a_{u,t}}^t)^c] \right] \geq \mathbf{q}_U T \left[ 1 - e^{(-N/\mathbf{q}_U)} \right].$$

*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  $\widetilde{\mathbf{q}}_U$  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  $\widetilde{\mathbf{q}}_U T$ . Applying Lemma A.5 shows that  $\mathbb{E}[\widetilde{\mathbf{q}}_U] \geq \mathbf{q}_U(1 - (1 - 1/\mathbf{q}_U)^N) \geq \mathbf{q}_U[1 - e^{(-N/\mathbf{q}_U)}]$ .  $\square$

We now complete the proof of Proposition 5.5.

*Proof of Prop 5.5.* Partitioning recommendations according to  $\mathcal{B}_{\tau_U(u),a_{u,t}}^t$  gives

$$\begin{aligned} N[T - \mathbf{regret}(T)] &= \mathbb{E} \left[ \sum \mathbb{1}[L_{u,a_{u,t}} = +1] \right] \\ &= \mathbb{E} \left[ \sum \mathbb{P}[L_{u,a_{u,t}} = +1, \mathcal{B}_{\tau_U(u),a_{u,t}}^t | \mathcal{H}_{t-1}] \right] + \mathbb{E} \left[ \sum \mathbb{P}[L_{u,a_{u,t}} = +1, (\mathcal{B}_{\tau_U(u),a_{u,t}}^t)^c | \mathcal{H}_{t-1}] \right] \\ &\stackrel{\text{Claim 5.6}}{\leq} \mathbb{E} \left[ \sum \left( \mathbb{P}[\mathcal{B}_{\tau_U(u),a_{u,t}}^t | \mathcal{H}_{t-1}] + \frac{1}{2} \mathbb{P}[(\mathcal{B}_{\tau_U(u),a_{u,t}}^t)^c | \mathcal{H}_{t-1}] \right) \right] \\ &= NT - \frac{1}{2} \mathbb{E} \left[ \sum \mathbb{1}[(\mathcal{B}_{\tau_U(u),a_{u,t}}^t)^c] \right] \stackrel{\text{Claim 5.7}}{\leq} NT - \frac{T}{2} \mathbf{q}_U [1 - e^{(-N/\mathbf{q}_U)}], \end{aligned}$$

where all the summations are over  $t \in [T]$  and  $u \in [N]$ . Rearranging shows that  $\mathbf{regret}(T) \geq [1 - e^{(-N/\mathbf{q}_U)}] \frac{\mathbf{q}_U}{2N} T$  for all  $T$ .  $\square$

## 6 Item-item algorithm and analysis

This section describes a version of item-item CF with explicit exploration steps and analyzes its performance within the setup specified in Section 2. The algorithm is quite different from the user-user algorithm. This is due to the inherent asymmetry between users and items: multiple users can rate a given item simultaneously but each user can rate only one item at each time-step.

## 6.1 Algorithm

Algorithm ITEM-ITEM performs the following steps (see Algorithm 3). 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 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  $r = \lceil 2 \log(\mathbf{q}_I/\epsilon) \rceil$  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  $\mathcal{M}^2$ , and are used to represent clusters of items from  $\mathcal{M}^1$ .

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  $\mathbf{q}_I$ . Each of the  $\ell$  items chosen from  $\mathcal{M}^2$  is thought of as a *representative* of its type. For each  $j = 1, \dots, \ell$ , all items in  $\mathcal{M}^2$  that appear to be of the same type as the representative item  $i_j$  are stored in a set  $\mathcal{S}_j^2$  and then removed from  $\mathcal{M}^2$ . This guarantees that at each time  $\mathcal{M}^2$  does not contain items with the same type as any of the previously selected representative items. For each  $j = 1, \dots, \ell$ , all items in  $\mathcal{M}^1$  that appear to be of the same type as  $i_j$  are stored in a set  $\mathcal{S}_j^1$  and then removed from  $\mathcal{M}^1$ .

For each user  $u$ , we add the items in the groups  $\mathcal{S}_j^1$  whose representative  $i_j$  were liked by  $u$  to the set of exploitable items  $\mathcal{R}_u$  by  $u$ . Finally, in the exploitation phase, each user is recommended items from  $\mathcal{R}_u$ . We choose the number  $M$  of items in each of  $\mathcal{M}^1$  and  $\mathcal{M}^2$  as a function of  $\ell$  to ensure that there are enough exploitable items in  $\mathcal{R}_u$  for all users  $u$  for the entire length- $T$  time-horizon. Then,  $\ell$  is chosen to minimize regret.

The algorithm description uses the following 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.

---

**Algorithm 3** ITEM-ITEM( $T, \mathbf{q}_I, N$ ) (fixed time horizon)

---

- 1:  $\ell = \min \left\{ \left\lceil 18 \log T + \sqrt{330 \frac{\mathbf{q}_I r}{N} T} \right\rceil, \mathbf{q}_I \right\}$
  - 2:  $\epsilon = \frac{1}{2\mathbf{q}_I N}$ ;  $r = \lceil 2 \log \frac{\mathbf{q}_I}{\epsilon} \rceil$ ;  $M = \left\lceil \frac{64\mathbf{q}_I}{\ell} T \right\rceil$
  - 3:  $\mathcal{M}^1 \leftarrow M$  random items;  $\mathcal{M}^2 \leftarrow M$  new random items
  - 4:  $\mathcal{R}_u \leftarrow \emptyset$ , for each  $u \in [N]$  (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 4** ITEMEXPLORE( $\mathcal{M}^1, \mathcal{M}^2, \ell$ )

---

1:  $\mathcal{S}_j^1 \leftarrow \emptyset, \quad j \in [\mathbf{q}_I]$  (initialize sets of items of type  $j$  in  $\mathcal{M}^1$ )  
2:  $\mathcal{S}_j^2 \leftarrow \emptyset, \quad j \in [\mathbf{q}_I]$  (initialize sets of items of type  $j$  in  $\mathcal{M}^2$ )  
*Rate items by a few users (preliminary exploration)*  
3: Recommend each item in  $\mathcal{M}^1$  and  $\mathcal{M}^2$  to  $\mathbf{r}$  users from  $[N]$  (details specified in Remark 6.2).  
*Partition items and learn item types*  
4: **for**  $j = 1, \dots, \ell$  **if**  $\mathcal{M}^2 \neq \emptyset$  **do**  
5:  $i_j \leftarrow$  a random item in  $\mathcal{M}^2$  (a representative item)  
6:  $a_{u,t} \leftarrow i_j$  if  $u \notin \text{rated}(i_j)$ , otherwise a random item not from  $\mathcal{M}^2$   
7:  $\mathcal{S}_j^1 \leftarrow \{i \in \mathcal{M}^1 : L_{u,i} = L_{u,i_j} \text{ for all } u \in \text{rated}(i) \cap \text{rated}(i_j)\},$  (users agree on  $i$  vs.  $i_j$ )  
8:  $\mathcal{M}^1 \leftarrow \mathcal{M}^1 \setminus \mathcal{S}_j^1$   
9:  $\mathcal{S}_j^2 \leftarrow \{i \in \mathcal{M}^2 : L_{u,i} = L_{u,i_j} \text{ for all } u \in \text{rated}(i) \cap \text{rated}(i_j)\}$   
10:  $\mathcal{M}^2 \leftarrow \mathcal{M}^2 \setminus \mathcal{S}_j^2$   
11:  $\mathcal{R}_u = \bigcup_{j \in [\ell]: L_{u,i_j} = +1} \mathcal{S}_j^1$  for each  $u \in [N]$   
**return**  $\{\mathcal{R}_u\}_{u \in [N]}$ 

---

---

**Algorithm 5** ITEMEXPLOIT( $\{\mathcal{R}_u\}_{u \in [N]}$ )

---

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

---

**Remark 6.1.** The set of items  $\mathcal{M}^1$  and  $\mathcal{M}^2$  are updated throughout algorithm ITEMEXPLORE. In the proof, we use the notation  $\mathcal{M}_0^1$  and  $\mathcal{M}_0^2$  to refer to the set of items  $\mathcal{M}^1$  and  $\mathcal{M}^2$  at the beginning of the algorithm ITEM-ITEM.

**Remark 6.2.** Assignment of users to items for Line 3 of ITEMEXPLORE: This is done over  $\lceil (|\mathcal{M}_0^1| + |\mathcal{M}_0^2|)\mathbf{r}/N \rceil \leq 2M\mathbf{r}/N + 1$  time-steps, with additional recommendations being random new items. The main requirement in assigning users to items is to make Claim 6.3 (which bounds the probability of mis-classification of each item) hold. Since the claim addresses each item separately, we may introduce dependencies between sets of users for different items. What is important is that the set of users assigned to each specific item is uniform at random among all sets of users (or at least contains a random subset with size  $\mathbf{r}$ ). For example, one can choose a random permutation over the users, repeat this list  $\lceil (|\mathcal{M}_0^1| + |\mathcal{M}_0^2|)\mathbf{r}/N \rceil \leq 2M\mathbf{r}/N + 1$  times, and then assign each item to a block of  $\mathbf{r}$  users.

**Theorem 6.1.** Let  $\mathbf{r} = \lceil 2\log(2Nq_I^2) \rceil$ . Suppose  $q_I > 13\log N$  and  $q_U > 4\mathbf{r}$ . Then ITEM-ITEM (Algorithm 3) obtains regret per user at time  $T$  upper bounded as

$$\text{regret}(T) < \min \left\{ Y(T), \frac{T}{2} \right\}$$

where we define

$$Y(T) = 4 + \max \left\{ 52 \log T, 48 \sqrt{\frac{q_I r}{N}} T, 270 \frac{r}{N} T \right\}. \quad (13)$$

The simplified version of this theorem in Section 3 is obtained  $2 \log(Nq_I) < r < 5 \log(Nq_I)$  and  $N > 5$ .

**Remark 6.3.** *The regret bound we get for the algorithm is actually  $\text{regret}(T) \leq Y(T)$ . One can obtain  $T/2$  by a trivial algorithm which recommends random items independent of feedback to all users. This trivial algorithm improves on our bound for the parameter range in which  $T/2 < Y(T)$ . Thus in our analysis we focus on the parameter range where  $Y(T) \leq T/2$ ; one consequence is that  $\ell < T$  as chosen in Algorithm 3.*

## 6.2 Proof of Theorem 6.1

We prove Theorem 6.1, deferring several lemmas and claims to the next subsection.

The basic error event is misclassification of an item. In ITEMEXPLORE (Alg. 4),  $\mathcal{S}_j^1$  is the set of items in  $\mathcal{M}^1$  that the algorithm posits are of the same type as the  $j$ -th representative  $i_j$ . Let  $\mathcal{E}_i^1$  be the event that item  $i$  was mis-classified,

$$\mathcal{E}_i^1 = \{\exists j: i \in \mathcal{S}_j^1, \tau_I(i) \neq \tau_I(i_j)\}. \quad (14)$$

Let  $T_0$  be the number of time-steps spent making recommendations in ITEMEXPLORE. Recall that  $\mathcal{R}_u$  is the set of items to be recommended to user  $u$  in the exploit phase. We partition the recommendations made by Algorithm ITEM-ITEM to decompose the regret as follows:

$$\begin{aligned} N\text{regret}(T) &= \mathbb{E} \left[ \sum_{u=1}^N \sum_{t=1}^{T_0} \mathbb{1}[L_{u,a_{u,t}} = -1] \right] \\ &\quad + \mathbb{E} \left[ \sum_{u=1}^N \sum_{t=T_0+1}^T \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \notin \mathcal{R}_u] \right] \\ &\quad + \mathbb{E} \left[ \sum_{u=1}^N \sum_{t=T_0+1}^T \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in \mathcal{R}_u, \mathcal{E}_{a_{u,t}}^1] \right] \\ &\quad + \mathbb{E} \left[ \sum_{u=1}^N \sum_{t=T_0+1}^T \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in \mathcal{R}_u, (\mathcal{E}_{a_{u,t}}^1)^c] \right] \\ &=: \mathbf{A1} + \mathbf{A2} + \mathbf{A3} + \mathbf{A4}. \end{aligned} \quad (15)$$

The first term, **A1**, is the regret from early time-steps up to  $T_0$ . The second term, **A2**, is the regret due to not having enough items available for the exploitation phase, which is proved to be small with high probability for sufficiently large  $M$ . The third term, **A3**, 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. The fourth term, **A4**, is the regret due to exploiting the correctly classified items. It is intuitively clear and will be checked later that **A4** = 0.

**Bounding A1.** Line 3 of ITEMEXPLORE takes at most  $\lceil \frac{2Mr}{N} \rceil$  units of time to rate every item in  $\mathcal{M}_0^1$  and  $\mathcal{M}_0^2$  by  $r$  users each, since  $N$  users provide feedback at each time-step. Remark 6.2 above discusses the assignment of users to items in this phase. 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

$$T_0 \leq \left\lceil \frac{2Mr}{N} \right\rceil + \ell \leq \frac{2Mr}{N} + \ell + 1.$$

For  $t \leq T_0$ , from the perspective of any user the items recommended to it are of random type and hence

$$\mathbf{A1} = \mathbb{E} \left[ \sum_{u=1}^N \sum_{t=1}^{T_0} \mathbb{1}[L_{u,a_{u,t}} = -1] \right] = NT_0 \leq 2Mr + (\ell + 1)N. \quad (16)$$

**Bounding A2.** Time-steps  $t > T_0$  are devoted to exploitation as described in ITEMEXPLOIT. During this phase, a random item  $a_{u,t} \notin \mathcal{R}_u$  is recommended to user  $u$  only when there are no items to exploit because 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 in the time interval  $T_0 < t \leq T$  is at most  $(T - |\mathcal{R}_u|)_+$ , and

$$\begin{aligned} \mathbf{A2} &= \mathbb{E} \left[ \sum_{u \in [N]} \sum_{t=T_0+1}^T \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \notin \mathcal{R}_u] \right] \leq \mathbb{E} \left[ \sum_{u \in [N]} \sum_{t=T_0+1}^T \mathbb{1}[a_{u,t} \notin \mathcal{R}_u] \right] \\ &\leq \sum_{u \in [N]} \mathbb{E} [(T - |\mathcal{R}_u|)_+] \leq 3TN \exp(-\ell/13) + 30T, \end{aligned} \quad (17)$$

where the last inequality is from Lemma 6.4 in Section 6.3.

**Bounding A3.** Term **A3** in (15) is the expected number of mistakes made by the Algorithm ITEMEXPLOIT as a result of misclassification. Claim 6.2 below upper bounds the expected number of “potential misclassifications” (defined in Equation (19)) in the algorithm to provide an upper bound for this quantity:

$$\begin{aligned} \mathbf{A3} &= \mathbb{E} \left[ \sum_{u=1}^N \sum_{t=T_0+1}^T \mathbb{1}[L_{u,a_{u,t}} = -1, a_{u,t} \in \mathcal{R}_u, \mathcal{E}_{a_{u,t}}^1] \right] \\ &\leq \mathbb{E} \left[ \sum_{u=1}^N \sum_{t=T_0+1}^T \sum_{i \in \mathcal{N}} \mathbb{1}[a_{u,t} = i, i \in \mathcal{R}_u, \mathcal{E}_i^1] \right] \stackrel{\text{Claim 6.2}}{\leq} 2NM\epsilon \leq \frac{M}{q_I}. \end{aligned} \quad (18)$$

**Bounding A4.** By definition of the mis-classification event,  $\mathcal{E}_i^1$ , given in Equation (14), if an item  $i \in \mathcal{S}_j^1$  is correctly classified, then  $\tau_I(i) = \tau_I(i_j)$ . Since  $L_{u,i} = \xi_{\tau_U(u), \tau_I(i)}$ , all users rate  $i$  the same

as  $i_j$ . By construction of the sets  $\mathcal{R}_u$  in Line 11 of ITEMEXPLOIT, for any item  $i \in \mathcal{R}_u$ , there is some  $j \in [\mathbf{q}_I]$  such that  $i \in \mathcal{S}_j^1$  and  $u$  likes item  $i_j$ . Hence,

$$\begin{aligned}\mathbb{P}[L_{u,i} = -1 | i \in \mathcal{R}_u, (\mathcal{E}_i^1)^c] &= \mathbb{P}[L_{u,i} = -1 | \exists j : L_{u,i_j} = +1, i \in \mathcal{S}_j^1, \tau_I(i) = \tau_I(i_j)] \\ &= \mathbb{P}[L_{u,i} = -1 | \xi_{\tau_U(u), \tau_I(i)} = +1] = 0.\end{aligned}$$

It follows that  $\mathbf{A4} = 0$ .

**Combining all the bounds.** Plugging in Equations (16), (17), (18) and  $\mathbf{A4} = 0$  into Equation (15) gives

$$N\text{regret}(T) \leq 2Mr + (\ell + 1)N + 3TN \exp(-\ell/13) + 30T + \frac{M}{\mathbf{q}_I}.$$

Setting  $M = \frac{64T\mathbf{q}_I}{\ell}$  gives

$$\text{regret}(T) \stackrel{(a)}{\leq} 2 + 134\frac{T\mathbf{q}_I r}{\ell N} + \ell + 3T \exp\left(-\frac{\ell}{13}\right) \stackrel{(b)}{\leq} \begin{cases} 4 + 7\log T + 24\sqrt{\frac{\mathbf{q}_I r}{N}T} & \text{if } \ell < \mathbf{q}_I \\ \mathbf{q}_I + 1 + 135\frac{r}{N}T & \text{if } \ell = \mathbf{q}_I \end{cases}$$

where (a) holds for  $\ell \geq 13$  (imposed by the algorithm for  $T \geq 2$ ) which gives  $M/\mathbf{q}_I \leq 5T$  and  $r > 6$  (which holds if  $\mathbf{q}_I, N \geq 3$ ). If  $\ell = \mathbf{q}_I$ , since  $\mathbf{q}_I > 13\log N$ , we have  $3T \exp(-\ell/13) \leq 3T/N \leq rT/N$ . (b) is obtained by choosing the parameter  $\ell$  to be

$$\ell = \min \left\{ \left\lceil 13\log T + \sqrt{134\frac{\mathbf{q}_I r}{N}T} \right\rceil, \mathbf{q}_I \right\}.$$

Next, we will show that  $\text{regret}(T) \leq Y(T)$  for  $Y(T)$  defined in (13). This is clearly true for  $\ell < \mathbf{q}_I$ .

To show  $\text{regret}(T) \leq Y(T)$  for  $\ell = \mathbf{q}_I$ , we consider two cases: If  $\mathbf{q}_I \leq 135Tr/N$ , then  $\text{regret}(T) \leq 270Tr/N + 1$ . If  $135Tr/N < \mathbf{q}_I$ , we know that  $\mathbf{q}_I \leq \left\lceil 13\log T + \sqrt{134\frac{\mathbf{q}_I r}{N}T} \right\rceil$  (which is true since  $\ell = \mathbf{q}_I$ ). This gives  $\text{regret}(T) \leq 2 \left\lceil 13\log T + \sqrt{134\frac{\mathbf{q}_I r}{N}T} \right\rceil \leq Y(T)$ .

### 6.3 Lemmas used in the proof

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

**Claim 6.2.** *Suppose that  $\mathbf{q}_U > 4r$ . The expected total number of times a misclassified item is recommended in the exploitation step is upper bounded as*

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

*Proof.* Event  $\mathcal{E}_i^1$ , defined in (14) to be the misclassification of item  $i \in \mathcal{M}^1$ , occurs if the preferences of random users classifying item  $i$  is the same as 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 in Line 3 of ITEMEXPLORE. Instead of directly analyzing  $\mathcal{E}_i^1$ , we define an event called “potential error event”  $\bar{\mathcal{E}}_{i,U_i}$ , which will be shown to satisfy  $\mathcal{E}_i^1 \subseteq \bar{\mathcal{E}}_{i,U_i}$ ; an upper bound for  $\mathbb{P}[\bar{\mathcal{E}}_{i,U_i}]$  is given in Claim 6.3.

For item  $i$  and subset of users  $U \subseteq [N]$ , define

$$\bar{\mathcal{E}}_{i,U} = \{\exists j \neq \tau_I(i) : L_{u,i} = \xi_{\tau_U(u),j}, \text{ for all } u \in U\} \quad (19)$$

to be the event that the ratings of users in  $U$  for item  $i$  agree with some other item type. For item  $i \in \mathcal{S}_j^1$ , if  $t$  is the time  $i$  is added to  $\mathcal{S}_j^1$  in the exploration phase, let  $U_i = \text{rated}_t(i) \cap \text{rated}_t(i_j)$  be the set of witness users whose ratings were used to conclude that  $i$  and  $i_j$  are of the same type. Item  $i$  is added to  $\mathcal{S}_j^1$  only if all users in  $U_i$  agree on  $i$  vs.  $i_j$ , so misclassification  $\mathcal{E}_i^1$  (defined in Equation (14)) implies  $\bar{\mathcal{E}}_{i,U_i}$ . We can now deduce the inequalities, justified below:

$$\begin{aligned} \mathbb{E} \left[ \sum_{u=1}^N \sum_{t=T_0+1}^T \sum_{i \in \mathcal{N}} \mathbb{1}[a_{u,t} = i, i \in \mathcal{R}_u, \mathcal{E}_i^1] \right] &\stackrel{(a)}{\leq} \mathbb{E} \left[ \sum_{u=1}^N \sum_{t=T_0+1}^T \sum_{i \in \mathcal{M}_0^1} \mathbb{1}[a_{u,t} = i, i \in \mathcal{R}_u, \bar{\mathcal{E}}_{i,U_i}] \right] \\ &\stackrel{(b)}{\leq} N \mathbb{E} \left[ \sum_{i \in \mathcal{M}^1} \mathbb{1}[\bar{\mathcal{E}}_{i,U_i}] \right] \stackrel{(c)}{\leq} 2NM\epsilon. \end{aligned}$$

Inequality (a) holds because according to Line 11 of ITEMEXPLORE for every user  $u$ , the set  $\mathcal{R}_u$  is a subset of  $\mathcal{M}_0^1$ , and also the containment  $\mathcal{E}_i \subseteq \bar{\mathcal{E}}_{i,U_i}$ ; (b) follows since each item  $i$  is recommended at most  $N$  times; (c) uses  $|\mathcal{M}_0^1| = M$  together with Claim 6.3 which shows that  $\mathbb{P}[\bar{\mathcal{E}}_{i,U_i}] \leq 2\epsilon$ .  $\square$

**Claim 6.3.** Suppose that  $q_U > 4r$ . Consider the “potential error” event  $\bar{\mathcal{E}}_{i,U_i}$  defined in (19) with the set of users  $U_i$  defined immediately after. Then  $\mathbb{P}[\bar{\mathcal{E}}_{i,U_i}] \leq 2\epsilon$  for all  $i \in \mathcal{M}_0^1$ .

*Proof.* Each representative item  $i_j$  chosen in ITEMEXPLORE is rated by all of the users. Other items in  $\mathcal{M}_0^1$  and  $\mathcal{M}_0^2$  are rated by at least  $r$  users in the exploration phase. Remark 6.2 describes how Line 3 in ITEMEXPLORE is performed to guarantee that the set of  $r$  users assigned to each specific item is uniformly at random among all subset of users of size  $r$ . By Lemma A.3, if  $q_U > 4r$ , then with probability at least  $1 - \exp(-r/2) \geq 1 - \epsilon/q_I$  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  $\tau_I(i)$  also have the same ratings for type  $j \neq \tau_I(i)$  with probability at most  $2^{-r/2}$ . (Any two item types  $j \neq j'$  have jointly independent columns in the preference matrix.) The choice  $r \geq 2 \log(q_I/\epsilon)$  and a union bound over item types  $j$  completes the proof.  $\square$

**Lemma 6.4.** For user  $u \in [N]$ , let  $\mathcal{R}_u$  be defined as in Line 10 of algorithm ITEMEXPLORE. Then

$$\mathbb{P}[|\mathcal{R}_u| \leq T] \leq 3e^{-\ell/18} + \frac{30}{N},$$

and

$$\mathbb{E}[(T - |\mathcal{R}_u|)_+] \leq 3Te^{-\ell/18} + \frac{30}{N}T.$$

*Proof.* We begin with a sketch. Any user  $u$  likes roughly half of the  $\ell$  item types that have representatives  $\{i_j\}_{j=1}^\ell$ . The total number of items in  $\mathcal{M}_0^1$  is  $\frac{64q_u}{\ell}T$ , so there are about  $\frac{64}{\ell}T$  items from each of the item types in  $\mathcal{M}_0^1$ . Adding over the  $\ell$  types, the set  $\mathcal{R}_u$  of items in  $\mathcal{M}_0^1$  that user  $u$  likes will typically have size at least  $T$ .

Making this argument rigorous requires some care for the following reasons: 1)  $\mathcal{R}_u$  is the union of the  $\mathcal{S}_j^1$ 's with  $L_{u,i_j} = +1$ , but  $\mathcal{S}_j^1$  can be missing items due to misclassification (there may be items of the same type as  $i_j$  that have been classified as being of the same type as some  $i_{j'}$  for which  $L_{u,i_{j'}} = -1$ ). 2) The distribution of the type of each representative item depends on the number of remaining items of each type in  $\mathcal{M}^2$  when the representative is chosen. Again, due to misclassification, this can be different from the actual number of items of each type initially present in  $\mathcal{M}^2$ . 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, bounding the size of  $\mathcal{R}_u$  by introducing a different set. For  $u \in [N]$ , let

$$\tilde{\mathcal{R}}_u^1 = \{i \in \mathcal{M}_0^1 : \tau_I(i) = \tau_I(i_j), \text{ for some } j \in [\ell] \text{ such that } L_{u,i_j} = +1\} \quad (20)$$

be the items in  $\mathcal{M}_0^1$  whose types are the same as one of the representatives  $i_j$  that are liked by  $u$ . Note that if an item  $i \in \tilde{\mathcal{R}}_u^1$  is correctly classified by the algorithm (i.e.  $(\mathcal{E}_i^1)^c$  occurs, where  $\mathcal{E}_i^1$  is defined in Equation (14)), then  $i \in \mathcal{R}_u$ . Hence,  $i \in \tilde{\mathcal{R}}_u^1 \setminus \mathcal{R}_u$  implies  $\mathcal{E}_i^1$  and since  $\mathcal{E}_i^1 \subseteq \bar{\mathcal{E}}_{i,\mathcal{U}_i}$  we have

$$|\mathcal{R}_u| \geq |\tilde{\mathcal{R}}_u^1| - \sum_{i \in \mathcal{M}_0^1} \mathbb{1}[\mathcal{E}_i^1] \geq |\tilde{\mathcal{R}}_u^1| - \sum_{i \in \mathcal{M}_0^1} \mathbb{1}[\bar{\mathcal{E}}_{i,\mathcal{U}_i}]. \quad (21)$$

It follows that

$$\mathbb{P}[|\mathcal{R}_u| \leq T] \leq \mathbb{P}\left[|\tilde{\mathcal{R}}_u^1| - \sum_{i \in \mathcal{M}_0^1} \mathbb{1}[\bar{\mathcal{E}}_{i,\mathcal{U}_i}] \leq T\right] \leq \mathbb{P}\left[|\tilde{\mathcal{R}}_u^1| < \frac{3T}{2}\right] + \mathbb{P}\left[\sum_{i \in \mathcal{M}_0^1} \mathbb{1}[\bar{\mathcal{E}}_{i,\mathcal{U}_i}] \geq \frac{T}{2}\right]. \quad (22)$$

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

$$\mathbb{E}\left[\sum_{i \in \mathcal{M}_0^1} \mathbb{1}[\bar{\mathcal{E}}_{i,\mathcal{U}_i}]\right] \leq 2|\mathcal{M}_0^1|\epsilon = \frac{64T}{N\ell}, \quad (23)$$

and by Markov's Inequality<sup>4</sup>

$$\mathbb{P}\left[\sum_{i \in \mathcal{M}_0^1} \mathbb{1}[\bar{\mathcal{E}}_{i,\mathcal{U}_i}] \geq \frac{T}{2}\right] \leq \frac{128}{\ell N}. \quad (24)$$

---

<sup>4</sup>Note that application of Markov's Inequality does not require the events  $\bar{\mathcal{E}}_{i,\mathcal{U}_i}$  to be jointly independent for various values of  $i \in \mathcal{M}_0^1$ . This is discussed in Remark 6.2.

We now bound the first term on the right-hand side of (22). To this end, let

$$\tilde{\mathcal{L}}_u^{(\ell)} = \{\tau_I(i_j) : j \in [\ell], L_{u,i_j} = +1\} \quad (25)$$

be the type of item representatives that are liked by user  $u$ . By definition of  $\tilde{\mathcal{R}}_u^1$  in (20),

$$|\tilde{\mathcal{R}}_u^1| = \sum_{i \in \mathcal{M}_0^1} \mathbb{1}[\tau_I(i) \in \tilde{\mathcal{L}}_u^{(\ell)}].$$

Now, we claim that the set  $\tilde{\mathcal{L}}_u^{(\ell)}$  is independent of all types and preferences for items in  $\mathcal{M}_0^1$ . To see this, note that  $\tilde{\mathcal{L}}_u^{(\ell)}$  is determined by row  $u$  of the type matrix  $\Xi$  and the items in  $\mathcal{M}_0^2$ , their types, and randomness in the algorithm, which determines the choice of  $i_j$  and  $\tau_I(i_j)$ ; these together determine  $L_{u,i_j} = \xi_{\tau_U(u), \tau_I(i_j)}$  and  $\tau_I(i_j)$ . So, conditioning on  $\tilde{\mathcal{L}}_u^{(\ell)}$  having cardinality  $\tilde{\ell}_u$ ,  $|\tilde{\mathcal{R}}_u^1|$  is the sum of  $|\mathcal{M}_0^1| = 64 \frac{q_I}{\ell} T$  i.i.d. Bernoulli variables with parameter  $\frac{\tilde{\ell}_u}{q_I}$  and hence  $|\tilde{\mathcal{R}}_u^1| \sim \text{Binom}(64 \frac{q_I}{\ell} T, \frac{\tilde{\ell}_u}{q_I})$ . Conditioning on  $\tilde{\ell}_u \geq \frac{\ell}{20}$ , by a Chernoff bound (Lemma A.1) and stochastic domination of binomials with increasing number of trials we obtain

$$\mathbb{P} \left[ |\tilde{\mathcal{R}}_u^1| < \frac{3T}{2} \mid \tilde{\ell}_u \geq \frac{\ell}{20} \right] \leq \exp \left( -\frac{T}{8} \right).$$

In Lemma 6.5 below, we will lower bound the probability that  $\tilde{\ell}_u \geq \ell/20$ . Combining the last displayed inequality with Lemma 6.5, we get

$$\mathbb{P} \left[ |\tilde{\mathcal{R}}_u^1| < \frac{3T}{2} \right] \leq \exp \left( -\frac{T}{8} \right) + \exp \left( -\frac{\ell}{13} \right) + \exp \left( -\frac{q_I}{8} \right) + \frac{20}{N}.$$

Plugging this and (24) into (22) gives

$$\mathbb{P} \left[ |\mathcal{R}_u| < T \right] \leq \exp \left( -\frac{T}{8} \right) + \exp \left( -\frac{\ell}{13} \right) + \exp \left( -\frac{q_I}{8} \right) + \frac{20}{N} + \frac{128}{\ell N},$$

and since  $\ell \leq T, q_I$ <sup>5</sup> and  $\ell \geq 13$  (for  $T \geq 13$ )

$$\mathbb{P} \left[ |\mathcal{R}_u| < T \right] \leq 3 \exp(-\ell/13) + \frac{30}{N}.$$

The bound on  $\mathbb{E}[(T - |\mathcal{R}_u|)_+]$  is an immediate consequence. □

**Lemma 6.5.** *Fix  $u \in [N]$  and suppose that  $q_I > 8r$ . With  $\tilde{\mathcal{L}}_u^{(\ell)}$  defined in (25) and  $\tilde{\ell}_u = |\tilde{\mathcal{L}}_u^{(\ell)}|$ ,*

$$\mathbb{P} \left[ \tilde{\ell}_u < \ell/20 \right] \leq \exp(-\ell/13) + \exp(-q_I/8) + 20/N.$$

*Proof.* For a given user  $u$ ,  $\tilde{\ell}_u$  is the number of item type of representatives liked by user  $u$ . Let  $\mathcal{L}_u$

---

<sup>5</sup>As discussed in Remark 6.3, the current analysis of regret assumes that  $Y(T) < \frac{T}{2}$ . It can be shown that this assumption implies  $\ell \leq T$ . The choice of  $\ell$  in the algorithm clearly guarantees  $\ell \leq q_I$ .

be the item types in  $[\mathbf{q}_I]$  that are liked by user  $u$ :

$$\mathcal{L}_u = \{j \in [\mathbf{q}_I] : \xi_{\tau_U(u),j} = +1\}. \quad (26)$$

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

$$\mathbb{P}[|\mathcal{L}_u| < \mathbf{q}_I/4] < \exp(-\mathbf{q}_I/8). \quad (27)$$

We will show that

$$\mathbb{P}[\tilde{\ell}_u \leq \ell/20 \mid |\mathcal{L}_u| \geq \mathbf{q}_I/4] \leq \exp(-\ell/13) + 20/N, \quad (28)$$

which will prove the lemma by combining with (27).

We now work towards defining a certain error event in Equation (31) below. Let the sequence of random variables  $X_1 = \tau_I(\mathbf{i}_1), X_2 = \tau_I(\mathbf{i}_2), \dots, X_\ell = \tau_I(\mathbf{i}_\ell)$  denote the types of the item representatives chosen by the algorithm, so that  $\tilde{\ell}_u = |\tilde{\mathcal{L}}_u^{(\ell)}| = \sum_{j \in [\ell]} \mathbb{1}[X_j \in \mathcal{L}_u]$  (with  $\tilde{\mathcal{L}}_u^{(\ell)}$  defined in (25)). Let  $\bar{\mathcal{R}}^2(j)$  be the set of items in  $\mathcal{M}_0^2$  of type  $j$

$$\bar{\mathcal{R}}^2(j) = \{i \in \mathcal{M}_0^2 : \tau_I(i) = j\}. \quad (29)$$

Later we will use the notation  $\bar{\mathcal{R}}^2(\cdot) = \{\bar{\mathcal{R}}^2(j)\}_{j=1}^{\mathbf{q}_I}$  for the collection of these sets. Now let the event  $\mathcal{E}_i^2$  denote misclassification of an item  $i \in \mathcal{M}_0^2$  (similar to event  $\mathcal{E}_i^1$  for  $i \in \mathcal{M}_0^1$  in (14)),

$$\mathcal{E}_i^2 = \{\exists j : i \in \mathcal{S}_j^2, \tau_I(i) \neq \tau_I(\mathbf{i}_j)\}. \quad (30)$$

Let **Err** be the event that for some item type  $j$ , more than a fraction 1/10 of the items in  $\mathcal{M}_0^2$  of type  $j$  are misclassified:

$$\mathbf{Err} = \left\{ \sum_{i \in \bar{\mathcal{R}}^2(j)} \mathbb{1}[\mathcal{E}_i^2] > \frac{|\bar{\mathcal{R}}^2(j)|}{10}, \text{ for some } j \in [\mathbf{q}_I] \right\}. \quad (31)$$

Conditioning on **Err** in the left-hand side of (28) gives

$$\begin{aligned} & \mathbb{P} \left[ \sum_{i=1}^{\ell} \mathbb{1}[X_i \in \tilde{\mathcal{L}}_u^{(\ell)}] < \ell/20 \mid |\mathcal{L}_u| \geq \mathbf{q}_I/4 \right] \\ & \leq \mathbb{P} \left[ \sum_{i=1}^{\ell} \mathbb{1}[X_i \in \tilde{\mathcal{L}}_u^{(\ell)}] < \ell/20 \mid \mathbf{Err}^c, |\mathcal{L}_u| \geq \mathbf{q}_I/4 \right] + \mathbb{P}[\mathbf{Err} \mid |\mathcal{L}_u| \geq \mathbf{q}_I/4]. \end{aligned} \quad (32)$$

**Bound on second term of (32)** We will use Markov's inequality to bound

$$\mathbb{P}[\mathbf{Err} \mid |\mathcal{L}_u| \geq \mathbf{q}_I/4] = \mathbb{E} \left[ \mathbb{P}[\mathbf{Err} \mid \bar{\mathcal{R}}^2(\cdot), \mathcal{L}_u] \mid |\mathcal{L}_u| \geq \mathbf{q}_I/4 \right].$$

We need to consider the effect of matrix  $\Xi$  (and in particular its  $\tau_U(u)$ th row, which determines  $\mathcal{L}_u$ ) on the probability of error in categorizing items. Notably, if users rate two item types similarly,



the probability of misclassifying items of these types is higher. As a result,  $\mathcal{L}_u$  contains some information about discriminability of distinct item types and we need to control this dependence.

Claim 6.2 shows that for  $i \in \mathcal{M}_0^1$ , the probability of  $\mathcal{E}_i^1$  (the event that item  $i$  is miscategorized) is  $\mathbb{P}[\mathcal{E}_i^1] \leq 2\epsilon$ . The same proof gives  $\mathbb{P}[\mathcal{E}_i^2] \leq 2\epsilon$  (with  $\mathcal{E}_i^2$  defined in (30) for  $i \in \mathcal{M}_0^2$ ). To show that, let the potential error event  $\bar{\mathcal{E}}_{i,U_i}$ , be the event that there exists an item type  $j \neq \tau_I(i)$  such that for all the users  $u$  in  $U_i$ , we have  $L_{u,i} = \xi_{\tau_U(u),j}$ . For  $i \in \mathcal{S}_j^2$ , let  $U_i = \text{rated}(i) \cap \text{rated}(i_j)$  at the time the algorithm added  $i$  to  $\mathcal{S}_j^2$ . Then the containment  $\mathcal{E}_i^2 \subseteq \bar{\mathcal{E}}_{i,U_i}$  holds. We will later use the inequality  $\mathbb{P}[\mathcal{E}_i^2 | \tau_I(i), \bar{\mathcal{R}}^2(\cdot), \mathcal{L}_u] \leq \mathbb{P}[\bar{\mathcal{E}}_{i,U_i} | \tau_I(i), \bar{\mathcal{R}}^2(\cdot), \mathcal{L}_u]$  and focus on the set of users  $\mathcal{U}_i \setminus \{u\}$ , using  $\bar{\mathcal{E}}_{i,\mathcal{U}_i} \subseteq \bar{\mathcal{E}}_{i,\mathcal{U}_i \setminus \{u\}}$ .

As specified in Algorithm ITEMEXPLORE, for each  $i \in \mathcal{M}_0^1 \cup \mathcal{M}_0^2$  the set  $\mathcal{U}_i$  of at least  $r$  users, chosen independently of feedback, rate item  $i$ . By Lemma A.3, if  $q_I > 4r$ , then there are users of at least  $r/2$  distinct user types in  $\mathcal{U}_i$  with probability at least  $1 - \exp(-r/2) \geq 1 - \epsilon/q_I$ . Conditional on  $\mathcal{U}_i$  having at least  $r/2$  distinct user types, there are at least  $r/2 - 1$  users of distinct types in  $\mathcal{U}_i \setminus \{u\}$ —also distinct from type of  $u$ —whose preferences for item type  $\tau_I(i)$  are independent of  $\mathcal{L}_u$  and  $\bar{\mathcal{R}}^2(\cdot)$ .

Conditional on  $\tau_I(i), \bar{\mathcal{R}}^2(\cdot), \mathcal{L}_u$ , any two item types  $j \neq j'$  have jointly independent user preferences by  $r/2 - 1$  users with distinct types in  $\mathcal{U}_i \setminus \{u\}$ ; they are rated in the same way by these users with probability at most  $2^{-(r/2-1)}$ . A union bound over item types  $j' \neq \tau_I(i)$  gives  $\mathbb{P}[\bar{\mathcal{E}}_{i,\mathcal{U}_i \setminus \{u\}} | \tau_I(i), \bar{\mathcal{R}}^2(\cdot), \mathcal{L}_u] \leq q_I 2^{-(r/2-1)}$ . The choice  $r > 2 \log(q_I/\epsilon)$  and  $\epsilon = \frac{1}{2q_I N}$  and the containment  $\mathcal{E}_i^2 \subseteq \bar{\mathcal{E}}_{i,\mathcal{U}_i} \subseteq \bar{\mathcal{E}}_{i,\mathcal{U}_i \setminus \{u\}}$  gives  $\mathbb{P}[\mathcal{E}_i^2 | \tau_I(i), \bar{\mathcal{R}}^2(\cdot), \mathcal{L}_u] \leq \frac{1}{q_I N}$  for all  $i \in \mathcal{M}_0^2$ . Knowing  $\bar{\mathcal{R}}^2(\cdot)$  determines  $\tau_I(i)$  for  $i \in \mathcal{M}_0^2$ . Hence,  $\mathbb{P}[\mathcal{E}_i^2 | \bar{\mathcal{R}}^2(\cdot), \mathcal{L}_u] \leq \frac{1}{q_I N}$  and Markov's inequality gives

$$\mathbb{P} \left[ \sum_{i \in \bar{\mathcal{R}}^2(j)} \mathbb{1}[\mathcal{E}_i^2] > \frac{|\bar{\mathcal{R}}^2(j)|}{10} \mid \bar{\mathcal{R}}^2(\cdot), \mathcal{L}_u \right] \leq \frac{20}{q_I N}.$$

Union bounding over  $j \in [q_I]$  and tower property gives the desired bound.

**Bound on first term of (32).** We start by showing that conditioned on the event  $\text{Err}^c$  conditional on  $\mathcal{L}_u$  and variables  $X_1, \dots, X_{m-1}$  and  $\bar{\mathcal{R}}^2(X_1), \dots, \bar{\mathcal{R}}^2(X_{m-1})$ , the type of the  $m$ -th representative item,  $X_m$ , is almost uniform over all the item types not learned yet,  $[q_I] \setminus \{X_1, \dots, X_{m-1}\}$ . Concretely, we will find upper and lower bounds on

$$\mathbb{P}[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \{\bar{\mathcal{R}}^2(X_n)\}_{n=1}^{m-1}, \mathcal{L}_u, \text{Err}^c] \quad (33)$$

for any  $j \in [q_I] \setminus \{X_1, \dots, X_{m-1}\}$ . Later, we will focus on  $\mathcal{L}_u$  such that  $|\mathcal{L}_u| \geq q_I/4$ .

*Lower bound for (33).* Let  $t$  be the time the  $m$ -th item type representative is chosen by the algorithm. For  $j \in [q_I] \setminus \{X_n\}_{n=1}^{m-1}$  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  $\mathcal{M}^2$  at time  $t-1$ . The number of items of type  $j$  in  $\mathcal{M}^2$  at time  $t-1$  is at least  $|\bar{\mathcal{R}}^2(j)| - \sum_{i \in \bar{\mathcal{R}}^2(j)} \mathbb{1}[\mathcal{E}_i^2]$ .

The number of items removed from  $\mathcal{M}^2$  by time  $t$  is at least  $\sum_{l=1}^{m-1} |\bar{\mathcal{R}}(X_l)|$  (there could be

additional items with types not in  $\{X_n\}_{n=1}^{m-1}$  that were removed from  $\mathcal{M}^2$  due to misclassification). Hence, the total number of remaining items in  $\mathcal{M}^2$  by time  $t$  is at most  $M - \sum_{n=1}^{m-1} |\bar{\mathcal{R}}^2(X_n)|$ . Let  $Z_{\mathcal{E}^2}$  be the collection of indicator variables  $Z_{\mathcal{E}^2} = \{\mathbb{1}[\mathcal{E}_i^2]\}_{i \in \mathcal{M}_0^2}$ . For any  $j \in [\mathbf{q}_I] \setminus \{X_n\}_{n=1}^{m-1}$ :

$$\mathbb{P}\left[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \{\bar{\mathcal{R}}^2(j')\}_{j'=1}^{\mathbf{q}_I}, \mathcal{L}_u, Z_{\mathcal{E}^2}\right] \geq \frac{|\bar{\mathcal{R}}^2(j)| - \sum_{i \in \bar{\mathcal{R}}^2(j)} \mathbb{1}[\mathcal{E}_i^2]}{M - \sum_{n=1}^{m-1} |\bar{\mathcal{R}}^2(X_n)|}.$$

On the event  $\text{Err}^c$  (defined in (31)),  $\sum_{i \in \bar{\mathcal{R}}^2(j)} \mathbb{1}[\mathcal{E}_i^2] \leq |\bar{\mathcal{R}}^2(j)|/10$  for any  $j \in [\mathbf{q}_I] \setminus \{X_n\}_{n=1}^{m-1}$ . Therefore,

$$\mathbb{P}\left[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \{\bar{\mathcal{R}}^2(j')\}_{j'=1}^{\mathbf{q}_I}, \mathcal{L}_u, \text{Err}^c\right] \geq \frac{9}{10} \frac{|\bar{\mathcal{R}}^2(j)|}{M - \sum_{n=1}^{m-1} |\bar{\mathcal{R}}^2(X_n)|}. \quad (34)$$

*Upper bound for (33).* This time we lower bound the number of items remaining in  $\mathcal{M}^2$  by the number of items in types  $[\mathbf{q}_I] \setminus \{X_n\}_{n=1}^{m-1}$  minus the number of possible mistakes which removed some items from them. This gives

$$\mathbb{P}\left[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \{\bar{\mathcal{R}}^2(j')\}_{j'=1}^{\mathbf{q}_I}, \mathcal{L}_u, Z_{\mathcal{E}^2}\right] \leq \frac{|\bar{\mathcal{R}}^2(j)|}{M - \sum_{n=1}^{m-1} |\bar{\mathcal{R}}^2(X_n)| - \sum_{i \in \mathcal{M}_0^2 \setminus \bigcup_{n=1}^{m-1} \bar{\mathcal{R}}^2(X_n)} \mathbb{1}[\mathcal{E}_i^2]}.$$

By definition of  $\text{Err}$  (in (31)), we have

$$\mathbb{P}\left[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \{\bar{\mathcal{R}}^2(j')\}_{j'=1}^{\mathbf{q}_I}, \mathcal{L}_u, \text{Err}^c\right] \leq \frac{10}{9} \frac{|\bar{\mathcal{R}}^2(j)|}{M - \sum_{n=1}^{m-1} |\bar{\mathcal{R}}^2(X_n)|}. \quad (35)$$

*Combining lower and upper bounds.* The expected value of  $|\bar{\mathcal{R}}^2(j)|$  conditional on variables  $\{X_n\}_{n=1}^{m-1}$ ,  $\{\bar{\mathcal{R}}^2(X_n)\}_{n=1}^{m-1}$ , and  $\mathcal{L}_u$  is invariant to choice of  $j \in [\mathbf{q}_I] \setminus \{X_n\}_{n=1}^{m-1}$ . So, the conditional expectation

$$C = \mathbb{E}\left[\frac{|\bar{\mathcal{R}}^2(j)|}{M - \sum_{n=1}^{m-1} |\bar{\mathcal{R}}^2(X_n)|} \mid \{X_n\}_{n=1}^{m-1}, \{\bar{\mathcal{R}}^2(X_n)\}_{n=1}^{m-1}, \mathcal{L}_u, \text{Err}^c\right],$$

is independent of  $j$  for  $j \in [\mathbf{q}_I] \setminus \{X_n\}_{n=1}^{m-1}$ . Note that in the above display, the conditional expectation is with respect to  $\bar{\mathcal{R}}^2(X_n)$  for  $n = 1, \dots, m-1$ . Hence, using tower property of expectation on (34) and (35) to remove the conditioning on  $\{\bar{\mathcal{R}}^2(j)\}_{j \notin \{X_n\}_{n=1}^{m-1}}$  along with the definition of  $C$  above gives

$$\frac{9}{10}C \leq \mathbb{P}\left[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \{\bar{\mathcal{R}}^2(X_n)\}_{n=1}^{m-1}, \mathcal{L}_u, \text{Err}^c\right] \leq \frac{10}{9}C.$$

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

gives for all  $j \in [\mathbf{q}_I] \setminus \{X_n\}_{n=1}^{m-1}$ ,

$$\mathbb{P}\left[X_m = j \mid \{X_n\}_{n=1}^{m-1}, \{\bar{\mathcal{R}}^2(X_n)\}_{n=1}^{m-1}, \mathcal{L}_u, \text{Err}^c\right] > \frac{2}{5} \frac{1}{\mathbf{q}_I - m + 1}.$$

Conditional on  $\mathcal{L}_u$  such that  $|\mathcal{L}_u| \geq \mathbf{q}_I/4$ , for  $s \leq m \leq \ell \leq \mathbf{q}_I$  and  $s \leq \ell/20$  we have

$$\mathbb{P}\left[X_m \in \mathcal{L}_u \mid \{X_n\}_{n=1}^{m-1}, \sum_{n=1}^{m-1} \mathbb{1}[X_n \in \mathcal{L}_u] < s, \mathcal{L}_u, |\mathcal{L}_u| \geq \frac{\mathbf{q}_I}{4}, \text{Err}^c\right] \geq \frac{2}{5} \frac{\frac{\mathbf{q}_I}{4} - \min\{m, \ell/20\}}{\mathbf{q}_I - m + 1} \geq \frac{1}{12}.$$

So, on the event  $\sum_{n=1}^{\ell} \mathbb{1}[X_n \in \mathcal{L}_u] \leq \frac{\ell}{20}$ , the random variable  $\sum_{n=1}^{\ell} \mathbb{1}[X_n \in \mathcal{L}_u]$  conditional on  $\mathcal{L}_u$  and events  $|\mathcal{L}_u| \geq \mathbf{q}_I/4$  and  $\text{Err}^c$ , stochastically dominates a Binomial random variable with mean  $\ell/12$ . Hence, by a Chernoff bound,

$$\mathbb{P}\left[\sum_{n=1}^{\ell} \mathbb{1}[X_n \in \mathcal{L}_u] \leq \frac{\ell}{20} \mid \mathcal{L}_u, |\mathcal{L}_u| \geq \frac{\mathbf{q}_I}{4}, \text{Err}^c\right] \leq \exp(-\ell/13). \quad \square$$

## 7 Item structure only: lower bound

In this section we prove a lower bound on the regret of any online recommendation system in the regime with item structure only where  $\mathbf{q}_U = N$  as described in Definition 2.2. Throughout this section, we will assume  $N > 32$ .

**Theorem 7.1.** *Let  $r = \lfloor 8 \log \mathbf{q}_I - 4 \log \log N \rfloor$  and  $\eta = 1/\log N$ . In the item structure model with  $N$  users and  $\mathbf{q}_I$  item types, any recommendation algorithm must incur regret*

$$\text{regret}(T) \geq \frac{1 - 3\eta}{2} \max\left\{1, \frac{1}{2} Z(T), \frac{r}{N} T\right\} - \frac{T}{N},$$

where we define

$$Z(T) = \begin{cases} \frac{T}{2}, & \text{if } T < \frac{2\sqrt{\mathbf{q}_I}}{3N} \\ \frac{T}{5 \log \mathbf{q}_I} [\log(8 \mathbf{q}_I \log \mathbf{q}_I) - \log(NT)], & \text{if } \frac{2\sqrt{\mathbf{q}_I}}{3N} \leq T < \frac{4 \mathbf{q}_I \log \mathbf{q}_I}{N} \\ \frac{T}{8 \log \mathbf{q}_I}, & \text{if } \frac{4 \mathbf{q}_I \log \mathbf{q}_I}{N} \leq T < \frac{16 \mathbf{q}_I (\log \mathbf{q}_I)^2}{N} \\ \frac{1}{2} \sqrt{\frac{T \mathbf{q}_I}{N}}, & \text{if } \frac{16 \mathbf{q}_I (\log \mathbf{q}_I)^2}{N} \leq T. \end{cases} \quad (36)$$

The assumption  $N > 32$  implies  $\eta < 1/5$ . Note that the function  $Z(T)$  is continuous up to a multiplicative constant factor<sup>6</sup>.

---

<sup>6</sup>We define a function  $Z(T)$  to be continuous up to a multiplicative constant factor  $C$  if for every  $T_0$ ,

$$C^{-1} \lim_{T \rightarrow T_0^-} Z(T) < \lim_{T \rightarrow T_0^+} Z(T) < C \lim_{T \rightarrow T_0^-} Z(T).$$

To get the simplified version in Section 3, we bounded the value of  $Z(T)$  in the second (in which  $\frac{2\sqrt{q_I}}{3N} \leq T < \frac{4q_I \log q_I}{N}$ ) by  $\frac{T}{8 \log q_I}$ . Also, the assumption  $q_I > 25(\log N)^5$  guarantees that  $\frac{1-3\eta}{2} Tr/N > cT/N$  for a constant  $c$  and we can remove the last term in the lower bound of regret.

## 7.1 Proof strategy

We call a recommendation  $a_{u,t}$  to user  $u$  at time  $t$  a *bad* (or uncertain) recommendation when the probability of  $L_{u,a_{u,t}} = +1$  given the history is close to (or smaller than)  $1/2$ . Conversely, recommendations for which the probability of  $L_{u,a_{u,t}} = +1$  is close to one (much greater than  $1/2$ ) are considered *good* recommendations. Good and bad refers only to the confidence that the recommendation is liked given the history at the moment the recommendation is made: 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 in the item structure only model introduced in Definition 2.2: (i) A good estimate of item types is necessary in order to make meaningful recommendations. To determine whether or not two items are of the same type, approximately  $\log q_I$  users should rate both of them. To formalize this, 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 are liked by any user with probability roughly half, even if the preference matrix is known. (ii) Even when we know the item types (i.e., clustering of items), if a given user  $u$  has not rated any item with the same type as item  $i$  before, the probability that user  $u$  likes item  $i$  is  $1/2$ . Lemma 7.4 shows this property.

In Lemma 7.5 we bound regret in terms of the number of good recommendations and in Lemma 7.6 we upper bound the number of good recommendations, which entails a counting argument. The theorem follows immediately from Lemmas 7.5 and 7.6.

## 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 (Definition 5.1). We write  $A^\top \in \Omega_{r, \eta}$ , 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  $A$  is  $(r, \eta)$ -row regular with probability at least*

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

Throughout this section we will fix

$$r = \lfloor 0.8 \log q_I - 4 \log \log N \rfloor \quad \text{and} \quad \eta = \frac{1}{\log N}. \quad (37)$$

Applying Lemma 7.2 with these choices of  $\mathbf{r}$  and  $\eta$ , we obtain that so long as  $N > 32$  the preference matrix  $\Xi$  is  $(\mathbf{r}, \eta)$  row-regular (*i.e.*,  $\Xi^\top \in \Omega_{\mathbf{r}, \eta}$ ) with probability

$$\mathbb{P}[\Xi^\top \in \Omega_{\mathbf{r}, \eta}] \geq 1 - \frac{1}{N}. \quad (38)$$

Let  $c_i^t$  be the number of times item  $i$  has been rated by any user by the end of time-step  $t - 1$ ,

$$c_i^t := \sum_{u=1}^N \sum_{s=1}^{t-1} \mathbb{1}[a_{u,s} = i]. \quad (39)$$

The following lemma shows that if  $c_i^t$  is small and the preference matrix is row-regular, then the outcome of recommending item  $i$  to any user at time  $t$  is uncertain.

**Lemma 7.3.** *Let  $\mathbf{r}$  and  $\eta$  be as in Equation (37). For any user  $u \in [N]$  and item  $i \in \mathbb{N}$ ,*

$$\mathbb{P}[L_{u,i} = +1 \mid a_{u,t} = i, c_i^t < \mathbf{r}, \Xi^\top \in \Omega_{\mathbf{r}, \eta}] \leq \frac{1 + 3\eta}{2}.$$

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

$$\mathcal{B}_{u, \tau_I(i)}^t = \{\exists i' \in \mathbb{N} : a_{u,s} = i' \text{ for some } s < t \text{ and } \tau_I(i) = \tau_I(i')\}. \quad (40)$$

**Lemma 7.4.** *We have that  $\mathbb{P}[L_{u,i} = -1 \mid a_{u,t} = i, (\mathcal{B}_{u, \tau_I(i)}^t)^c, c_i^t \geq \mathbf{r}] = \frac{1}{2}$ .*

Lemmas 7.3 and 7.4 (proved in Section 7.3) identify scenarios in which recommendations are bad. In the complementary scenario, recommendations are not necessarily bad (and may be good). We denote the number of such recommendations by

$$\text{good}(T) = \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{1}[c_{a_{u,t}}^t \geq \mathbf{r}, \mathcal{B}_{u, \tau_I(a_{u,t})}^t]. \quad (41)$$

The proof of the following lemma uses Lemmas 7.3 and 7.4 to lower bound regret in terms of expectation of  $\text{good}(T)$ .

**Lemma 7.5.** *For  $\eta$  and  $\mathbf{r}$  defined in Equation (37),*

$$N \text{regret}(T) \geq \frac{1 - 3\eta}{2} (TN - \mathbb{E}[\text{good}(T)]) - T.$$

*Proof.* We partition the liked recommendations based on  $c_{a_u,t}^t$ ,  $\mathcal{B}_{u,\tau_I(a_u,t)}^t$ , and row regularity of  $\Xi$ :

$$\begin{aligned}
N(T - \text{regret}(T)) &= \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{P}[L_{u,a_u,t} = +1] \\
&= \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{P}[L_{u,a_u,t} = +1, c_{a_u,t}^t < r, \Xi^\top \notin \Omega_{r,\eta}] + \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{P}[L_{u,a_u,t} = +1, c_{a_u,t}^t < r, \Xi^\top \in \Omega_{r,\eta}] \\
&\quad + \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{P}[L_{u,a_u,t} = +1, c_{a_u,t}^t \geq r, (\mathcal{B}_{u,\tau_I(a_u,t)}^t)^c] + \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{P}[L_{u,a_u,t} = +1, c_{a_u,t}^t \geq r, \mathcal{B}_{u,\tau_I(a_u,t)}^t] \\
&=: \mathbf{A1} + \mathbf{A2} + \mathbf{A3} + \mathbf{A4}.
\end{aligned}$$

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

**Bounding A1.** Plugging in the probability of  $\Xi$  being row regular from (38) gives

$$\mathbf{A1} = \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{P}[L_{u,a_u,t} = +1, c_{a_u,t}^t < r, \Xi^\top \notin \Omega_{r,\eta}] \leq NT \cdot \mathbb{P}[\Xi^\top \notin \Omega_{r,\eta}] \leq T.$$

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

$$\begin{aligned}
\mathbf{A2} &= \sum_{\substack{t \in [T] \\ u \in [N]}} \sum_{i \geq 1} \mathbb{P}[L_{u,i} = +1, a_{u,t} = i, c_i^t < r, \Xi^\top \in \Omega_{r,\eta}] \\
&\leq \frac{1+3\eta}{2} \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{P}[c_{a_u,t}^t < r, \Xi^\top \in \Omega_{r,\eta}] \leq \frac{1+3\eta}{2} \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{P}[c_{a_u,t}^t < r].
\end{aligned}$$

**Bounding A3.** Lemma 7.4 gives

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

Multiplying this by  $\mathbb{P}[a_{u,t} = i, c_i^t \geq r, (\mathcal{B}_{u,\tau_I(i)}^t)^c]$  and summing over  $i$  gives

$$\mathbf{A3} = \frac{1}{2} \sum_{u \in [N], t \in [T]} \mathbb{P}[c_{a_u,t}^t \geq r, (\mathcal{B}_{u,\tau_I(a_u,t)}^t)^c].$$

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

$$\mathbf{A4} = \sum_{u \in [N], t \in [T]} \mathbb{P}[L_{u,a_u,t} = +1, c_{a_u,t}^t \geq r, \mathcal{B}_{u,\tau_I(a_u,t)}^t] \leq \sum_{u \in [N], t \in [T]} \mathbb{P}[c_{a_u,t}^t \geq r, \mathcal{B}_{u,\tau_I(a_u,t)}^t] = \mathbb{E}[\text{good}(T)]. \quad \square$$

Next, we upper bound the expected number of *good* recommendations made by the algorithm in terms of parameters of the model.

**Lemma 7.6** (Upper Bound for expected number of good recommendations). *For any algorithm,*

$$\mathbb{E}[\text{good}(T)] \leq TN - \max \left\{ \frac{1}{2}NZ(T), Tr, N \right\}.$$

where  $Z(T)$  is defined in (36).

We prove this lemma in the next subsection. Theorem 7.1 is an immediate consequence of Lemmas 7.5 and 7.6.

### 7.3 Proofs of lemmas

#### 7.3.1 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 even if the preference matrix is known. For a row-regular preference matrix, uncertainty in the type of an item makes it impossible to accurately predict user preferences for that item.

Consider item  $i$  at time  $t$ . Let  $w = \{u \in [N] : a_{u,s} = i, s < t\}$  be the ordered tuple corresponding to 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  $K_{b,w}(\Xi^\top)$  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  $\Xi$ ,  $w$ , and  $b$ , the type  $\tau_I(i)$  of item  $i$  at the end of time  $t - 1$  is uniformly distributed over the set of item types  $K_{b,w}(\Xi^\top)$ . This allows us to relate the posterior probability of  $i$  being liked to row regularity of  $\Xi$  as follows. Let  $b^+ = [b \ 1] \in \{-1, +1\}^{|w|+1}$  be obtained from  $b$  by appending  $+1$ . For a given user  $u \notin w$ , we have  $L_{u,i} = +1$  precisely when  $\tau_I(i) \in K_{b^+, \{w, u\}}(\Xi^\top)$ , 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  $\Xi$  and any user  $u$  which has not rated  $i$  up to time  $t - 1$ ,

$$\mathbb{P}[L_{u,i} = +1 \mid \mathcal{H}_{t-1}, \Xi] = \mathbb{P}[\tau_I(i) \in K_{b^+, \{w, u\}}(\Xi^\top) \mid \mathcal{H}_{t-1}, \Xi] = \frac{k_{b^+, \{w, u\}}(\Xi^\top)}{k_{b, w}(\Xi^\top)}. \quad (42)$$

The second equality is due to: (i)  $w$  and  $b$  are functions of the history up to time  $t - 1$ ,  $\mathcal{H}_{t-1}$ ; (ii) for fixed  $w$  and  $b$ , the set  $K_{b,w}(\Xi^\top)$  is determined by  $\Xi^\top$ ; (ii)  $\tau_I(i)$  is uniformly distributed on  $K_{b,w}(\Xi^\top)$  given  $\mathcal{H}_{t-1}$ .

Recall that  $\Xi^\top \in \Omega_{r,\eta}$  if the preference matrix  $\Xi$  is  $(r, \eta)$ -row regular. We have

$$\begin{aligned} \mathbb{P}[L_{u,i} = +1, a_{u,t} = i \mid c_i^t \leq r, \Xi^\top \in \Omega_{r,\eta}] &\stackrel{(a)}{=} \mathbb{E} \left[ \mathbb{P}[L_{u,i} = +1 \mid \mathcal{H}_{t-1}, \Xi] \mathbb{P}[a_{u,t} = i \mid \mathcal{H}_{t-1}] \mid c_i^t \leq r, \Xi^\top \in \Omega_{r,\eta} \right] \\ &\stackrel{(b)}{=} \mathbb{E} \left[ \frac{k_{b^+, \{w, u\}}(\Xi^\top)}{k_{b, w}(\Xi^\top)} \mathbb{1}[u \notin w] \mathbb{P}[a_{u,t} = i \mid \mathcal{H}_{t-1}] \mid c_i^t \leq r, \Xi^\top \in \Omega_{r,\eta} \right] \\ &\stackrel{(c)}{\leq} \frac{(1 + \eta) \frac{\mathbf{q}_I}{2^{|w|+1}}}{(1 - \eta) \frac{\mathbf{q}_I}{2^{|w|}}} \mathbb{P}[a_{u,t} = i \mid c_i^t \leq r, \Xi^\top \in \Omega_{r,\eta}] \stackrel{(d)}{\leq} \frac{1}{2} (1 + 3\eta) \mathbb{P}[a_{u,t} = i \mid c_i^t \leq r, \Xi^\top \in \Omega_{r,\eta}], \end{aligned}$$

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 (42) and the fact that  $\mathbb{P}[a_{u,t} = i \mid \mathcal{H}_{t-1}]$  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  $\Xi^\top \in \Omega_{r,\eta}$ , then by Claim 5.2,  $\Xi^\top \in \Omega_{r-1,\eta}$ . By Definition 5.1, this means that  $k_{b,w}(\Xi^\top) \geq (1 - \eta) \mathbf{q}_I / 2^{|w|}$  and  $k_{b^+, \{w, u\}}(\Xi^\top) \leq (1 + \eta) \mathbf{q}_I / 2^{|w|+1}$ . (d) If  $N > 32$ , then  $\eta$  in (37) is less than  $1/\log 32$  and  $(1 + \eta)/(1 - \eta) \leq 1 + 3\eta$ .

### 7.3.2 Proof of Lemma 7.4

At a high level, we make two observations: (i) if user  $u$  has not rated any item with type  $\tau_I(i)$  before, the feedback in the 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) the types of items (function  $\tau_I(\cdot)$ ) are independent of matrix  $\Xi$ , and the elements of  $\Xi$  are independent. Hence, conditional on  $(\mathcal{B}_{u, \tau_I(i)}^t)^c$ , the posterior distribution at time  $t$  of  $L_{u,i}$  is uniform on  $\{-1, +1\}$ .

Concretely, we may think of revealing the entries of  $\Xi$  on a “need-to-know” basis. Conditional on  $(\mathcal{B}_{u, \tau_I(i)}^t)^c$ , entry  $\xi_{\tau_U(u), \tau_I(i)}$  has not yet been touched, so

$$\begin{aligned} \mathbb{P}[L_{u,i} = +1 \mid a_{u,t} = i, \tau_I(i) = j, (\mathcal{B}_{u, \tau_I(i)}^t)^c, c_i^t \geq r, \tau_U(u)] \\ = \mathbb{P}[\xi_{\tau_U(u), j} = +1 \mid a_{u,t} = i, \tau_I(i) = j, (\mathcal{B}_{u, \tau_I(i)}^t)^c, c_i^t \geq r, \tau_U(u)] = \frac{1}{2}. \end{aligned}$$

The lemma now follows by the tower property.

### 7.3.3 Proof of Lemma 7.6

Lemma 7.6 upper bounds the expected number of good recommendations. To prepare for the proof of this lemma we introduce some notation. Recall that  $c_i^t$  (defined in (39)) is the number of users who have rated item  $i$  before time  $t$ . Let  $\mathcal{F}_t = \{i : c_i^t \geq r\}$  be the set of items with at least  $r$  ratings before time  $t$  and  $f_t := |\mathcal{F}_t|$  their number,

$$f_t = \sum_{i \in \mathbb{N}} \mathbb{1}[c_i^t \geq r].$$



Let  $\mathcal{G}_t = \{i : 0 < c_i^t < r\}$  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,

$$g_t = \sum_{i \in \mathbb{N}} \mathbb{1}[0 < c_i^t < r].$$

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

**Claim 7.7.** *The number of good recommendations  $\text{good}(T)$ , defined in (41), satisfies  $\text{good}(T) \leq TN - g_T - f_T r$ .*

*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  $r$  recommendations in which  $i$  has been rated fewer than  $r$  previous times:

$$\sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{1}[c_i^t < r, a_{u,t} = i] = r, \quad \text{for } i \in \mathcal{F}_T. \quad (43)$$

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

$$\sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{1}[c_i^t < r, a_{u,t} = i] \geq 1, \quad \text{for } i \in \mathcal{G}_T.$$

All recommended items are either in  $\mathcal{F}_T$  or  $\mathcal{G}_T$ . So, the total number of recommendations satisfies

$$\begin{aligned} NT &= \sum_{\substack{t \in [T] \\ u \in [N]}} \sum_{i \geq 1} \mathbb{1}[a_{u,t} = i] = \sum_{\substack{t \in [T] \\ u \in [N]}} \sum_{i \geq 1} \mathbb{1}[c_i^t \geq r, a_{u,t} = i] + \sum_{t=1}^T \sum_{u=1}^N \sum_{i \geq 1} \mathbb{1}[c_i^t < r, a_{u,t} = i] \\ &\geq \text{good}(T) + \sum_{\substack{t \in [T] \\ u \in [N]}} \sum_{i \in \mathcal{F}_T} \mathbb{1}[c_i^t < r, a_{u,t} = i] + \sum_{\substack{t \in [T] \\ u \in [N]}} \sum_{i \in \mathcal{G}_T} \mathbb{1}[c_i^t < r, a_{u,t} = i] \\ &\geq \text{good}(T) + f_T r + g_T. \end{aligned}$$

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

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

*Proof of Lemma 7.6.* The lemma consists of three bounds on  $\text{good}(T)$ .

**Proof of inequality**  $\mathbb{E}[\text{good}(T)] \leq TN - Tr$ . This bound is the easiest, so we start with this.

$$\begin{aligned}
\text{good}(T) &\stackrel{(a)}{\leq} \sum_{\substack{t \in [T] \\ u \in [N]}} \sum_{i \geq 1} \mathbb{1}[c_i^t \geq r, a_{u,t} = i] = \sum_{\substack{t \in [T] \\ u \in [N]}} \sum_{i: c_i^T \geq r} \mathbb{1}[c_i^t \geq r, a_{u,t} = i] \\
&\stackrel{(b)}{\leq} \sum_{i \in \mathcal{F}_T} \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{1}[a_{u,t} = i] - \mathbb{1}[c_i^t < r, a_{u,t} = i] \stackrel{(c)}{\leq} f_T(N - r).
\end{aligned}$$

(a) uses the definition of  $\text{good}(T)$ . For any item  $i$ , the sequence  $c_i^t$  is nondecreasing in  $t$ . This shows that if for an item  $c_i^t > r$  at  $t \leq T$ , then item  $i$  satisfies  $c_i^T \geq r$ . (b) is derived by changing the order of summations, the definition of  $\mathcal{F}_T$  and equality  $\mathbb{1}[c_i^t \geq r, a_{u,t} = i] = \mathbb{1}[a_{u,t} = i] - \mathbb{1}[c_i^t < r, a_{u,t} = i]$ . (c) Because each item is recommended at most once to each user, each item (and specifically items in  $\mathcal{F}_T$ ) are recommended at most  $N$  times. This gives for  $i \in \mathcal{F}_T$ ,  $\sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{1}[a_{u,t} = i] \leq N$  as the bound for the first term in (c). The second term in (c) is bounded by Equation (43).

Plugging this into the statement of Claim 7.7 gives for any values of  $f_T$  and  $g_T$ ,

$$\text{good}(T) \leq T(N - r). \quad (44)$$

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

$$\Gamma_u^T = \left\{ j \in [q_I] : \sum_{\substack{t \in [T] \\ i: \tau_I(i) = j}} \mathbb{1}[a_{u,t} = i] \neq 0 \right\}, \quad \text{and} \quad \gamma_u^T = |\Gamma_u^T|. \quad (45)$$

**Claim 7.8.**  $\text{good}(T) \leq N[T - \min_u \gamma_u^T]$ .

*Proof.* For user  $u$ , the number of times an item type is rated for the first time by this user is equal to the number of item types rated by user  $u$ . Hence,  $\sum_{t \in [T]} \mathbb{1}[(\mathcal{B}_{u, \tau_I(a_{u,t})}^t)^c] = \gamma_u^T$  and  $\sum_{t \in [T]} \mathbb{1}[\mathcal{B}_{u, \tau_I(a_{u,t})}^t] = T - \gamma_u^T$ . Summing over  $u$  and using  $\text{good}(T) \leq \sum_{\substack{t \in [T] \\ u \in [N]}} \mathbb{1}[\mathcal{B}_{u, \tau_I(a_{u,t})}^t]$  proves the claim.  $\square$

**Proof of inequality**  $\mathbb{E}[\text{good}(T)] \leq NT - N$ . This follows from Claim 7.8 since  $\gamma_u^T \geq 1$  for all  $u \in [N]$ .

**Proof of inequality**  $\mathbb{E}[\text{good}(T)] \leq NT - \frac{1}{2}NZ(T)$ . This last inequality is more involved than the others. Let  $r_j^t$  be the number of items with type  $j$  that have been recommended (to any user) by time  $t$ :

$$r_j^t = |\{i : \tau_I(i) = j, a_{u,s} = i \text{ for some } u \in [N] \text{ and } s \in [t]\}|. \quad (46)$$

By time  $T$ ,  $T$  items with types in  $\Gamma_u^T$  (defined in (45)) are recommended to user  $u$ . Hence,  $T \leq$

$\sum_{j \in \Gamma_u^T} r_j^T \leq \gamma_u^T \max_j r_j^T$ . This implies that

$$\min_{u \in [N]} \gamma_u^T \geq \frac{T}{\max_j r_j^T}. \quad (47)$$

We get a lower bound on  $\min_u \gamma_u^T$  via an upper bound on  $\max_j r_j^T$ , which is in turn obtained via martingale concentration bounds 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 random. 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.

**Claim 7.9.** *Let  $\delta > 0$ . Define  $k_0, k_1, k_2$  and the function  $\Theta_{q_I}(k)$  as follows:*

$$\Theta_{q_I}(k) := \begin{cases} 2, & \text{if } k < k_0 := \sqrt{q_I}/3 \\ 3 \frac{\log q_I}{\log q_I - \log k}, & \text{if } k_0 \leq k < k_1 := q_I/2 \\ 8 \log q_I, & \text{if } k_1 \leq k < k_2 := 2q_I \log q_I \\ 4k/q_I, & \text{if } k_2 \leq k. \end{cases} \quad (48)$$

For  $r_j^T$  defined in (46),  $\mathbb{P} \left[ \max_{j \in [q_I]} r_j^T \geq \Theta_{q_I}(f_T + g_T) \right] \leq 1/2$ .

*Proof.* First, we define a useful martingale. Let  $r^t = (r_1^t, \dots, r_{q_I}^t)$  where  $r_j^t$  is defined in (46). Note that  $f_t + g_t = \sum_j r_j^t$  is the total number of recommended items at the end of time  $t$ . Any new item has type uniformly distributed on  $[q_I]$ ; as a consequence, the sequence  $r_j^t - (f_t + g_t)/q_I$  is a martingale with respect to filtration  $\mathcal{F}_t = \sigma(r^0, r^1, \dots, r^t)$ , because  $f_t + g_t$  is incremented whenever a new item is recommended and each new item increases  $r_j^t$  by one with probability  $1/q_I$ .

It turns out to be easier to work with a different martingale that considers recommendations to 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  $v \leq u$  (such that the recommendation to user  $v$  at time  $s$  occurred before that of user  $u$  at time  $t$ ). For  $j \in [q_I]$ , let

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

Let  $r^{t,u} = (r_1^{t,u}, \dots, r_{q_I}^{t,u})$  and define  $\rho^{t,u} = \sum_j r_j^{t,u}$  to be the total number of items recommended by  $(t, u)$ , e.g.,  $\rho^{T,N} = f_T + g_T$ . We now define a sequence of stopping times  $Z_k \in \mathbb{N} \times [N]$ ,

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

where  $(t, 0)$  is interpreted as  $(t-1, N)$  and  $Z_0 = (0, N)$ .  $Z_k$  is the first  $(t, u)$  such that a new item is recommended by the algorithm for the  $k$ -th time, so  $\rho^{Z_k} = k$ . The  $Z_k$  are stopping times with respect to  $(\rho^{t,u})$ , and observe that  $k^* = \max\{k : Z_k \leq (T, N)\} = \rho^{T,N} = f_T + g_T$  since  $f_T + g_T$  is the total number of items recommended by the algorithm by the end of time  $T$ . Also,  $\rho^{Z_{k^*}} = f_T + g_T$  and  $r_j^{Z_{k^*}} = r_j^{(T,N)}$  for all  $j \in [q_I]$ .

Fix item type  $j \in [\mathbf{q}_I]$ . The sequence  $M_j^{t,u} = r_j^{t,u} - \rho^{t,u}/\mathbf{q}_I$  is a martingale with respect to the filtration  $\mathcal{F}^{t,u} = \sigma(r^{1,1}, \dots, r^{t,u})^\dagger$ . It follows that  $\widetilde{M}_j^k := M_j^{(T,N) \wedge Z_k}$  is martingale as well, this time with respect to  $\widetilde{\mathcal{F}}^k := \mathcal{F}^{(T,N) \wedge Z_k}$ . Since  $Z_{k^*} \leq (T, N)$ , we have  $\widetilde{M}_j^{k^*} = M_j^{Z_{k^*}}$ . We will use this notation to prove statement of the claim in three different regimes. First, we would like to apply martingale concentration (Lemma A.2) to  $\widetilde{M}_j^k$ , and to this end observe that  $\text{Var}(\widetilde{M}_j^k | \widetilde{\mathcal{F}}^{k-1}) \leq 1/\mathbf{q}_I$  and  $|\widetilde{M}_j^k - \widetilde{M}_j^{k-1}| \leq 1$  almost surely.

**Step 1** For any  $k \geq k_2 := 2\mathbf{q}_I \log \mathbf{q}_I$ , Lemma A.2 gives

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

This gives

$$\begin{aligned} \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k^*}} \geq \Theta_{\mathbf{q}_I}(k^*), k^* \geq k_2\right] &\leq \mathbb{P}\left[\exists k \geq k_2 \text{ s.t. } \max_{j \in [\mathbf{q}_I]} r_j^{Z_k} \geq \frac{4k}{\mathbf{q}_I}\right] \\ &\stackrel{(a)}{\leq} \mathbb{P}\left[\exists k \geq k_2 \text{ s.t. } \widetilde{M}_j^k \geq \frac{3k}{\mathbf{q}_I}\right] \stackrel{(b)}{\leq} \sum_{k \geq k_2} \exp\left(-\frac{2k}{\mathbf{q}_I}\right) = \frac{\exp\left(-\frac{2k_2}{\mathbf{q}_I}\right)}{1 - \exp(-2/\mathbf{q}_I)} \stackrel{(c)}{\leq} \frac{1}{\mathbf{q}_I^2}. \end{aligned} \quad (49)$$

where (a) uses  $\rho^{Z_k} = k$ . (b) uses a union bound and the inequality in the last display. (c) uses definition of  $k_2$  and  $1 - \exp(-2/\mathbf{q}_I) > \mathbf{q}_I^{-2}$  (which is derived using  $e^{-a} \leq 1 - a + a^2/2$  and  $\mathbf{q}_I > 1$ ).

**Step 2** For any  $k_2 > k$  we get

$$\mathbb{P}\left[\widetilde{M}_j^k \geq 6 \log \mathbf{q}_I\right] \leq \exp\left(\frac{-36 \log^2 \mathbf{q}_I}{2(k/\mathbf{q}_I + 2 \log \mathbf{q}_I)}\right) \leq \exp\left(\frac{-36 \log \mathbf{q}_I}{2k_2/(\mathbf{q}_I \log \mathbf{q}_I) + 4}\right) \leq \frac{1}{\mathbf{q}_I^4}.$$

This gives

$$\begin{aligned} \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k^*}} \geq \Theta_{\mathbf{q}_I}(k^*), k^* < k_2\right] &\leq \mathbb{P}\left[\exists k < k_2 \text{ s.t. } \max_{j \in [\mathbf{q}_I]} r_j^{Z_k} \geq 8 \log \mathbf{q}_I\right] \\ &\stackrel{(a)}{\leq} \mathbb{P}\left[\exists k < k_2 \text{ s.t. } \widetilde{M}_j^k \geq 6 \log \mathbf{q}_I\right] \stackrel{(b)}{\leq} \sum_{k < k_2} \frac{1}{\mathbf{q}_I^4} \leq \frac{k_2}{\mathbf{q}_I^4} \leq \frac{1}{\mathbf{q}_I^7}. \end{aligned} \quad (50)$$

(a) uses  $\rho^{Z_k} = k$ . (b) uses the inequality in the above display.

---

<sup>†</sup>To see that, define the event  $\mathcal{E}_{u,t}^{\text{new}} = \{\text{the item } a_{u,t} \text{ has not been recommended before to anybody}\}$  where the order is based on the lexicographic order we define in the proof of Claim 7.9. Then,

$$\mathbb{E}\left[r_j^{t,u} - \frac{\rho^{t,u}}{\mathbf{q}_I} \mid \mathcal{F}^{t,u-1}\right] - \left[r_j^{t,u-1} - \frac{\rho^{t,u-1}}{\mathbf{q}_I}\right] \stackrel{(a)}{=} \mathbb{E}\left[\mathbb{1}[\tau_I(a_{u,t}) = j] - \frac{1}{\mathbf{q}_I} \mid \mathcal{F}^{t,u-1}, \mathcal{E}_{u,t}^{\text{new}}\right] \mathbb{P}\left[\mathcal{E}_{u,t}^{\text{new}} \mid \mathcal{F}^{t,u-1}\right] \stackrel{(b)}{=} 0,$$

(a) uses the fact that condition on event  $(\mathcal{E}_{u,t}^{\text{new}})^c$ , we have  $\rho^{t,u} = \rho^{t,u-1}$  and  $r_j^{t,u} = r_j^{t,u-1}$ . Equality (b) uses the assumption in the model which states that the prior distribution of type of an item which has not been recommended before is uniform over  $[\mathbf{q}_I]$ . Hence,  $\mathbb{P}\left[\tau_I(a_{u,t}) = j \mid \mathcal{F}^{t,u-1}, \mathcal{E}_{u,t}^{\text{new}}\right] = 1/\mathbf{q}_I$ .

**Step 3** This step,  $k < k_1 := \mathbf{q}_I/2$ , corresponds to bounding the number of balls in the fullest bin when the number of balls,  $k$ , is sublinear in the number of bins,  $\mathbf{q}_I$  (since  $k = \mathbf{q}_I^{1-3\delta}$  with  $\delta > \frac{1}{8 \log \mathbf{q}_I}$ ). We will show that in this regime, the number of balls in the fullest bin is bounded by  $1/\delta$ . For given  $k < k_1$ , define  $\delta = \frac{1}{3} \frac{\log \mathbf{q}_I - \log k}{\log \mathbf{q}_I}$  (such that  $k = \mathbf{q}_I^{1-3\delta}$ ). Then,

$$\mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_k} \geq 1/\delta\right] \stackrel{(a)}{\leq} \mathbf{q}_I \mathbb{P}\left[r_1^{Z_k} \geq 1/\delta\right] \stackrel{(b)}{\leq} \mathbf{q}_I \binom{k}{1/\delta} \frac{1}{\mathbf{q}_I^{1/\delta}} \stackrel{(c)}{\leq} \frac{5}{\mathbf{q}_I^2} \quad (51)$$

(a) is a union bound over  $j \in [\mathbf{q}_I]$ . (b) uses the fact that  $r_1^{Z_{k+1}} = r_1^{Z_k} + 1$  with probability  $1/\mathbf{q}_I$  and  $r_1^{Z_{k+1}} = r_1^{Z_k}$  with probability  $1 - 1/\mathbf{q}_I$  independently of  $r_1^{Z_k}$ . (c) holds for every  $\delta > \frac{1}{8 \log \mathbf{q}_I}$  (which is due to  $k < k_1$ ) using  $\binom{k}{1/\delta} \leq (ke\delta)^{1/\delta}$ .

$$\begin{aligned} \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{T,N} \geq 3 \frac{\log \mathbf{q}_I}{\log \mathbf{q}_I - \log k^*}, k^* < k_1\right] &\stackrel{(a)}{=} \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k^*}} \geq 3 \frac{\log \mathbf{q}_I}{\log \mathbf{q}_I - \log k^*}, k^* < k_1\right] \\ &\leq \mathbb{P}\left[\exists k < k_1 \text{ s.t. } \max_{j \in [\mathbf{q}_I]} r_j^{Z_k} \geq 3 \frac{\log \mathbf{q}_I}{\log \mathbf{q}_I - \log k}\right] \stackrel{(b)}{\leq} k_1 \frac{5}{\mathbf{q}_I^2} \leq \frac{5}{\mathbf{q}_I} \end{aligned} \quad (52)$$

(a) uses  $r_j^{Z_{k^*}} = r_j^{(T,N)}$ . (b) uses a union bound and (51). Last inequality uses  $k_1 < \mathbf{q}_I$ .

**Step 4** This step uses a variation of the Birthday Paradox to bound  $\max_{j \in [\mathbf{q}_I]} r_j^{T,N}$ .

$$\begin{aligned} \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{T,N} \geq 2, k^* < k_0\right] &= \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k^*}} \geq 2, k^* < k_0\right] \leq \mathbb{P}\left[\exists k < k_0 \text{ s.t. } \max_{j \in [\mathbf{q}_I]} r_j^{Z_k} \geq 2\right] \\ &\stackrel{(a)}{\leq} \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k_0}} \geq 2\right] = 1 - \mathbb{P}\left[r_j^{Z_{k_0}} \leq 1 \text{ for all } j \in [\mathbf{q}_I]\right] \\ &\stackrel{(b)}{=} 1 - \mathbb{P}\left[r_j^{Z_k} \leq 1 \text{ for all } j \in [\mathbf{q}_I] \text{ and } k \leq k_0\right] \\ &= 1 - \prod_{m=1}^{k_0} \mathbb{P}\left[r_j^{Z_m} \leq 1 \text{ for all } j \in [\mathbf{q}_I] \mid r_j^{Z_{m-1}} \leq 1 \text{ for all } j \in [\mathbf{q}_I]\right] \\ &\stackrel{(c)}{=} 1 - \prod_{m=1}^{k_0} \left(1 - \frac{m-1}{\mathbf{q}_I}\right) \leq 1 - \left(1 - \frac{k_0}{\mathbf{q}_I}\right)^{k_0} \stackrel{(d)}{\leq} \frac{2k_0^2}{\mathbf{q}_I} \leq \frac{2}{9}. \end{aligned} \quad (53)$$

(a) and (b) use the fact that  $r_j^{Z_k}$  is a nondecreasing function of  $k$ . We define  $r_j^{Z_0} = 0$ . The type of the  $(k+1)$ -th drawn item is independent of the type of the previous  $k$  drawn items. Hence, conditional on  $r^{Z_k} = (r_1^{Z_k}, \dots, r_{\mathbf{q}_I}^{Z_k})$ , the random variable  $r^{Z_{k+1}}$  is independent of  $r^{Z_{k-1}}$ . This gives equality (c). (d) uses  $\exp(k \log(1 - k/\mathbf{q}_I)) \geq \exp(-k^2/(\mathbf{q}_I - k)) \geq \exp(-2k^2/\mathbf{q}_I) \geq 1 - 2k^2/\mathbf{q}_I$  for  $k \leq k_0 \leq \sqrt{\mathbf{q}_I} \leq \mathbf{q}_I/2$ .

We put it all together,

$$\begin{aligned}
\mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^T \geq \Theta_{\mathbf{q}_I}(f_T + g_T)\right] &\stackrel{(a)}{=} \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k^*}} \geq \Theta_{\mathbf{q}_I}(k^*)\right] \\
&\leq \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k^*}} \geq \Theta_{\mathbf{q}_I}(k^*), k^* \geq k_2\right] + \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k^*}} \geq \Theta_{\mathbf{q}_I}(k^*), k_0 \leq k^* < k_2\right] \\
&\quad + \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k^*}} \geq \Theta_{\mathbf{q}_I}(k^*), \leq k^* < k_1\right] + \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^{Z_{k^*}} \geq \Theta_{\mathbf{q}_I}(k^*), k_0 \leq k^* < k_0\right] \\
&\stackrel{(b)}{\leq} \frac{2}{\mathbf{q}_I^2} + \frac{5}{\mathbf{q}_I} + \frac{2}{9} \leq \frac{1}{2}.
\end{aligned}$$

(a) uses the definition of  $Z_k$  and  $k^*$ . (b) uses (49), (50), (52) and (53). Last inequality uses  $\mathbf{q}_I > 10$ .  $\square$

Plugging  $\min_u \gamma_u^T \geq T / \max_{j \in [\mathbf{q}_I]} r_j^T$ , given in Equation (47) in Claim 7.9 proves

$$\mathbb{P}\left[\min_{u \in [N]} \gamma_u^T \leq \frac{T}{\Theta_{\mathbf{q}_I}(f_T + g_T)}\right] \leq \mathbb{P}\left[\max_{j \in [\mathbf{q}_I]} r_j^T \geq \Theta_{\mathbf{q}_I}(f_T + g_T)\right] \leq \frac{1}{2}.$$

This statement and Claim 7.8 imply that with probability at least 1/2 we have

$$\text{good}(T) \leq NT - N \frac{T}{\Theta_{\mathbf{q}_I}(f_T + g_T)}.$$

Combining this bound and Claim 7.7 gives that with probability at least 1/2,

$$\text{good}(T) \leq NT - \max\left\{f_T r + g_T, N \frac{T}{\Theta_{\mathbf{q}_I}(f_T + g_T)}\right\}. \quad (54)$$

Using the definition of the function  $\Theta_{\mathbf{q}_I}(k)$  in Claim 7.9 and some algebra, one can show that for any  $k > 0$ ,

$$\max\left\{k, \frac{NT}{\Theta_{\mathbf{q}_I}(k)}\right\} \geq \begin{cases} \frac{NT}{2}, & \text{if } T < \frac{2\sqrt{\mathbf{q}_I}}{3N} \\ \frac{NT}{5 \log \mathbf{q}_I} [\log(8\mathbf{q}_I \log \mathbf{q}_I) - \log(NT)], & \text{if } \frac{2\sqrt{\mathbf{q}_I}}{3N} \leq T < \frac{4\mathbf{q}_I \log \mathbf{q}_I}{N} \\ \frac{NT}{8 \log \mathbf{q}_I}, & \text{if } \frac{4\mathbf{q}_I \log \mathbf{q}_I}{N} \leq T < \frac{16\mathbf{q}_I (\log \mathbf{q}_I)^2}{N} \\ \frac{1}{2} \sqrt{NT \mathbf{q}_I}, & \text{if } \frac{16\mathbf{q}_I (\log \mathbf{q}_I)^2}{N} \leq T, \end{cases}$$

or alternatively,  $\max\left\{k, \frac{NT}{\Theta_{\mathbf{q}_I}(k)}\right\} \geq NZ(T)$  where  $Z(T)$  is defined in (36). To prove this, we show that if  $k < Z(T)$ , then  $\Theta_{\mathbf{q}_I}(k) \leq \frac{NT}{Z(T)}$  for each regime of parameter  $T$ . Note that the above bound is not tight, but it is chosen such that this lower bound (and consequently the function  $Z(T)$  which is a scaling of the above lower bound as defined in (36)) is continuous up to a multiplicative constant factor.

Equation (41) shows that  $\text{good}(T) \leq NT$  with probability one. Hence,

$$\mathbb{E}[\text{good}(T)] \leq \frac{1}{2}NT + \frac{1}{2}(\text{RHS of (54)}) \leq NT - \frac{1}{2}NZ(T).$$

This completes the proof of the lemma.  $\square$

## 8 Discussion

In this paper, we analyzed the performance of online collaborative filtering within a latent variable model for the preferences of users for items. We proposed variants of user-user CF and item-item CF that explicitly explore the preference space. We also proved lower bounds for regret in the extreme regimes of parameters corresponding to user-structure only (no structure in item space) and item-structure only (no structure in user space). The lower bounds showed that the proposed algorithms are almost information-theoretically optimal in these parameter regimes.

Adaptivity to unknown time horizon  $T$ , as required to bound the anytime regret, is achieved via a doubling trick whereby the algorithm is run afresh at a growing sequence of epochs. In practice one would surely benefit from using knowledge gained from exploration in earlier epochs instead of starting from scratch at each epoch. We mentioned how the user-user algorithm can be modified to be adaptive to the number  $q_U$  of user types, but it is less obvious how to make the item-item algorithm adaptive without resorting to an impractical trick analogous to the doubling trick used for  $T$ .

It is possible to modify all of the proposed algorithms to handle i.i.d. noise in the user feedback. We did this only for the user-user algorithm, but it is straightforward to do so also for the item-item algorithm. A hybrid algorithm, exploiting structure in both user space and item space, that is nearly information-theoretically optimal in all regimes appears in a forthcoming paper.

While various insights were obtained through the analysis carried out in the paper, the assumed randomly generated user preference matrix is unrealistic. A reasonable next objective is to perform a similar analysis with a more flexible model for user preferences, perhaps described by a low-rank matrix or a graphical model.

## Acknowledgment

We are indebted to Devavrat Shah, George Chen, and Luis Velasco for many inspiring discussions and thank 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 [28].

**Lemma A.1** (Chernoff bound). *Let  $X_1, \dots, X_n \in [0, 1]$  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  $\epsilon > 0$ ,*

$$\begin{aligned}\mathbb{P}[X \geq (1 + \epsilon)\bar{X}] &\leq \exp\left(-\frac{\epsilon^2}{2 + \epsilon}\bar{X}\right) \leq \max\left\{\exp\left(-\frac{\epsilon^2}{3}\bar{X}\right), \exp\left(-\frac{\epsilon}{2}\bar{X}\right)\right\} \\ \mathbb{P}[X \leq (1 - \epsilon)\bar{X}] &\leq \exp\left(-\frac{\epsilon^2}{2}\bar{X}\right) \\ \mathbb{P}[|X - \bar{X}| \geq \epsilon\bar{X}] &\leq 2 \max\left\{\exp\left(-\frac{\epsilon^2}{3}\bar{X}\right), \exp\left(-\frac{\epsilon}{2}\bar{X}\right)\right\}\end{aligned}$$

**Lemma A.2** (McDiarmid [29]). *Let  $X_1, \dots, X_n$  be a martingale adapted to filtration  $(\mathcal{F}_n)$  satisfying*

- (i)  $\text{Var}(X_i | \mathcal{F}_{i-1}) \leq \sigma_i^2$ , 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, using union bound, the probability of at most  $m/2$  nonempty bins is bounded by

$$\binom{n}{m/2} \left(\frac{m/2}{n}\right)^m \leq \left(\frac{n \cdot e}{m/2}\right)^{m/2} \left(\frac{m/2}{n}\right)^m \leq \left(\frac{me}{2n}\right)^{m/2} \leq \exp(-m/2),$$

where we used  $m \leq n/4$ . □

The following generalization of the above lemma is used in the analysis of the recommendation system in noisy setup.

**Lemma A.4** (Generalized Balls and bins: tail bound for number of nonempty bins). *Fix  $0 < a < 1$  and  $c > 0$ . Define  $b \triangleq \min\{t : -(1 - a/t) \log a + (1 - a)/t \log(1 - a) > c\}$ . Suppose  $m \leq n/b$ . If  $m$  balls are placed into  $n$  bins each independently and uniformly at random, then with probability at least  $1 - \exp(-cm)$  at least  $na$  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

$$\binom{n}{m/2} \left(\frac{m/2}{n}\right)^m \leq \left(\frac{n \cdot e}{m/2}\right)^{m/2} \left(\frac{m/2}{n}\right)^m \leq \left(\frac{me}{2n}\right)^{m/2} \leq \exp(-m/2),$$

where we used  $m \leq n/4$ .  $\square$

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

**Lemma A.5** (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 that does not require knowledge of the time horizon and yet achieves the same regret (up to multiplicative constant) at any time [27] (i.e., *anytime regret*).

The trick is to divide time into intervals and restart algorithm at the beginning of each interval. Let  $A(T)$  be an online algorithm taking the known time horizon as input and achieving regret  $R(T)$  at time  $T$ . There are two regret scalings of interest. (1) If  $R(T) = O(T^\alpha)$  for some  $0 < \alpha < 1$ , then to achieve anytime regret, the doubling trick uses time intervals of length  $2, 2^2, 2^3, \dots, 2^m$ . This achieves regret of at most  $R(T)/(1 - 2^\alpha)$  at time  $T$  for any  $T$ . (2) Alternatively, if  $R(T) = O(\log T)$ , then using intervals of length  $2^2, 2^{2^2}, \dots, 2^{2^m}$  achieves regret of at most  $4R(T)$  at time  $T$  for any  $T$ .

Clearly, different scalings can be used before and after  $T_1$  if the algorithm achieves regret  $O(\log T)$  for  $T < T_1$  and  $O(\sqrt{T})$  if  $T \geq T_1$ , as is the case for the proposed item-item CF algorithm.

## C Proof of Lemma 4.5

We show that  $\mathbb{P}[B_{uv}^c] \leq 2\epsilon/(N^2)$  for any pair of users  $u, v \in [N]$ . Using a union bound over the  $\binom{N}{2}$  pairs of users gives  $\mathbb{P}[B^c] \leq \epsilon$ . According to Line 7 of the Algorithm 4.5,  $g_{u,v} = \mathbb{1}\{\sum_{s=1}^r L_{u,a_{u,s}} L_{v,a_{v,s}} \geq \lambda r\}$ . At  $s \leq r$ , the same item is recommended to all users:  $a_s := a_{u,s} = a_{v,s}$ . Hence,  $L_{u,a_{u,s}} = \xi_{\tau_U(u), \tau_I(a_s)} z_{u,a_s}$  and  $L_{v,a_{v,s}} = \xi_{\tau_U(v), \tau_I(a_s)} z_{v,a_s}$ . First, we look at the users of the same type:

$$\begin{aligned} \mathbb{P}[g_{u,v} \neq 1 | \tau_U(u) = \tau_U(v)] &\stackrel{(a)}{=} \mathbb{P}\left[\sum_{s=1}^r L_{u,a_{u,s}} L_{v,a_{v,s}} < \lambda r | \tau_U(u) = \tau_U(v)\right] \\ &\stackrel{(b)}{=} \mathbb{P}\left[\sum_{s=1}^r \xi_{\tau_U(u), \tau_I(a_s)} z_{u,a_s} \xi_{\tau_U(v), \tau_I(a_s)} z_{v,a_s} < \lambda r | \tau_U(u) = \tau_U(v)\right] \\ &\stackrel{(c)}{=} \mathbb{P}\left[\sum_{s=1}^r z_{u,a_s} z_{v,a_s} < \lambda r | \tau_U(u) = \tau_U(v)\right] \stackrel{(d)}{=} \mathbb{P}\left[\sum_{s=1}^r z_{u,a_s} z_{v,a_s} < \lambda r\right] \stackrel{(e)}{\leq} \exp\left(-\frac{(1-2\gamma)^2 r}{18}\right) \stackrel{(f)}{\leq} \frac{2\epsilon}{N^2}, \end{aligned}$$

where (a) uses the definition of  $g_{u,v}$  according to the Line 7 of Algorithm 2. (b) uses the noise model (11). If  $\tau_U(u) = \tau_U(v)$  then  $\xi_{\tau_U(u),\tau_I(a_s)} = \xi_{\tau_U(v),\tau_I(a_s)}$  which gives (c). The variables  $z_{u,i}$  are independent of other variables in the model which implies (d). For users  $u \neq v$  and any item  $i$ ,  $\mathbb{E}[z_{u,i}z_{v,i}] = (1 - 2\gamma)^2$ . Lemma A.1 and the choice of  $\lambda$  in Algorithm 2 gives (e). The choice of  $r$  gives the result (f).

Now, we look at the pair of users  $u$  and  $v$  such that  $\tau_U(u) \neq \tau_U(v)$ . Define  $\mathcal{L}_{u,v} = |\{j : \xi_{\tau_U(u),j} \neq \xi_{\tau_U(v),j}\}|$  to be the set of item types on which user types  $\tau_U(u)$  and  $\tau_U(v)$  disagree.

$$\begin{aligned}
\mathbb{P}[g_{u,v} \neq 0 | \tau_U(u) \neq \tau_U(v)] &\stackrel{(a)}{=} \mathbb{P}\left[\sum_{s=1}^r \xi_{\tau_U(u),\tau_I(a_s)} z_{u,a_s} \xi_{\tau_U(v),\tau_I(a_s)} z_{v,a_s} \geq \lambda r | \tau_U(u) \neq \tau_U(v)\right] \\
&\stackrel{(b)}{\leq} \mathbb{P}\left[|\mathcal{L}_{u,v}| < 5\mathbf{q}_I/12 | \tau_U(u) \neq \tau_U(v)\right] \\
&\quad + \mathbb{P}\left[\sum_{s=1}^r \xi_{\tau_U(u),\tau_I(a_s)} \xi_{\tau_U(v),\tau_I(a_s)} z_{u,a_s} z_{v,a_s} \geq \lambda r | |\mathcal{L}_{u,v}| > 5\mathbf{q}_I/12, \tau_U(u) \neq \tau_U(v)\right] \\
&\stackrel{(c)}{\leq} \exp(-\mathbf{q}_I/144) + \exp(-r(1-2\gamma)^2/4) \leq 2\epsilon/N^2.
\end{aligned}$$

(a) uses the definition of  $g_{u,v}$  in Line 7 of Algorithm 2. Total probability lemma gives (b). Conditional on  $\tau_U(u) \neq \tau_U(v)$ , the variables  $\xi_{\tau_U(u),j}$  and  $\xi_{\tau_U(v),j}$  are independently uniformly distributed. Using the definition on  $\mathcal{L}_{u,v}$ , this implies that  $|\mathcal{L}_{u,v}|$  is the sum of  $\mathbf{q}_I$  i.i.d. uniform Bernoulli random variables. Hence, Chernoff bound in Lemma A.1 gives the bound on the first term in (c). To bound the second term, note that the items  $a_s$  are chosen independently of feedback for  $s \leq r$ . Hence, conditional on  $|\mathcal{L}_{u,v}| \geq 5\mathbf{q}_I/12$ , the variables  $\xi_{\tau_U(u),\tau_I(a_s)} \xi_{\tau_U(v),\tau_I(a_s)} = -1$  with probability at least  $5/12$ . The variables  $z_{u,a_s}$  and  $z_{v,a_s}$  are independent of other parameters of the model and algorithm and  $z_{u,a_s} z_{v,a_s} = -1$  with probability  $2\gamma(1-\gamma)$ . Hence,  $\xi_{\tau_U(u),\tau_I(a_s)} \xi_{\tau_U(v),\tau_I(a_s)} z_{u,a_s} z_{v,a_s} = -1$  with probability at least  $[5 + 4\gamma(1-\gamma)]/12$ . Using Chernoff bound in Lemma A.1 gives the second term in (c). The assumption  $\mathbf{q}_I > 144 \log(N^2/\epsilon)$  and the choice of  $r$  gives the result.

## References

- [1] S. Aditya, O. Dabeer, and B. K. Dey, “A channel coding perspective of collaborative filtering,” *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2327–2341, 2011.
- [2] G. Biau, B. Cadre, and L. Rouviere, “Statistical analysis of k-nearest neighbor collaborative recommendation,” *The Annals of Statistics*, vol. 38, no. 3, pp. 1568–1592, 2010.
- [3] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, vol. 9, no. 6, p. 717, 2009.
- [4] 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*. ACM, 2013, pp. 665–674.

- [5] R. H. Keshavan, A. Montanari, and S. Oh, “Matrix completion from a few entries,” *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2980–2998, 2010.
- [6] S. Negahban and M. J. Wainwright, “Restricted strong convexity and weighted matrix completion: Optimal bounds with noise,” *Journal of Machine Learning Research*, vol. 13, no. May, pp. 1665–1697, 2012.
- [7] A. Rohde and A. B. Tsybakov, “Estimation of high-dimensional low-rank matrices,” *The Annals of Statistics*, vol. 39, no. 2, pp. 887–930, 2011.
- [8] 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*, 2005, pp. 1321–1328.
- [9] S. Bubeck and N. Cesa-Bianchi, “Regret analysis of stochastic and nonstochastic multi-armed bandit problems,” *Foundations and Trends in Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.
- [10] T. L. Lai and H. Robbins, “Asymptotically efficient adaptive allocation rules,” *Advances in applied mathematics*, vol. 6, no. 1, pp. 4–22, 1985.
- [11] D. Russo and B. Van Roy, “Learning to optimize via information-directed sampling,” in *Advances in Neural Information Processing Systems*, 2014, pp. 1583–1591.
- [12] G. Bresler, G. H. Chen, and D. Shah, “A latent source model for online collaborative filtering,” in *Advances in Neural Information Processing Systems*, 2014, pp. 3347–3355.
- [13] G. Bresler, D. Shah, and L. F. Voloch, “Collaborative filtering with low regret,” in *SIGMETRICS Performance Evaluation Review*, vol. 44, no. 1. New York, NY, USA: ACM, Jun. 2016, pp. 207–220. [Online]. Available: <http://doi.acm.org/10.1145/2964791.2901469>
- [14] 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*. ACM, 2009, pp. 757–766.
- [15] 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*, vol. 42, no. 1. ACM, 2014, pp. 29–41.
- [16] 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*. ACM, 2012, pp. 213–216.
- [17] 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*. ACM, 2007, pp. 271–280.

- [18] G. Linden, B. Smith, and J. York, “Amazon.com recommendations: Item-to-item collaborative filtering,” *IEEE Internet computing*, vol. 7, no. 1, pp. 76–80, 2003.
- [19] 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*. ACM, 2001, pp. 285–295.
- [20] O. Dabeer, “Adaptive collaborating filtering: The low noise regime,” in *International Symposium on Information Theory Proceedings (ISIT)*. IEEE, 2013, pp. 1197–1201.
- [21] I. Kerenidis and A. Prakash, “Quantum recommendation systems,” *arXiv preprint arXiv:1603.08675*, 2016.
- [22] K. Barman and O. Dabeer, “Analysis of a collaborative filter based on popularity amongst neighbors,” *IEEE Transactions on Information Theory*, vol. 58, no. 12, pp. 7110–7134, 2012.
- [23] 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*, 2016, pp. 2155–2163.
- [24] 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*. ACM, 2006, pp. 501–508.
- [25] 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*. IEEE, 2010, pp. 462–466.
- [26] C. Borgs, J. Chayes, C. E. Lee, and D. Shah, “Thy friend is my friend: Iterative collaborative filtering for sparse matrix estimation,” in *Advances in Neural Information Processing Systems*, 2017, pp. 4715–4726.
- [27] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge University Press, 2006.
- [28] F. Chung and L. Lu, “Concentration inequalities and martingale inequalities: a survey,” *Internet Mathematics*, vol. 3, no. 1, pp. 79–127, 2006.
- [29] C. McDiarmid, “Concentration,” in *Probabilistic methods for algorithmic discrete mathematics*. Springer, 1998, pp. 195–248.