Optimal Algorithms for Latent Bandits with Cluster Structure

Soumyabrata Pal
Google Research

Arun Sai Suggala
Google Research

Karthikeyan Shanmugam
Google Research

Prateek Jain
Google Research

Abstract

We consider the problem of latent bandits with cluster structure where there are multiple users, each with an associated multi-armed bandit problem. These users are grouped into *latent* clusters such that the mean reward vectors of users within the same cluster are identical. At each round, a user, selected uniformly at random, pulls an arm and observes a corresponding noisy reward. The goal of the users is to maximize their cumulative rewards. This problem is central to practical recommendation systems and has received wide attention of late Gentile et al. [2014], Maillard and Mannor [2014]. Now, if each user acts independently, then they would have to explore each arm independently and a regret of $\Omega(\sqrt{\mathsf{MNT}})$ is unavoidable, where M, N are the number of arms and users, respectively. Instead, we propose LATTICE (Latent bAndiTs via maTrIx ComplEtion) which allows exploitation of the latent cluster structure to provide the minimax optimal regret of $\widetilde{O}(\sqrt{(\mathsf{M}+\mathsf{N})\mathsf{T}})$, when the number of clusters is $\tilde{O}(1)$. This is the first algorithm to guarantee such strong regret bound. LATTICE is based on a careful exploitation of arm information within a cluster while simultaneously clustering users. Furthermore, it is computationally efficient and requires only $O(\log \mathsf{T})$ calls to an offline matrix completion oracle across all T rounds.

1 INTRODUCTION

Bandit optimization is a very general framework for sequential decision making when the dynamics of the underlying environment are unknown a priori. It has been well studied over the past few decades, and has shown great empirical success in areas including ad placement, clinical trials [Lat-

timore and Szepesvári, 2020, Mate et al., 2022]. Such standard bandit methods often assume the decision maker has access to the entire user context or user state. However, this assumption rarely holds in practice. For example, in movie-recommendation scenarios, users can be clustered according to their taste in movies, but the observed features like user's demographic information might only be a noisy indicator of their taste. Similarly, in educational settings, the true cognitive state of a user is unknown. Instead we only get to observe a noisy estimate of the cognitive state through assessments. This shows that in practice, we need to optimize reward in the presence of observed as well as unobserved/latent context. Naturally, one option is to completely ignore the latent context as it can be thought of as part of the reward structure itself, but that generally leads to a significant increase in the sample complexity compared to the scenario when the latent structure is known a priori.

Several recent works have considered bandit optimization frameworks that explicitly factor in the latent data, and designed algorithms to maximize the cumulative rewards provided by the environment [Maillard and Mannor, 2014, Hong et al., 2020, Zhou and Brunskill, 2016]. However, as detailed below, even for simple latent structure like cluster of users, these papers either require strong assumptions or require additional side information, both of which are impractical.

In this work, we consider the problem of multi-user multi-armed bandits with latent clusters (MAB-LC). This is a simple yet powerful setting that captures several practically important multi-user scenarios like recommendation systems, and was introduced in Maillard and Mannor [2014]. Let there be N users, M arms and T rounds ($\mathsf{N}, \mathsf{M} \approx 10^6$ in recommendation systems such as Youtube). The N users can be partitioned into C latent clusters where users in the same cluster have identical reward distributions; in other words, users in the same cluster have similar preferences for arms. In every round, one of the users, sampled uniformly at random, pulls one of the arms and obtains certain feedback. The goal of the decision maker is to maximize the cumulative reward of all the users. This problem was first introduced and studied in Maillard and Mannor [2014] who provided theoretical guarantees for certain special settings (such as known cluster rewards or known cluster assign-

ments) but not for the general problem. Gentile et al. [2014], Li et al. [2016, 2019], Gentile et al. [2017], Qi et al. [2022] considered a contextual bandit variant of MAB-LC, which is a generalization of our setting. But, most of the existing methods either provide sub-optimal regret bounds, or require strong assumptions on context vectors that might not hold in practice. Hong et al. [2020] considered more general latent structures than the cluster structure we consider in this work, but required access to offline data for estimating the latent states. In another line of related work, Jain and Pal [2022] studied the online low rank matrix completion problem (a generalization of our setting). Jain and Pal [2022] could only obtain minimax optimal regret bounds in the special case of rank-1 setting. Obtaining optimal regret for the general rank-r problem is still an open problem [Jain and Pal, 2022]. To summarize, while the MAB-LC problem has been widely studied, to the best of our knowledge, designing an efficient method with nearly optimal regret bound is open.

Before moving ahead, it is instructive to consider two hypothetical scenarios that illustrate the complexity of MAB-LC. If the cluster assignment (*i.e.,* mapping between users and clusters) is known, we could have treated users within a cluster as a single super-user and solved a separate multi-armed bandit for each super-user. This leads to a regret of $\widetilde{O}(\sqrt{\mathsf{MCT}})$, which is minimax optimal. On the other hand, suppose the cluster assignment is unknown but the reward distributions of arms in a cluster are known. Then we could have played a separate multi-armed bandit problem for each user with the best arm from each cluster as a candidate arm. This leads to a regret of $\widetilde{O}(\sqrt{\mathsf{NCT}})$, which is minimax optimal. However, in MAB-LC, both cluster assignment and reward distributions are unknown. Consequently, the *reference* regret guarantee that one can hope to achieve is $O(\sqrt{(\mathsf{M}+\mathsf{N})\mathsf{CT}})$.

In this work, we propose a novel algorithm (LATTICE) for the problem of MAB-LC that achieves the above reference regret bound. The key challenge in solving this problem is to simultaneously cluster users, and quickly identify the optimal arms within each cluster. LATTICE addresses these problems using two key insights: (a) **(user clustering)** it uses low-rank matrix completion as an algorithmic tool to cluster users, and (b) **(arm elimination)** within each identified cluster, it discards sub-optimal arms by a careful exploitation of the accrued arm information. LATTICE runs in phases and performs both user clustering and arm elimination in each phase. Computationally, our algorithm is efficient and requires only $O(\log \mathsf{T})$ calls to an offline matrix completion oracle across all T rounds. Furthermore, under certain incoherence assumptions on the user-arm matrix, we show that our algorithm achieves the minimax optimal regret of $\widetilde{O}(\sqrt{(\mathsf{M}+\mathsf{N})\mathsf{T}})$, when $\mathsf{C} = \widetilde{O}(1)$. We note the incoherence assumptions seems unavoidable for statistical recovery with partial observations. In addition to minimax

optimal bounds, we also derive distribution-dependent regret bounds for our algorithm that inversely depend on the minimum gap between mean rewards of arms thus obtaining the optimal scaling with respect to gaps.

We also consider a more general and practical setting where we relax the cluster definition as follows: (a) for any two users in the same cluster, we let their reward vectors be ν entry-wise close to each other, and (b) for any two users from different clusters, their respective best arm rewards are separated by more than $c\nu$, for some $c > 1$. Note that, we don't require large separation between users in the mean rewards of sub-optimal arms across clusters. We show that a modification of LATTICE obtains similar regret bounds as before in this general setting.

1.1 Other Related Work

Contextual Bandits with Latent Structure. An extensive line of work [Gentile et al., 2014, Li et al., 2019, Gentile et al., 2017, Li et al., 2016, Qi et al., 2022] studies a variant of MAB-LC where every arm is associated with a context vector of dimension d and expected arm reward for any user in a fixed cluster is a unknown linear function (depending only on the cluster and has unit norm) of the context vector. Importantly, in our setting, the arm contexts are not observed i.e. the context is hidden. In theory, one could apply the results in these works to MAB-LC by associating standard basis vectors to the arms and converting it into a contextual bandit problem. However, such a trivial conversion results in a highly sub-optimal regret of $\widetilde{O}(\sqrt{\mathsf{M}^2\mathsf{CT}} + \mathsf{M}^3\mathsf{N})$ due to a strong singular value assumption (see Appendix B for a detailed discussion). In other words, the guarantees in this line of work is only useful when the dimension d is much smaller than the number of arms. Furthermore, these papers also assume that the unknown parameter vectors corresponding to the clusters are significantly separated - this makes clustering easy with a few initial rounds. Importantly, our results/algorithm do not need such a condition - Assumption 1 allows the unknown cluster parameters to be as close as possible. In a similar line of work, Zhou and Brunskill [2016] proposed an explore-then-commit style algorithm, but with sub-optimal regret bound, which in some cases is linear in T.

Online Low rank Matrix Completion (O-LRMC). This is a more general problem than MAB-LC, but the existing results are significantly sub-optimal. As mentioned earlier, Jain and Pal [2022]'s result applies to only rank-1 case. For the general rank-r case, which corresponds to our C clusters, the algorithm can be significantly suboptimal in terms of dependence on T. In a related work, Sen et al. [2017] studied an epsilon-greedy algorithm, and derived sub-optimal distribution-dependent bounds scaling inversely in the square of the gap between mean rewards. In addition, their distribution-free regret bounds have sub-

optimal dependence in T; $T^{2/3}$ instead of $\sqrt{T}$ provided by our method. Dadkhahi and Negahban [2018] provided an online alternating minimization heuristic for the general rank-r problem, but do not provide any regret bounds. In a separate line of work, [Kveton et al., 2017, Trinh et al., 2020, Katariya et al., 2017, Hao et al., 2020, Jun et al., 2019, Huang et al., 2020, Lu et al., 2021] study a similar low-rank reward matrix setting but they consider a significantly easier objective of identifying the largest entry in the entire reward matrix/tensor instead of finding the most rewarding arms for each user/agent.

Online Collaborative Filtering. A closely related line of work studies the user-based online Collaborative Filtering (CF) Bresler et al. [2014, 2016], Heckel and Ramchandran [2017], Bresler and Karzand [2019], Huleihel et al. [2021]. These works study the MAB-LC problem with the additional constraint that the same arm cannot be pulled by any particular user more than once. While this model is strictly more restricted than MAB-LC, no theoretical bounds are known on the regret in this setting. Instead, these works minimize pseudo-regret: assuming the mean rewards or arms lie in $[0, 1]$, these works aim to maximize the number of arms pulled with reward more than $1/2$. We note that this is a simpler metric than cumulative regret because maximizing the latter requires identifying the best arm, whereas maximizing the former only requires identifying arms with reward more than $1/2$.

2 PROBLEM FORMULATION

Notations: We write $[n]$ to denote the set $\{1, 2, \dots, n\}$. For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, we will write $\mathbf{A}_i, \mathbf{A}_{|i}$ to denote the ith row and column of matrix $\mathbf{A}$ respectively. We will write $\mathbf{A}_{ij}$ to denote the entry of the matrix $\mathbf{A}$ in the ith row and j^{th} column. We will write $||\mathbf{A}||_\infty = \max_{ij} |\mathbf{A}_{ij}|$ to denote the largest entry of the matrix $\mathbf{A}$. For a subset $\mathcal{S} \subseteq [m]$ of indices, we will write $\mathbf{A}_\mathcal{S}$ and $\mathbf{A}_{|\mathcal{S}}$ to denote the sub-matrix of $\mathbf{A}$ restricted to the rows in $\mathcal{S}$ and columns in $\mathcal{S}$ respectively. Extending the above notations, $\mathbf{A}_{i|\mathcal{S}}$ denotes the ith row of $\mathbf{A}$ restricted to the columns in $\mathcal{S}$; $\mathbf{A}_{\mathcal{S},\mathcal{S}'}$ denotes the sub-matrix of $\mathbf{A}$ restricted to the rows in $\mathcal{S}$ and columns in $\mathcal{S}'$. We write $\mathbf{e}_i$ to denote the ith standard basis vector that is zero everywhere except in the ith position where it has a 1. $\widetilde{O}(\cdot)$ notation hides logarithmic factors.

Consider a multi-user multi-armed bandit (MAB) problem where we have a set of M arms (denoted by the set $[M]$), N users (denoted by the set $[N]$) and T rounds. In each round, a user $u(t)$ is sampled independently from a distribution P_N over $[N]$ (for much of the paper, we assume P_N is the uniform distribution). The sampled user $u(t)$ pulls an arm $\rho(t)$ from the set $[M]$ and receives a reward $\mathbf{R}^{(t)}$, s.t.,

$$\mathbf{R}^{(t)} = \mathbf{P}_{u(t)\rho(t)} + \eta^{(t)} \tag{1}$$

where $\eta^{(t)}$ denotes the additive noise that is added to

each observation. We will assume that the noise sequence $\{\eta^{(t)}\}_{t \in [T]}$ is composed of i.i.d zero-mean sub-Gaussian random variables with variance proxy at most $\sigma^2 > 0$. Also, $\mathbf{P} \in \mathbb{R}^{N \times M}$ is the user-arm reward matrix. We study the MAB-LC problem under two assumptions on the reward matrix $\mathbf{P}$.

Cluster Structure (CS). Here, we assume the set of N users can be partitioned into C unknown clusters $\mathcal{C}^{(1)}, \mathcal{C}^{(2)}, \dots, \mathcal{C}^{(C)}$. Furthermore, in a particular cluster $\mathcal{C}^{(i)}$ for any $i \in [C]$, each user $u \in \mathcal{C}^{(i)}$ has an identical reward vector $\mathbf{P}_u$. Let $\mathbf{X} \in \mathbb{R}^{C \times M}$ be the sub-matrix of $\mathbf{P}$ that has the distinct rows of $\mathbf{P}$ corresponding to each cluster. $\tau := \max_{i,j \in [C]} |\mathcal{C}^{(i)}|/|\mathcal{C}^{(j)}|$ denotes the ratio of the maximum and minimum cluster size. For each user u, $\pi_u : [M] \to [M]$ denotes a permutation that sorts the arms in descending order of their reward for user u, i.e., $\mathbf{P}_{u\pi_u(i)} \geq \mathbf{P}_{u\pi_u(j)}$ for $i \leq j$. Also, $\pi_u(1) \triangleq \text{argmax}_{j \in [M]} \mathbf{P}_{uj}$ is the arm with the highest reward for user $u \in [N]$.

Relaxed Cluster Structure (RCS): Here, we relax the cluster definition so that the users in the same cluster might not have identical reward vectors. That is, the assumption is that the set of N users can be partitioned into C clusters $\mathcal{C}^{(1)}, \mathcal{C}^{(2)}, \dots, \mathcal{C}^{(C)}$ s.t. the following holds for some known $\nu > 0$: 1) For any two users u, v in the *same cluster*, $\pi_u(1) = \pi_v(1)$ and $||\mathbf{P}_u - \mathbf{P}_v||_\infty \leq \nu$, 2) For any two users u, v in different clusters, we will have either $|\mathbf{P}_{u\pi_u(1)} - \mathbf{P}_{v\pi_u(1)}| > 20\nu$ or $|\mathbf{P}_{u\pi_v(1)} - \mathbf{P}_{v\pi_v(1)}| > 20\nu$. Note that the CS structure is a special case of RCS with $\nu = 0$.

Remark 1. *The constant 20 in* RCS *model formulation is arbitrary and can be replaced by any constant* > 1.

Thus, in the RCS model formulation, users in the same cluster have the same best arm and the reward vectors are entry-wise close; users in different cluster have rewards corresponding to one of the best arms to be separated.

Now, the goal is to minimize the regret assuming either CS or RCS structure on the reward matrix $\mathbf{P}$:

$$\text{Reg}(T) \triangleq \mathbb{E}\left(\sum_{t \in [T]} \mathbf{P}_{u(t)\pi_{u(t)}(1)} - \sum_{t \in [T]} \mathbf{P}_{u(t)\rho(t)}\right). \tag{2}$$

Here the expectation is over the randomness in the algorithm and the sampled users.

Remark 2. *Note that a trivial approach is to treat each user as a separate multi-armed bandit problem. Such a strategy does not utilize the low rank structure and leads to a regret of* $O(\sqrt{MNT})$ *assuming* $||\mathbf{P}||_\infty, \sigma = O(1)$. *Another trivial approach is to recommend random arms to each user (exploration) and subsequently use low rank matrix completion guarantees to estimate* $\mathbf{P}$ *and exploit. This will lead to a regret guarantee of* $O((M + N)^{1/3}T^{2/3})$ *[Jain and Pal, 2022]. The goal is to obtain a significantly smaller regret guarantee of* $O(\sqrt{(M + N)T})$ *with* $C = O(1)$.

3 PRELIMINARIES

As mentioned earlier, the key algorithmic tool that we use is low rank matrix completion - a statistical estimation problem where the goal is to recover a low rank matrix from partially observed randomly sampled entries. Since the reward matrix $\mathbf{P}$ in MAB-LC is low rank, our strategy is to call the offline matrix completion oracle for relevant sub-matrices of $\mathbf{P}$ after we accumulate a sufficient number of random observations in each sub-matrix. In this work, we are interested in low rank matrix completion with non-trivial entry-wise guarantees that has been studied recently in Chen et al. [2019], Abbe et al. [2020]. Below, we state a low rank matrix completion result that is adapted from Jain and Pal [2022] which is in turn obtained with minor modifications from Chen et al. [2019][Theorem 1] and is more suited to our setting:

Lemma 1 (Lemma 2 in Jain and Pal [2022]). *Consider rank* $\mathsf{C} = O(1)$ *reward matrix* $\mathbf{P} \in \mathbb{R}^{\mathsf{N} \times \mathsf{M}}$ *with SVD decomposition* $\mathbf{P} = \bar{\mathbf{U}}\boldsymbol{\Sigma}\bar{\mathbf{V}}^{\mathsf{T}}$ *satisfying* $\|\bar{\mathbf{U}}\|_{2,\infty} \leq \sqrt{\mu\mathsf{C}/\mathsf{N}}, \|\bar{\mathbf{V}}\|_{2,\infty} \leq \sqrt{\mu\mathsf{C}/\mathsf{M}}$ *and condition number* $\kappa = O(1)$. *Let* $d_1 = \max(\mathsf{N}, \mathsf{M})$, $d_2 = \min(\mathsf{N}, \mathsf{M})$, *and let* p *be such that* $1 \geq p \geq c\mu^2 d_2^{-1} \log^3 d_1$ *for some constant* $c > 0$. *For any positive integer* $s > 0$ *satisfying* $\frac{\sigma}{\sqrt{s}} = O\left(\sqrt{\frac{pd_2}{\mu^3 \log d_2}}\|\mathbf{P}\|_\infty\right)$, *Algorithm 5 with input* s, p, σ *that uses* $m = \widetilde{O}\left(sp\mathsf{MN}\right)$ *observations to output* $\widehat{\mathbf{P}} \in \mathbb{R}^{\mathsf{N} \times \mathsf{M}}$ *for which, with probability at least* $1 - \widetilde{O}(\delta)$, *we have*

$$\|\mathbf{P} - \widehat{\mathbf{P}}\|_\infty = O\left(\frac{\sigma\sqrt{\mu^3 \log d_1}}{\sqrt{spd_2}}\right) = \widetilde{O}\left(\frac{\sigma\sqrt{\mathsf{MN}\mu^3}}{\sqrt{md_2}}\right).$$

We now introduce a definition characterizing a *nice* subset of users that we often use in the analysis.

Definition 1. *A subset of users* $\mathcal{S} \subseteq [\mathsf{N}]$ *will be called "nice" if* $\mathcal{S} \equiv \bigcup_{j \in \mathcal{A}} \mathcal{C}^{(j)}$ *for some* $\mathcal{A} \subseteq [\mathsf{C}]$. *In other words,* $\mathcal{S}$ *can be represented as the union of some subset of clusters.*

4 LATTICE ALGORITHM FOR CS

4.1 Algorithm and Proof Overview

LATTICE runs in phases of exponentially increasing length. In each phase, the goal is to divide the set of users into *nice* subsets. Moreover, for each subset of users, we have an active subset of arms that must contain the best arm for all users in the corresponding subset. That is, at the start of ℓth phase, we aim to create a list (of size $a_\ell \leq \mathsf{C}$) of *nice* subsets of users $\mathcal{M}^{(\ell)} \equiv \{\mathcal{M}^{(\ell,i)}\}_{i \in [a_\ell]}$ and the corresponding subsets of arms $\mathcal{N}^{(\ell)} \equiv \{\mathcal{N}^{(\ell,i)}\}_{i \in [a_\ell]}$, s.t. $\cup_{i \in [a_\ell]} \mathcal{M}^{(\ell,i)} = [\mathsf{N}], \mathcal{N}^{(\ell,i)} \supseteq \{\pi_u(1) \mid u \in \mathcal{M}^{(\ell,i)}\}$, and

$$\left|\mathbf{P}_{u\pi_u(1)} - \min_{j \in \mathcal{N}^{(\ell,i)}} \mathbf{P}_{uj}\right| \leq \epsilon_\ell, \ \forall \ u \in \mathcal{M}^{(\ell,i)}. \quad (3)$$

Above, ϵ_ℓ is a fixed exponentially decreasing sequence in ℓ. As we eliminate arms at each phase, the number of user subsets with more than $\gamma\mathsf{C}$ active arms goes on shrinking with each phase. Since LATTICE is random, we define event $\mathcal{E}^{(\ell)}$ to be true if LATTICE maintains a list of user subsets and arm subsets satisfying the above properties.

Now, in round t, the sampled user $u(t)$ pulls arm $\rho(t)$ where $\rho(t)$ is sampled from $\mathcal{N}^{(\ell,i)}$ assuming $|\mathcal{N}^{(\ell,i)}| \geq \gamma\mathsf{C}$, where i is the index of subset $\mathcal{M}^{(\ell,i)}$ to which $u(t)$ belongs. If, $|\mathcal{N}^{(\ell,i)}| < \gamma\mathsf{C}$, then the cluster structure is ignored and arm $\rho(t)$ is selected from the active set of arms ($\mathcal{N}^{(\ell,i)}$) as determined by the Upper Confidence Bound (UCB) algorithm. Conditioned on $\mathcal{E}^{(\ell)}$ being true, our goal is to ensure $\mathcal{E}^{(\ell+1)}$ with high probability. Due to the arm pull strategy described above, for each subset of users $\mathcal{M}^{(\ell,i)}$ in $\mathcal{M}^{(\ell)}$ and their corresponding subset of active arms $\mathcal{N}^{(\ell,i)}$ (such that $|\mathcal{N}^{(\ell,i)}| \geq \gamma\mathsf{C}$) we observe random noisy entries from the sub-matrix $\mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$. Subsequently, we use low rank matrix completion (Step 4 in Alg. 1) and Lemma 1 to obtain $\widetilde{\mathbf{P}} \in \mathbb{R}^{\mathsf{N} \times \mathsf{M}}$ such that $\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$ is an entry-wise good estimate of $\mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$, i.e.,

$$\left\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\right\|_\infty \leq \Delta_{\ell+1} \quad (4)$$

with high probability where $\Delta_{\ell+1} = \epsilon_{\ell+1}/40\mathsf{C}$. We define the event $\mathcal{E}_2^{(\ell)}$ to be true if eq. (4) is satisfied for all relevant sub-matrices.

Next, conditioning on $\mathcal{E}_2^{(\ell)}$, consider a subset of users $\mathcal{M}^{(\ell,i)}$ for which the corresponding subset of active arms is large $|\mathcal{N}^{(\ell,i)}| \geq \gamma\mathsf{C}$. For next phase, the intuitive goal is to further partition $\mathcal{M}^{(\ell,i)}$ into subsets $\{\mathcal{M}^{(\ell,i,j)}\}_j$, each of which is *nice* and find a list of corresponding subsets of active arms $\{\mathcal{N}^{(\ell,i,j)}\}_j \subseteq \mathcal{N}^{(\ell,i)}$ such that all arms in $\mathcal{N}^{(\ell,i,j)}$ have high reward (as in eq. 3) for all users in $\mathcal{M}^{(\ell,i,j)}$. To do so, for each user in the set $\mathcal{M}^{(\ell,i)}$, we find a subset of *good arms* among the active arms $\mathcal{T}_u^{(\ell)} \subseteq \mathcal{N}^{(\ell,i)}$ such that

$$\mathcal{T}_u^{(\ell)} \equiv \{j \in \mathcal{N}^{(\ell,i)} \mid \max_{j' \in \mathcal{N}^{(\ell,i)}} \widetilde{\mathbf{P}}_{uj'}^{(\ell)} - \widetilde{\mathbf{P}}_{uj}^{(\ell)} \leq 2\Delta_{\ell+1}\} \quad (5)$$

i.e. arms which have a high estimated reward for user u.

Subsequently, we design a graph whose nodes are users in $\mathcal{M}^{(\ell,i)}$ and an edge is drawn between two users $u, v \in \mathcal{M}^{(\ell,i)}$ if the following conditions are satisfied:

$$\mathcal{T}_u^{(\ell)} \cap \mathcal{T}_v^{(\ell)} \neq \Phi, \left|\widetilde{\mathbf{P}}_{ux}^{(\ell)} - \widetilde{\mathbf{P}}_{vx}^{(\ell)}\right| \leq 2\Delta_{\ell+1} \forall x \in \mathcal{N}^{(\ell,i)} \quad (6)$$

In other words, eq. (6) defines an edge between two users in the same subset if reward estimates of active arms for the two users are close; secondly, there are common arms in their respective set of *good arms* as defined in eq. (5). We partition the set of users $\mathcal{M}^{(\ell,i)}$ into smaller *nice* sets $\{\mathcal{M}^{(\ell,i,j)}\}$ by considering the connected components of the aforementioned graph and for users in each component

$\mathcal{M}^{(\ell,i,j)}$, the updated trimmed common subset of arms

$$\mathcal{N}^{(\ell,i,j)} \equiv \bigcup_{u \in \mathcal{M}^{(\ell,i,j)}} \mathcal{T}_u^{(\ell)} \qquad (7)$$

with high reward is the union of set of good arms for all users in the connected component (see Step 8 in Alg. 1). We can show the following crucial and interesting lemma:

Lemma. *Fix any $i \in [a_\ell]$ such that $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma \mathsf{C}$. Consider two users $u, v \in \mathcal{M}^{(\ell,i)}$ having a path in the constructed graph. Conditioned on $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, we have*

$$\max_{x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_v^{(\ell)}} |\mathbf{P}_{ux} - \mathbf{P}_{uy}| \leq 32\mathsf{C}\Delta_{\ell+1}$$

and $\max_{x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_v^{(\ell)}} |\mathbf{P}_{vx} - \mathbf{P}_{vy}| \leq 32\mathsf{C}\Delta_{\ell+1}.$

This lemma shows that good arms for one user is good for another if they are connected by a path. If the number of active arms for a subset of users become less than $\gamma \mathsf{C}$, then we start UCB (Upper Confidence Bound Lattimore and Szepesvári [2020][Ch. 7]) for each user in that subset with the active arms until end of algorithm. Therefore, conditioned on $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, the event $\mathcal{E}^{(\ell+1)}$ is true w.h.p. Hence, conditioned on $\mathcal{E}^{(\ell)}$, we can bound the regret in each round of the ℓth phase by ϵ_ℓ; roughly speaking, the number of rounds in the ℓth phase is $1/\epsilon_\ell^2$ and therefore the regret is $1/\epsilon_\ell$. By setting Δ_ℓ as in Step 3 of Alg. 1 (and $\epsilon_\ell = \Delta_\ell/64\mathsf{C}$), we can bound the regret of LATTICE and achieve the guarantee in Theorems 1 and 2.

Remark 3. *The low rank matrix completion oracle is obtained with slight modifications from Jain and Pal [2022]. In line 6 of Algorithm 1, we require the matrix completion oracle (Algorithm 5) to be stateful (wait appropriately). This is because vanilla low rank matrix completion results work under the assumption of Bernoulli sampling i.e. each entry in the matrix is observed once with some probability p. So to mimic Bernoulli sampling, we require a stateful version of the algorithm. We create a Bernoulli mask at the beginning of the phase and pull arms to observe only the masked entries in sequence. We also make multiple observations corresponding to the same mask and take the average in each of the mask indices to reduce the variance; similarly, we also compute several estimates of the same matrix with independently sampled mask and take an entry-wise median to reduce error probability. Using these tricks appropriately (see Appendix D for a detailed proof and discussion) can allow us to obtain the guarantee in Lemma 1. In between subsequent arrivals of users belonging to a cluster, the algorithm remains stateful and waits at line 8.*

4.2 Theoretical Guarantees

To obtain regret bounds, we first make the following assumptions on the matrix $\mathbf{X} \in \mathbb{R}^{\mathsf{C} \times \mathsf{M}}$ whose rows correspond to cluster reward vectors in the CS setting:

Assumption 1 (Assumptions on $\mathbf{X}$). *Let $\mathbf{X} = \mathbf{U\Sigma V}^{\mathsf{T}}$ be the SVD of $\mathbf{X}$. Also, let $\mathbf{X}$ satisfy the following: 1) Condition number: $\mathbf{X}$ is full-rank and has non zero singular values $\lambda_1 > \cdots > \lambda_{\mathsf{C}}$ with condition number $\lambda_1/\lambda_{\mathsf{C}} = O(1)$, 2) μ-incoherence: $||\mathbf{V}||_{2,\infty} \leq \sqrt{\mu \mathsf{C}/\mathsf{M}}$, 3) Subset Strong Smoothness: For some $\alpha = O(1)$, $\gamma = \widetilde{O}(1)$, for any subset of indices $\mathcal{S} \subseteq [\mathsf{M}], |\mathcal{S}| = \gamma\mathsf{C}$, we must have $\mathbf{x}^{\mathsf{T}} \mathbf{V}_{\mathcal{S}}^{\mathsf{T}} \mathbf{V}_{\mathcal{S}} \mathbf{x} \geq \alpha\gamma\mathsf{C}/\mathsf{M}$ for all unit norm vectors $\mathbf{x} \in \mathbb{R}^{\mathsf{C}}$.*

Feasibility of Assumption 1. Assuming $\mathsf{C} \ll \mathsf{M}$, one can show that Assumption 1 is satisfied by a large subset of matrices. That is, suppose each entry of $\mathbf{X}$ is generated according to $\mathcal{N}(0,1)$. Then, it is easy to show that $\mathbf{X}$ satisfies Assumption 1 (see Appendix).

Justification of Assumption 1. Note that the assumption is similar to the assumptions required by standard low-rank matrix completion methods Candès and Recht [2009], Bhojanapalli and Jain [2014]. Intuitively, these assumptions are necessary to obtain small regret because to get small regret we require an arm pull for i-th user to provide good information for j-th user. That is, the matrix $\mathbf{X}$ should have information "well-spread" out instead of information being concentrated in a few entries or in a few directions only. To see this, consider an extreme example (when these assumptions are not satisfied) when $\mathbf{X} = [\mathbf{I}_{\mathsf{C} \times \mathsf{C}} \ \ \mathbf{0}]$. In that case, most of the arms will give no information when pulled; all the arms need to be sampled for all users to get a good estimate of the reward matrix. Further exploration into necessity of Assumption 1 is left for future work.

Assumption 2. *We will assume that $\tau, \mathsf{C} = O(1)$ and does not scale with the number of rounds T.*

Note that the above assumption is just for simplicity of exposition. Our algorithm is indeed polynomial in C and τ, so we can incorporate more general τ and C. But for simplicity, we ignore these factors by assuming them to be constants.

Next, we characterize some properties namely the condition number and incoherence of sub-matrices of $\mathbf{P}$ restricted to a *nice* subset of users in the CS setting

Lemma 2. *Suppose Assumption 1 is true. Consider a sub-matrix $\mathbf{P}_{\mathsf{sub}}$ of $\mathbf{P}$ having non-zero singular values $\lambda_1' > \cdots > \lambda_{\mathsf{C}'}'$, (for $\mathsf{C}' \leq \mathsf{C}$). Then, if the rows of $\mathbf{P}_{\mathsf{sub}}$ correspond to a nice subset of users, we have $\frac{\lambda_1'}{\lambda_{\mathsf{C}'}'} \leq \frac{\lambda_1}{\lambda_{\mathsf{C}}} \sqrt{\tau}$.*

Lemma 3. *Suppose Assumption 1 is true. Consider a sub-matrix $\mathbf{P}_{\mathsf{sub}} \in \mathbb{R}^{\mathsf{N}' \times \mathsf{M}'}$ (with SVD decomposition $\mathbf{P}_{\mathsf{sub}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}$) of $\mathbf{P}$ whose rows correspond to a nice subset of users. Then, $\left|\left|\widetilde{\mathbf{U}}\right|\right|_{2,\infty} \leq \sqrt{\frac{\mathsf{C}\tau}{\mathsf{N}'}}$ and $\left|\left|\widetilde{\mathbf{V}}\right|\right|_{2,\infty} \leq \sqrt{\frac{\mu\mathsf{C}}{\alpha\mathsf{M}'}}$.*

Lemmas 2 and 3 allow us to apply low rank matrix completion (Lemma 1) to relevant sub-matrices of the reward matrix $\mathbf{P}$. Now, we are ready to present our main theorem:

Theorem 1. *Consider the MAB-LC problem in CS frame-*

Algorithm 1 LATTICE (Latent bAndiTs via maTrIx ComplEtion)

Require: Number of users N, arms M, clusters C, rounds T, noise $\sigma^2 > 0$, $\gamma \geq 1$ and constant $C' \geq 0$.

1: Set $\mathcal{M}^{(1)} \equiv [\mathcal{M}^{(1,1)}]$ where $\mathcal{M}^{(1,1)} = [N]$ and $\mathcal{N}^{(1)} \equiv [\mathcal{N}^{(1,1)}]$ where $\mathcal{N}^{(1,1)} = [M]$. Set round index t to be a global parameter; initialize $t = 0$.

2: **for** $\ell = 1, 2, \ldots,$ **do**

3:     Set $\Delta_{\ell+1} = C'2^{-\ell}$ for some appropriate $C' > 0$. Initialize $\mathcal{M}^{(\ell+1)} = []$ and $\mathcal{N}^{(\ell+1)} = []$ to be empty lists.

4:     Collect data by running Alg. 2 i.e. Alg. COLLECT DATA AND COMPUTE ESTIMATE$(C, \gamma, \sigma^2, \Delta_{\ell+1}, \mathcal{M}^{(\ell)}, \mathcal{N}^{(\ell)}, T)$ for the ℓth phase. Subsequently, for $\mathcal{M}^{(\ell,i)} \in \mathcal{M}^{(\ell)}$ such that $|\mathcal{N}^{(\ell,i)}| \geq \gamma C$, compute an estimate $\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}}$ of matrix $\mathbf{P}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}}$ from the data such that with high probability $\left\| \widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}} - \mathbf{P}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}} \right\|_{\infty} \leq \Delta_{\ell+1}$.

5:     **for** $i : |\mathcal{N}^{(\ell,i)}| \geq \gamma C$ **do**

6:         For every user $u \in \mathcal{M}^{(\ell,i)}$, compute $\mathcal{T}_u^{(\ell)} \equiv \{j \in \mathcal{N}^{(\ell,i)} \mid \max_{j' \in \mathcal{N}^{(\ell,i)}} \widetilde{\mathbf{P}}_{uj'}^{(\ell)} - \widetilde{\mathbf{P}}_{uj}^{(\ell)} \leq 2\Delta_{\ell+1}\}$.

7:         Construct graph $\mathcal{G}^{(\ell,i)}$ whose nodes are users in $\mathcal{M}^{(\ell,i)}$ and an edge exists between two users $u, v \in \mathcal{M}^{(\ell,i)}$ if $\mathcal{T}_u^{(\ell)} \cap \mathcal{T}_v^{(\ell)} \neq \Phi$ and $\left| \widetilde{\mathbf{P}}_{ux}^{(\ell)} - \widetilde{\mathbf{P}}_{vx}^{(\ell)} \right| \leq 2\Delta_{\ell+1}$ for all arms $x \in \mathcal{N}^{(\ell,i)}$.

8:         For each connected component $\mathcal{M}^{(\ell,i,j)}$ $(\cup_j \mathcal{M}^{(\ell,i,j)} \equiv \mathcal{M}^{(\ell,i)})$, compute $\mathcal{N}^{(\ell,i,j)} \equiv \cup_{u \in \mathcal{M}^{(\ell,i,j)}} \mathcal{T}_u^{(\ell)}$. Append $\mathcal{M}^{(\ell,i,j)}$ into $\mathcal{M}^{(\ell+1)}$ and $\mathcal{N}^{(\ell,i,j)}$ into $\mathcal{N}^{(\ell+1)}$.

9:     **end for**

10:     For each pair of sets $(\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)})$ such that $|\mathcal{N}^{(\ell,i)}| \leq \gamma C$, append $\mathcal{M}^{(\ell,i)}$ to $\mathcal{M}^{(\ell+1)}$ and $\mathcal{N}^{(\ell,i)}$ to $\mathcal{N}^{(\ell+1)}$.

11: **end for**

---

work with M arms, N users, C clusters and T rounds such that at every round $t \in [T]$, we observe reward $\mathbf{R}^{(t)}$ as defined in eq. (1) with noise variance proxy $\sigma^2 > 0$. Let $\mathbf{P} \in \mathbb{R}^{N \times M}$ be the expected reward matrix and $\mathbf{X} \in \mathbb{R}^{C \times M}$ be the sub-matrix of $\mathbf{P}$ with distinct rows. Suppose Assumption 1 is satisfied by $\mathbf{X}$ and Assumption 2 is true. Then Alg. 1 with $C' = cC^{-1} \min \left(\|\mathbf{P}\|_{\infty}, \frac{\sigma\sqrt{\mu}}{\log M} \right)$ for some appropriate constant $c > 0$ guarantees the regret $\text{Reg}(T)$ to be:

$$\widetilde{O}(\sigma\sqrt{\mu^3 T(M + N)}) + \sigma\sqrt{NT}).$$

To better understand the theorem, let's remove the scaling factors. Dividing the regret by σ gives us a scale-free regret of $\sqrt{T(M + N)}$. Now, even if we know the clustering structure apriori, the regret would be $\sqrt{TM}$, so we are only paying an additive factor of $\sqrt{NT}$ for the latent cluster structure, which is tight.

Also, dividing the the regret by N we get per-user regret of $\sqrt{\left(1 + \frac{M}{N}\right)}\sqrt{T'}$ where $T' = T/N$ is the average number of arm-pulls for each user. That is, when $N \gg M$, we need only $\log(M + N)$ arm pulls to get reasonable estimate. Hence, per user, the number of arms that needs to be pulled decreases exponentially from $\Omega(M)$ to $\log(M + N)$. On the other hand, when the number of users is small, each user has to explore at least M/N arms to collaboratively provide information about all the arms. This also matches the intuition, especially when the number of users is 1 where the bound matches the standard single-user MAB bound.

We would like to add the following two remarks:

Remark 4 (Generalization). *Our results can be generalized to the setting when the users are sampled according to a known non-uniform distribution in different ways. We can simulate the uniform distribution in each phase of the Alg. 1 by ignoring several observations; this approach is disadvantageous since users with very low probability of getting sampled will increase the number of sufficient observations significantly. Another approach is to partition the set of users into disjoint buckets such that the probabilities of getting sampled for users in the same bucket are within a factor of 2 of each other. Now, in each bucket, we can run Alg. 1 separately and simulate the uniform distribution in each phase. Since the number of buckets will be logarithmic in $[N]$, the regret remains same up to logarithmic factors.*

Remark 5 (Generalization Continued). *We can generalize our results to the setting where $\tau, C, \kappa = \lambda_1/\lambda_C$ scales with the number of rounds T by modifying Thm. 2 in Chen et al. [2019] appropriately. This will lead to the first term of regret guarantee in Thm. 1 being $\sqrt{VT}$ where $V = \widetilde{O}\left(\text{poly}(r, \tau, C, \kappa)\left(N + M\right)\right)$; hence we will have a $\text{poly}(r, \tau, C, \kappa)$ additional multiplicative factor in the regret. See Appendix H for details on this generalization.*

In the CS framework, we can also provide instance-dependent regret bounds that are sharper than worst case guarantees in Theorem 1. Let us introduce some definitions: for every cluster $c \in [C]$, define the subset of arms $\mathcal{G}_{c,\ell}$ for all users $u \in \mathcal{C}^{(c)}$ and for all $\ell > 1$ as

$$\mathcal{G}_{c,\ell} \equiv \{j \in [M] \mid \epsilon_{\ell} \leq \left| \mathbf{P}_{uj} - \mathbf{P}_{u\pi_u(1)} \right| \leq \epsilon_{\ell-1}\}$$

and $\mathcal{G}_{c,1} \equiv \{j \in [M] \mid \epsilon_1 \leq \left| \mathbf{P}_{uj} - \mathbf{P}_{u\pi_u(1)} \right|\}$ for $\ell = 1$; $\mathcal{G}_{c,\ell}$ ($\mathcal{G}_{c,1}$) corresponds to the subset of arms having a sub-optimality gap that is between $\epsilon_{\ell-1}$ and ϵ_{ℓ} (greater than ϵ_1 respectively) for all users belonging to the cluster $\mathcal{C}^{(c)}$. There is no ambiguity in the definition since all users in the

---

Algorithm 2 COLLECT DATA AND COMPUTE ESTIMATE

Require: Number of clusters C, parameter $\gamma > 0$, noise variance $\sigma^2 > 0$, desired error guarantee $\Delta_{\ell+1} > 0$, partition of users $\mathcal{M}^{(\ell)}$ comprising *nice* subsets of users, corresponding list of sets of arms $\mathcal{N}^{(\ell)}$, rounds T.

1: **for** the ith set of users $\mathcal{M}^{(\ell,i)} \in \mathcal{M}^{(\ell)}$ and i^{th} set of arms $\mathcal{N}^{(\ell,i)} \in \mathcal{N}^{(\ell)}$ **do**
2:     **if** $\mathcal{N}^{(\ell,i)} \geq \gamma C$ **then**
3:         Create a stateful instance of Algorithm LOW RANK MATRIX ESTIMATE with parameters - { $\mathcal{V} = \mathcal{N}^{(\ell,i)}$ for users, $\mathcal{U} = \mathcal{M}^{(\ell,i)}$ (whenever a user $u \in \mathcal{U}$ is sampled to pull an arm), desired error $\Delta_{\ell+1}$ in estimate, noise σ^2 and total rounds T}. {All instances of Algorithm 5 will wait until entire data is collected (Line 5 in Alg. 6) or finish.}
4:     **else**
5:         For *each* $u \in \mathcal{M}^{(\ell,i)}$, if not created already, create a stateful instance of Algorithm UCB(u) with arm set $\mathcal{N}^{(\ell,i)}$
6:     **end if**
7: **end for**
8: **while** There exists an instance of Algorithm 5 that waits (i.e. Alg. 6 invoked by Alg. 5 waits at Line 5.) **do**
9:     Sample a user $u(t)$ from the environment. Determine $i_t \leftarrow \{i : u(t) \in \mathcal{M}^{(\ell,i)}\}$.
10:     **if** Algorithm 5's instance with $\mathcal{U} = \mathcal{M}^{(\ell,i_t)}$ is still waiting **then**
11:         Execute one step of the DATACOLLECTIONSUBROUTINE (Line 5 of Alg. 6 with $u(t)$) invoked by Algorithm 5's instance (with $\mathcal{U} = \mathcal{M}^{(\ell,i_t)}$)
12:     **else if** Instance of Algorithm UCB$(u(t))$ (i.e. Alg. 7) exists **then**
13:         Invoke one time step of Algorithm UCB$(u(t))$.
14:     **else**
15:         pull a random arm from $\mathcal{N}^{(\ell,i_t)}$.
16:     **end if**
17:     $t \leftarrow t + 1$. Stop the Algorithm when $t = T$.
18: **end while**
19: **for** the ith set of users $\mathcal{M}^{(\ell,i)} \in \mathcal{M}^{(\ell)}$ and i^{th} set of arms $\mathcal{N}^{(\ell,i)} \in \mathcal{N}^{(\ell)}$ **do**
20:     **if** $\mathcal{N}^{(\ell,i)} \geq \gamma C$ **then**
21:         Store estimate $\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$ obtained as output from Algorithm LOW RANK MATRIX ESTIMATE for users $\mathcal{U} = \mathcal{M}^{(\ell,i)}$ and arms $\mathcal{V} = \mathcal{N}^{(\ell,i)}$.
22:     **end if**
23: **end for**
24: Return all stored estimates $\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$ for $i : \left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma C$.

---

same cluster $\mathcal{C}^{(c)}$ have the same mean rewards over all arms. Let us also define $\mathcal{H}_c \equiv \bigcup_{\ell > 1} \operatorname{argmin}_{j \in \mathcal{G}_{c,\ell}} \left| \mathbf{P}_{uj} - \mathbf{P}_{u\pi_u(1)} \right|$ with the understanding that whenever $\mathcal{G}_{c,\ell} = \Phi$, there is no argmin to be counted in the set. For brevity of notation, let $\Psi_{c,a} \triangleq \mathbf{P}_{u\pi_u(1)} - \mathbf{P}_{ua}$ be the sub-optimality gap in the reward of arm a for any user u in cluster c.

Theorem 2. *Consider the setting in Theorem 1 and the sets $\{\mathcal{G}_{c,\ell}, \mathcal{H}_c\}_{c,\ell}$ as defined above. Then Alg. 1 with $C' = cC^{-1} \min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log M}\right)$ for some appropriate constant $c > 0$ guarantees the regret* $\operatorname{Reg}(T)$ *to be*

$$\widetilde{O}\left(\frac{\|\mathbf{P}\|_\infty \mathsf{V}}{C'^2}\mathbf{1}[\mathcal{G}_{c,1} \neq \emptyset] + \sum_{c \in [C], a \in \mathcal{H}_c} \frac{\Psi_{c,a}T^{-2}}{C} + \frac{C\mathsf{V}}{\Psi_{c,a}}\right)$$

$$+ \mathsf{N}^{-1}\widetilde{O}\left(\sum_{\substack{c \in [C] \\ a \in \{\pi_c(s)\}_{s=1}^{\gamma C}}} \left|\mathcal{C}^{(c)}\right|\left(\frac{\sigma}{\Psi_{c,a}} + 3\Psi_{c,a}\right)\right).$$

where $\mathsf{V} = \widetilde{O}\left(\sigma^2\mu^3\left(\mathsf{N} + \mathsf{M}\right)\right)$, $\{\pi_c(s)\}_{s=1}^{\gamma C}$ *are the* γC *arms for cluster c with the smallest sub-optimality gap.*

Loosely speaking, the regret bound in Thm. 2 scales as

$\widetilde{O}((\mathsf{M} + \mathsf{N})/\Psi)$ where Ψ is the minimum sub-optimality gap across all the N users involved. This is because arms with large sub-optimality gaps are quickly eliminated by LATTICE in the initial phases itself; therefore if most competing arms for a user has large sub-optimality gaps, then the user will end up pulling high reward arms more often . Again, this guarantee improves over the $\widetilde{O}((\mathsf{MN})/\Psi)$ regret trivially obtained without collaboration across users.

4.2.1 Lower Bounds

Theorem 3 (Distribution-free). *Let $C \leq \min\{\mathsf{M}, \mathsf{N}\}$. Suppose the distributions of arm rewards are Bernoulli and suppose the user at round t is sampled independently from a distribution P_N. Moreover, suppose the weighted fraction of users in the ith cluster is τ_i. Let $\sup$ be the supremum over all problem instances and $\inf$ be the infimum over all algorithms with knowledge of $\mathsf{M}, \mathsf{N}, C$. Then*

$$\inf \sup \operatorname{Reg}(T) \geq 0.02(R_1 + R_2),$$

where $R_1 = \sum_{n \in [\mathsf{N}]} \mathbb{E}\left[\min\{\sqrt{CT_n}, \mathsf{T}_n\}\right]$, $R_2 = \sum_{c \in [C]} \mathbb{E}\left[\min\{\sqrt{\mathsf{MT}_c}, \mathsf{T}_c\}\right]$. *Here,* $\mathsf{T}_n \sim Bin(\mathsf{T}, P_\mathsf{N}(n))$,

---

Algorithm 3 LATTICE ALGORITHM FOR GCS

Require: Number of users N, arms M, clusters C, rounds T, noise σ^2, separation $\nu > 0$, parameters $\gamma \geq 1$, $C' \geq 0$.
1: Initialize as in Step 1 of Alg. 1.
2: **for** $\ell = 1, 2, \ldots,$ **do**
3:    Run Steps 3-4 as in Algorithm 1.
4:    **if** $\Delta_{\ell+1} \geq 2\nu$ and $\left|\mathcal{M}^{(\ell)}\right| < C$ **then**
5:       Run Steps 5-9 as in Algorithm 1
6:    **else**
7:       For each pair of sets $(\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)})$ s.t. $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma C$, append $\mathcal{M}^{(\ell,i)}$ to $\mathcal{M}^{(\ell+1)}$ and its corresponding set of arms $\bigcap_{u \in \mathcal{M}^{(\ell,i)}} \mathcal{T}_u^{(\ell)}$ to $\mathcal{N}^{(\ell+1)}$.
8:    **end if**
9:    Run Step 10 in Algorithm 1.
10: **end for**

$T_c \sim Bin(T, \tau_c)$ *are binomial random variables.*

We now specialize the above result to the case where P_N is a uniform distribution, and the cluster sizes are uniform.

Corollary 1. *Consider the setting of Theorem 3. Suppose* $T \geq 10(M + N)C$, P_N *is the uniform distribution, and suppose each cluster has the same number of users. Then* $\inf \sup \text{Reg}(T) = \Omega(\sqrt{(M + N)CT})$

Together with Theorem 1, the above result shows that LAT-TICE achieves minimax optimal regret when $C = O(1)$, and the reward matrix $\mathbf{X}$ satisfies the incoherence condition.

Theorem 4 (Distribution-dependent)**.** *Consider the setting of Theorem 3. Suppose there is a unique best arm for each cluster. Moreover, suppose our algorithm is uniformly efficient, i.e., for any sub-optimal arm a of any user u, $\mathbb{E}[N_{a,u}(T)] = o(T^{\alpha})$ for all $\alpha \in (0, 1)$. Then*

$$\lim_{T \to \infty} \frac{\text{Reg}(T)}{\log T} \geq \sum_{c \in [C]} \sum_{a} \frac{\mathbf{P}_{u(c)\pi_{u(c)}(1)}(1 - \mathbf{P}_{u(c)\pi_{u(c)}(1)})}{\Psi_{c,a}},$$

where the inner summation is over the set of all sub-optimal arms in cluster c. Here, $u(c)$ is any user in cluster c, and $\mathbf{P}_{u(c)\pi_{u(c)}(1)}$ is the mean reward of the best arm in cluster c, and $\Psi_{c,a} = \mathbf{P}_{u(c)\pi_{u(c)}(1)} - \mathbf{P}_{u(c)a}$.

The lower bound in Theorem 4 can be tightened a bit more, albeit at the expense of readability. We provide this improved bound in the Appendix.

5 LATTICE ALGORITHM FOR RCS

LATTICE for RCS (Alg. 3) is very similar to Alg. 1 with the main novelty being cluster-wise elimination of arms in Steps 7 that needs a more aggressive approach. In essence, Alg. 3 has three components:

- **Joint Arm Elimination:** As in Algorithm 1, we run a phased algorithm where in the ℓth phase, we maintain a partition of users $\mathcal{M}^{(\ell)}$ and a family of subsets of active arms $\mathcal{N}^{(\ell)}$ having a one-to-one mapping. For any set of users in $\mathcal{M}^{(\ell)}$ that has more than γC active arms, we use Matrix Completion techniques to jointly shrink their set of active arms and partition them even further. We stop this component if we end up with C groups of users for the first time or if $\Delta_{\ell+1} \leq 2\nu$. In essence, in each phase, we eliminate arms for multiple clusters of users together.

- **Cluster-wise Arm Elimination:** In the second part, we no longer seek to partition each subset of users any further since users in the same subset provably correspond to the same cluster. Here, for elimination of bad arms, we pursue an intersection-based approach of good arms over all users in the same subset (Step 7 in Alg. 3); this is more aggressive elimination as compared to the union-based approach (Step 8 in Alg. 1) that was pursued in the previous component.

- **Upper Confidence Bound:** If number of active arms $\left|\mathcal{N}^{(\ell,i)}\right|$ for users in a subset $\mathcal{M}^{(\ell,i)} \in \mathcal{M}^{(\ell)}$ falls below γC, then we start/continue the Upper Confidence Bound (UCB) algorithm for each user in $\mathcal{M}^{(\ell,i)}$ separately with their subset of active arms $\mathcal{N}^{(\ell,i)}$ (Step 10 in Alg. 1).

Theoretical guarantees: We make similar assumptions on the reward matrix $\mathbf{P} \in \mathbb{R}^{N \times M}$ as in the CS framework:

Assumption 3 (Assumptions on reward matrix $\mathbf{P}$)**.** *We assume that $\mathbf{P}$ with SVD decomposition $\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}}$ satisfies the following properties 1) (Condition Number) $\mathbf{P}$ has rank C and has non zero singular values $\lambda_1 > \lambda_2 > \cdots > \lambda_C$ with $\lambda_1/\lambda_C = O(1)$ 2) (μ-incoherence) $\|\mathbf{U}\|_{2,\infty} \leq \sqrt{\mu C/N}$ and $\|\mathbf{V}\|_{2,\infty} \leq \sqrt{\mu C/M}$. 3) (Subset Strong Smoothness (a)) For some constant $\beta > 0$ and for any subset of indices $\mathcal{S} \subseteq [N]$, $\mathcal{S} = \mathcal{C}^{(j)}$ (corresponding to some cluster of users), we must have $\mathbf{x}^{\mathsf{T}} \mathbf{U}_{\mathcal{S}}^{\mathsf{T}} \mathbf{U}_{\mathcal{S}} \mathbf{x} \geq \beta\tau/C$ for all unit norm vectors $\mathbf{x} \in \mathbb{R}^C$. 4) (Subset Strong Smoothness (b)) For some $\alpha = O(1)$ and $\gamma = \widetilde{O}(1)$, for any subset of indices $\mathcal{S} \subseteq [M], |\mathcal{S}| = \gamma C$, we must have $\mathbf{x}^{\mathsf{T}} \mathbf{V}_{\mathcal{S}}^{\mathsf{T}} \mathbf{V}_{\mathcal{S}} \mathbf{x} \geq \alpha\gamma C/M$ for all unit norm vectors $\mathbf{x} \in \mathbb{R}^C$.*

Remark 6. *Note that the Subset Strong Smoothness (a) of Assumption 3 (used for proving incoherence guarantees of relevant sub-matrices of $\mathbf{P}$-Lemma 5) will be satisfied only if the separation ν is bounded from below (since for $\mathcal{S} = \mathcal{C}^{(j)}$, $\mathbf{U}_{\mathcal{S}}$ loses rank when $\nu = 0$). However, when $\nu = 0$, RCS reduces to the CS framework; here, we do not need (Subset Strong Smoothness (a)) since we have a different analysis for proving incoherence guarantees (Lemma 3) of relevant sub-matrices. For extremely small ν, we can combine the two analyses to obtain similar sufficient guarantees (by using triangle inequality for instance).*

As before, we characterize the condition number and the incoherence of the relevant sub-matrices of $\mathbf{P}$ that will allow us to apply low rank matrix completion techniques and provide theoretical guarantees (see Lemma 1).

**Soumyabrata Pal, Arun Sai Suggala, Karthikeyan Shanmugam, Prateek Jain**
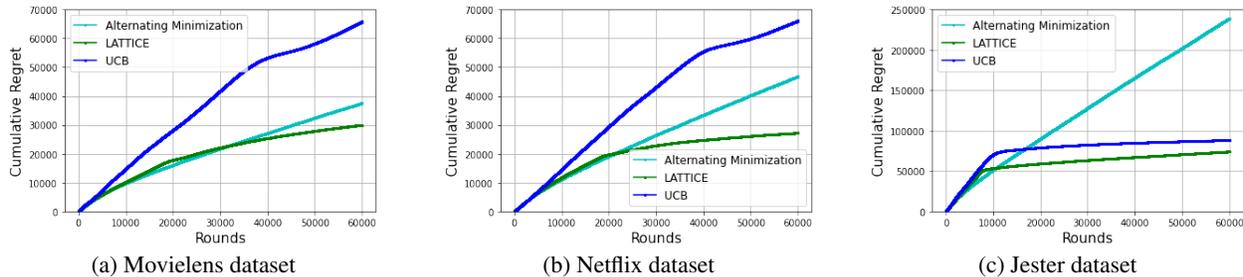
Figure 1: Cumulative Regret of the Alternating Minimization (AM) algorithm in Dadkhahi and Negahban [2018], LATTICE (Alg. 3) and Upper Confidence Bound (UCB) algorithm with $\mathsf{T} = 60000$ rounds for 3 datasets a) **Movielens 10m dataset:** $\mathsf{N} = 200$ users and $\mathsf{M} = 200$ arms b) **Netflix dataset:** $\mathsf{N} = 200$ users and $\mathsf{M} = 200$ arms c) **Jester dataset:** $\mathsf{N} = 100$ users and $\mathsf{M} = 100$ arms.

Lemma 4. *Suppose Assumption 3 is true. Consider a sub-matrix* $\mathbf{P}_{\mathsf{sub}}$ *of* $\mathbf{P}$ *having non-zero singular values* $\lambda'_1 > \cdots > \lambda'_{\mathsf{C}'}$ *(for* $\mathsf{C}' \le \mathsf{C}$*). Then, provided* $\mathbf{P}_{\mathsf{sub}}$ *is non-zero, we have* $\frac{\lambda'_1}{\lambda'_{\mathsf{C}'}} \le \frac{\lambda_1}{\lambda_{\mathsf{C}}}$.

Lemma 5. *Suppose Assumption 3 is true. Consider a sub-matrix* $\mathbf{P}_{\mathsf{sub}} \in \mathbb{R}^{\mathsf{N}' \times \mathsf{M}'}$ *(with SVD decomposition* $\mathbf{P}_{\mathsf{sub}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}$*) of* $\mathbf{P}$ *whose rows correspond to a nice subset of users. Then,* $\left\|\widetilde{\mathbf{U}}\right\|_{2,\infty} \le \sqrt{\frac{\mu\mathsf{C}}{\beta\mathsf{N}'}}$ *and* $\left\|\widetilde{\mathbf{V}}\right\|_{2,\infty} \le \sqrt{\frac{\mu\mathsf{C}}{\alpha\mathsf{M}'}}$.

Now, we are ready to state our main theorems

Theorem 5. *Consider the MAB-LC problem in* RCS *framework with* M *arms,* N *users,* C *clusters and* T *rounds such that at every round* $t \in [\mathsf{T}]$, *we observe reward* $\mathbf{R}^{(t)}$ *as defined in eq. (1) with noise variance proxy* $\sigma^2 > 0$. *Let* $\mathbf{P} \in \mathbb{R}^{\mathsf{N} \times \mathsf{M}}$ *be the expected reward matrix such that Assumption 3 is satisfied by* $\mathbf{P}$. *Moreover, suppose Assumption 2 is true. Then Alg. 3 with* $\mathsf{C}' = c\mathsf{C}^{-1}\min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log\mathsf{M}}\right)$ *for some appropriate constant* $c > 0$ *guarantees the regret* $\mathsf{Reg}(\mathsf{T})$ *to be:*

$$\widetilde{O}(\sigma\sqrt{\mu^3\mathsf{T}(\mathsf{M}+\mathsf{N})} + \sigma\sqrt{\mathsf{NT}}). \qquad (8)$$

Note that the regret bound above is similar to that of Theorem 1, despite the stricter setting. Here again, the "scale-free" regret is $\sqrt{\mathsf{T}(\mathsf{M}+\mathsf{N})}$ which as discussed in remarks below Theorem 1, is intuitive, is practical in realistic regimes, and is nearly optimal.

Remark 7. *Note that the generalization remarks 4,5 in Sec. 4.2 also extend to Theorem 5. Also, recall the definitions of* $\mathcal{H}_c, \mathcal{G}_{c,\ell}$ *for clusters* $c \in [\mathsf{C}]$ *and phases indexed by* $\ell \ge 1$ *depending on the sub-optimality gap from Section 4.2. With equivalent definitions for the* RCS *setting, the gap dependent regret bounds in Theorem 2 can be achieved by Algorithm 3 as well provided Assumptions 2 and 3 are true.*

6 EXPERIMENTS

We have provided detailed experiments on synthetic datasets (deferred to Appendix A) and popular real world recommendation data-sets namely 1) Movielens 10m dataset 2) Netflix

dataset and 3) Jester dataset. For simplicity, we implement a significantly simplified version of our algorithm (Alg. 4 in Appendix A). In addition, we have also compared with a highly competitive heuristic - the Alternating Minimization (AM) algorithm described in Dadkhahi and Negahban [2018] and the standard Upper Confidence Bound (see Lattimore and Szepesvári [2020]) algorithm individually for each user. However, we stress that the AM algorithm does not have any theoretical guarantees. For the Movielens dataset, we restricted ourselves to the 200 users (N) who have rated most movies and 200 movies (M) that have been rated the most. For Netflix and Jester, with a similar pre-processing, the values of N, M are $(200, 200)$ and $(100, 100)$ respectively. We compared the performance of our algorithm LATTICE (for GCS - i.e. after a few phases, we run UCB individually for each user with their active items) with the AM algorithm in Dadkhahi and Negahban [2018] (with the hyper-parameters provided in Dadkhahi and Negahban [2018] for Movielens and Jester datasets; for Netflix dataset, we used the hyperparameters provided for Movielens). In Figures 1a,1b and 1c, we have shown the cumulative regret of the three algorithms -clearly, LATTICE outperforms the other baselines empirically as well. In particular, LATTICE successfully removes large chunks of bad items for large groups of users jointly in few rounds. Further details about implementation are deferred to Appendix A.

7 CONCLUSION

For the multi-user multi-armed latent bandit problem introduced in Maillard and Mannor [2014] we provided a novel, computationally efficient algorithm LATTICE. Ours is the first algorithm to obtain $\widetilde{O}(\sqrt{(\mathsf{M}+\mathsf{N})\mathsf{T}})$ regret guarantee in this challenging and practically important setting, as latent cluster structure in users/agents is commonplace and is a standard modeling tool for practitioners. Our work also resolves open problems posed in Jain and Pal [2022] and Sen et al. [2017] for online low rank matrix completion in certain special case. Finally, it would be interesting to optimize the regret dependence on other factors such as the number of clusters (C), RCS-gap (ν), as well as other parameters incoherence and condition number.

References

Claudio Gentile, Shuai Li, and Giovanni Zappella. Online clustering of bandits. In *International Conference on Machine Learning*, pages 757–765. PMLR, 2014.

Odalric-Ambrym Maillard and Shie Mannor. Latent bandits. In *International Conference on Machine Learning*, pages 136–144. PMLR, 2014.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Aditya Mate, Lovish Madaan, Aparna Taneja, Neha Madhiwalla, Shresth Verma, Gargi Singh, Aparna Hegde, Pradeep Varakantham, and Milind Tambe. Field study in deploying restless multi-armed bandits: Assisting nonprofits in improving maternal and child health. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 12017–12025, 2022.

Joey Hong, Branislav Kveton, Manzil Zaheer, Yinlam Chow, Amr Ahmed, and Craig Boutilier. Latent bandits revisited. *Advances in Neural Information Processing Systems*, 33: 13423–13433, 2020.

Li Zhou and Emma Brunskill. Latent contextual bandits and their application to personalized recommendations for new users. *arXiv preprint arXiv:1604.06743*, 2016.

Shuai Li, Alexandros Karatzoglou, and Claudio Gentile. Collaborative filtering bandits. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 539–548, 2016.

Shuai Li, Wei Chen, and Kwong-Sak Leung. Improved algorithm on online clustering of bandits. *arXiv preprint arXiv:1902.09162*, 2019.

Claudio Gentile, Shuai Li, Purushottam Kar, Alexandros Karatzoglou, Giovanni Zappella, and Evans Etrue. On context-dependent clustering of bandits. In *International Conference on Machine Learning*, pages 1253–1262. PMLR, 2017.

Yunzhe Qi, Tianxin Wei, Jingrui He, et al. Neural collaborative filtering bandits via meta learning. *arXiv preprint arXiv:2201.13395*, 2022.

Prateek Jain and Soumyabrata Pal. Online low rank matrix completion. *arXiv preprint arXiv:2209.03997*, 2022.

Rajat Sen, Karthikeyan Shanmugam, Murat Kocaoglu, Alex Dimakis, and Sanjay Shakkottai. Contextual bandits with latent confounders: An nmf approach. In *Artificial Intelligence and Statistics*, pages 518–527. PMLR, 2017.

Hamid Dadkhahi and Sahand Negahban. Alternating linear bandits for online matrix-factorization recommendation. *arXiv preprint arXiv:1810.09401*, 2018.

Branislav Kveton, Csaba Szepesvári, Anup Rao, Zheng Wen, Yasin Abbasi-Yadkori, and S Muthukrishnan. Stochastic low-rank bandits. *arXiv preprint arXiv:1712.04644*, 2017.

Cindy Trinh, Emilie Kaufmann, Claire Vernade, and Richard Combes. Solving bernoulli rank-one bandits with unimodal thompson sampling. In *Algorithmic Learning Theory*, pages 862–889. PMLR, 2020.

Sumeet Katariya, Branislav Kveton, Csaba Szepesvári, Claire Vernade, and Zheng Wen. Bernoulli rank-1 bandits for click feedback. *arXiv preprint arXiv:1703.06513*, 2017.

Botao Hao, Jie Zhou, Zheng Wen, and Will Wei Sun. Low-rank tensor bandits. *arXiv preprint arXiv:2007.15788*, 2020.

Kwang-Sung Jun, Rebecca Willett, Stephen Wright, and Robert Nowak. Bilinear bandits with low-rank structure. In *International Conference on Machine Learning*, pages 3163–3172. PMLR, 2019.

Xiao-Yu Huang, Bing Liang, and Wubin Li. Online collaborative filtering with local and global consistency. *Information Sciences*, 506:366–382, 2020.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Low-rank generalized linear bandit problems. In *International Conference on Artificial Intelligence and Statistics*, pages 460–468. PMLR, 2021.

Guy Bresler, George H. Chen, and Devavrat Shah. A latent source model for online collaborative filtering. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pages 3347–3355, 2014.

Guy Bresler, Devavrat Shah, and Luis Filipe Voloch. Collaborative filtering with low regret. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, SIGMETRICS '16, pages 207–220, New York, NY, USA, 2016. ACM.

Reinhard Heckel and Kannan Ramchandran. The sample complexity of online one-class collaborative filtering. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pages 1452–1460. JMLR.org, 2017.

Guy Bresler and Mina Karzand. Regret bounds and regimes of optimality for user-user and item-item collaborative filtering. *arXiv:1711.02198*, 2019.

Wasim Huleihel, Soumyabrata Pal, and Ofer Shayevitz. Learning user preferences in non-stationary environments. In *International Conference on Artificial Intelligence and Statistics*, pages 1432–1440. PMLR, 2021.

Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *arXiv preprint arXiv:1902.07698*, 2019.

Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matri-

ces with low expected rank. *Annals of statistics*, 48(3): 1452, 2020.

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.

Srinadh Bhojanapalli and Prateek Jain. Universal matrix completion. In *International Conference on Machine Learning*, pages 1881–1889. PMLR, 2014.

Jaya Kawale, Hung H Bui, Branislav Kveton, Long Tran-Thanh, and Sanjay Chawla. Efficient thompson sampling for online matrix-factorization recommendation. *Advances in neural information processing systems*, 28, 2015.

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, New York, NY, USA, 2006.

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

Emilie Kaufmann. Contributions to the optimal solution of several bandits problems, 2020.

Aurélien Garivier and Emilie Kaufmann. Nonasymptotic sequential tests for overlapping hypotheses applied to near-optimal arm identification in bandit models. *Sequential Analysis*, 40(1):61–96, 2021.

Algorithm 4 Simplified LATTICE

Require: Number of users N, arms M, clusters C, rounds T, noise $\sigma^2 > 0$, phase lengths $\{\Delta_\ell\}_{\ell \geq 1}$ satisfying $\sum_\ell \Delta_\ell = T$. Gap Parameters $\{\nu_\ell\}_{\ell \geq 1}$. Parameter $\lambda \geq 0$ for nuclear norm minimization. Phase parameter $L \geq 0$ and robust intersection parameter $0 \leq \rho \leq 1$.

1: Partition entire time period into phases $[1, \Delta_1], [\Delta_1 + 1, \Delta_1 + \Delta_2], \ldots$.
2: Set $\mathcal{M}^{(1)} \equiv [\mathcal{M}^{(1,1)}]$ where $\mathcal{M}^{(1,1)} = [N]$ and $\mathcal{N}^{(1)} \equiv [\mathcal{N}^{(1,1)}]$ where $\mathcal{N}^{(1,1)} = [M]$.
3: **for** $t = 1, 2, \ldots, T$ **do**
4:     Sample user $u(t)$ from $[N]$, determine phase ℓ in which round t belongs
5:     Determine the set $\mathcal{M}^{(\ell,i)}$ in $\mathcal{M}^{(\ell)}$ in which $u(t)$ belongs.
6:     User $u(t)$ pulls an arm $\rho(t)$ uniformly at random from $\mathcal{N}^{(\ell,i)}$ and observes feedback $\mathbf{R}^{(t}$.
7:     **if** last round of phase and $\ell \leq L$ **then**
8:         Initialize $\mathcal{M}^{(\ell+1)} = [], \mathcal{N}^{(\ell+1)} = []$
9:         **for** each set $\mathcal{M}^{(\ell,i)} \in \mathcal{M}^{(\ell)}$ **do**
10:             Consider Matrix $\mathbf{Q} \in \mathbb{R}^{N \times M}$. Fill entries in the sub-matrix $\mathbf{Q}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}}$ by assigning $\mathbf{Q}_{ij} = \text{mean}\{\mathbf{R}^{(t)} \mid t \in$ phase $\ell, i = u(t), j = \rho(t), i \in \mathcal{M}^{(\ell,i)}\}$.
11:             Let Ω be the filled entries in $\mathbf{Q}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}}$. Complete the matrix $\mathbf{Q}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}}$ by solving the convex program

$$\min_{\mathbf{T} \in \mathbb{R}^{N \times M}} \frac{1}{2} \sum_{(i,j) \in \Omega} \left(\mathbf{Q}_{ij} - \mathbf{T}_{ij}\right)^2 + \lambda \|\mathbf{T}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}}\|_\star, \tag{9}$$

12:             **if** $\ell \leq L$ **then**
13:                 Solve k-means for users in $\mathcal{M}^{(\ell,i)}$ using the vector embedding formed by the rows in $\mathbf{T}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}}$. Choose best $k \leq C$ by using ELBOW method. Denote the cluster of users by $\{\mathcal{M}^{(\ell,i,j)}\}_j$.
14:                 **for** each cluster of users $\mathcal{M}^{(\ell,i,j)}$ **do**
15:                     Compute $\mathcal{N}^{(\ell,i,j)}$ as $\{s \in \mathcal{N}^{(\ell,i)} \mid |\mathbf{T}_{us} - \max_{s' \in \mathcal{N}^{(\ell,i)}} \mathbf{T}_{us'}| \leq \nu_\ell$ for some $u \in \mathcal{M}^{(\ell,i,j)}\}$.
16:                     Append $\mathcal{M}^{(\ell,i,j)}$ to $\mathcal{M}^{(\ell+1)}$ and $\mathcal{N}^{(\ell,i,j)}$ to $\mathcal{N}^{(\ell+1)}$.
17:                 **end for**
18:             **else**
19:                 Compute set of active arms $\mathcal{N}^{(\ell,i)}$ as $\{s \in \mathcal{N}^{(\ell,i)} \mid |\mathbf{T}_{us} - \max_{s' \in \mathcal{N}^{(\ell,i)}} \mathbf{T}_{us'}| \leq \nu_\ell$ for at least ρ-fraction of users in $\mathcal{M}^{(\ell,i)}\}$.
20:                 Append $\mathcal{M}^{(\ell,i)}$ to $\mathcal{M}^{(\ell+1)}$ and $\mathcal{N}^{(\ell,i)}$ to $\mathcal{N}^{(\ell+1)}$. *#Instead of Steps 20,21, we can also start running UCB individually for each user in $\mathcal{M}^{(\ell,i)}$ with the set of active items $\mathcal{N}^{(\ell,i)}$ for the remaining rounds. This can be more practical since cluster structures are not always satisfied exactly.*
21:             **end if**
22:   **end for**
23:     **end if**
24: **end for**

Organization: The Appendix is organized as follows: in Section A, we provide detailed synthetic experiments with a simplified version of the LATTICE algorithm. In Section B, we provide a more detailed comparison with the online clustering line of work. In Section C, we provide detailed proof for feasibility of Assumption 1. In Section D, we provide details on proofs of results presented in Section 3. In Section E, we provide detailed proof of Theorems 1 and 2. In Section F, we provide detailed proofs of the lower bounds on cumulative regret. In Section G, we provide detailed proof of Theorem 5. Finally in Section H, we provide a proof of a general version of Lemma 1 and the regret guarantee claimed in Remark 5.

A   Further Experiments

A.1   Synthetic Datasets

For experimentation, we will run Algorithm 4 that involves the following simplifications of a) *for each matrix completion step in Algorithm 1* - every user randomly pulls arms in the active set of arms (see Step 7 in Alg. 4) and subsequently, a single optimization problem with nuclear norm minimizer is solved at Step 12 b) *the clustering step using graphs in Steps 7-8 of Alg. 1* - we use k-means to cluster the users (see Step 14) where the vector embedding of each user is the row in the
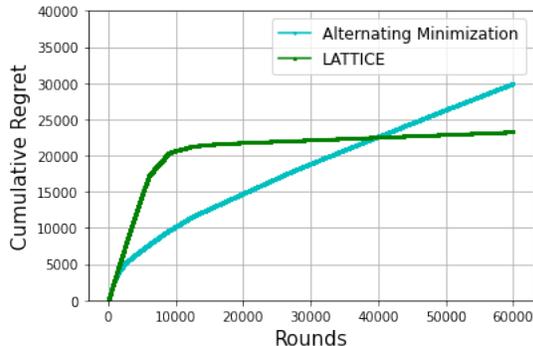
(a) $\mathbf{V}$ has entries distributed according to standard normal.
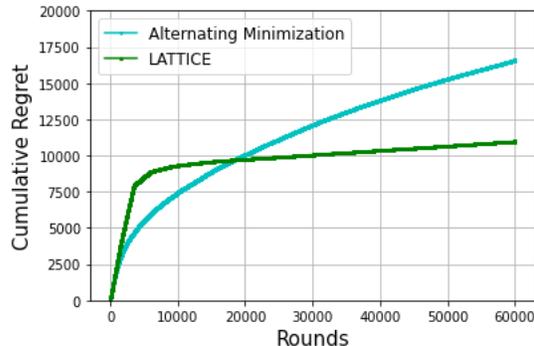
(b) $\mathbf{V}$ has entries generated according to a uniform distribution

Figure 2: Cumulative Regret of the Alternating Minimization (AM) algorithm in Dadkhahi and Negahban [2018] and LATTICE (Alg. 4) with $\mathsf{N} = 200$ users, $\mathsf{M} = 200$ arms, $\mathsf{C} = 4$ clusters and $\mathsf{T} = 60000$ rounds. The reward matrix $\mathbf{P} = \mathbf{U}\mathbf{V}^\mathsf{T}$ is generated in the following way: each row of $\mathbf{U}$ is a standard basis vector while each entry of $\mathbf{V}$ is sampled independently from a standard normal distribution $\mathcal{N}(0,1)$ in (a) and each entry of $\mathbf{V}$ is sampled independently from $\mathcal{U}[0,5]$ in (b). Notice that LATTICE becomes superior to AM as the number of rounds increase.

sub-matrix estimate that we computed by completing the sub-matrix corresponding to the subset of users (that the said user belongs to) and its active subset of arms.

Next we perform detailed experiments with Algorithm 4 on synthetic datasets that are generated as described below. Note that we compare against the Alternating Minimization (AM) algorithm presented in Dadkhahi and Negahban [2018] for solving the online multi-user multi-armed bandit problem when the reward matrix is of low-rank. Note that the AM algorithm is a very strong baseline in practice for our problem setting. In Dadkhahi and Negahban [2018], it was experimentally demonstrated for both synthetic and real datasets that the AM algorithm outperforms previously designed algorithms in the literature that can be applied to our problem setting (Sen et al. [2017] and Kawale et al. [2015]) by a significant margin.

Dataset Generation: We take $\mathsf{N} = 200$ users, $\mathsf{M} = 200$ arms, $\mathsf{C} = 4$ clusters and the number of rounds $\mathsf{T} = 60000$. We generate the ground truth matrix $\mathbf{P} = \mathbf{U}\mathbf{V}^\mathsf{T}$ where $\mathbf{U} \in \mathbb{R}^{\mathsf{N} \times \mathsf{C}}$, $\mathbf{V} \in \mathbb{R}^{\mathsf{M} \times \mathsf{C}}$ in the following manner: in the ith row of $\mathbf{U}$, the $(i\%\mathsf{C})^{\text{th}}$ entry is set to be 1 while the other entries are 0, each entry of $\mathbf{V}$ is sampled uniformly at random from (a) standard normal distribution $\mathcal{N}(0,1)$ (b) uniform distribution $[0,5]$. For (a), we assume that the noise added to each observed entry is sampled independently from $\mathcal{N}(0,0.5)$. For (b), we assume that the noise added is uniformly distributed in $[-0.5, 0.5]$.

Algorithm Details: We tune both the AM and the LATTICE algorithm (Alg. 4). The AM algorithm Dadkhahi and Negahban [2018] has two hyper-parameters λ_1, λ_2 which are set to be 0.5 and 0.01 respectively for both data-sets generated according to (a) and (b). Alg. 4 has several hyperparameters - we set the phase length $\Delta_\ell = 1500 + 500 * (\ell - 1)$, the gap parameters $\nu_\ell = ||\mathbf{P}||_\infty / 6 \cdot 8^\ell$ and $\mathsf{L} = 5$. We also take $\lambda = 5\sqrt{\Delta_\ell/200}$ for the convex relaxation problem in 9. In Step 14 of Alg. 4, we choose the best k using the following heuristic: we go on increasing k by 1 if the objective function of k-means reduces by a factor of at least 0.6; also if the objective is less than 100, we do not split the cluster anymore. Again, these hyperparameters remain same for both datasets generated according to (a) and (b).

Results and Insights: The cumulative regret (averaged over 5 independent runs) is plotted for both the AM and LATTICE algorithms (Alg. 4) in Figures 2a (Gaussian) and 2b (Uniform) respectively. Notice that for both synthetic datasets (Gaussian and Uniform), in the initial periods, AM has a better performance while in latter stages LATTICE improves significantly and eventually beats it. The reason is that the AM algorithm starts creating confidence sets for arm pulls for every user from the first few rounds itself. However, in almost all the runs, the AM algorithm fails to converge to the best arm for many users (although it does converge to arms with very small sub-optimality gap for each user). On the other hand, LATTICE, in the initial few phases mimics pure exploration but it converges to the best arm for most users almost always. Therefore, the cumulative regret of LATTICE hardly increases after a certain number of rounds whereas the cumulative regret of AM goes on increasing. Therefore, we can conclude that in synthetic datasets where our assumptions namely the cluster structure is satisfied, LATTICE is not only competitive with the AM algorithm but also has superior performance when the number of rounds is large. One disadvantage of the AM algorithm that it is quite sensitive to the choice of hyperparameters - a slightly

incorrect choice leads to diverging of the regret guarantees from the first few rounds itself; in comparison, LATTICE is much more stable with respect to the choice of hyperparameters.

A.2 Real-world datasets (Implementation details)

For all three datasets 1) Movielens 10m 2) Netflix and 3) Jester, we set the phase length in Algorithm 4 to be $\Delta_\ell = 2000 + 500 * (\ell - 1)$, the gap parameters $\nu_\ell = ||\mathbf{P}||_\infty / 4 \cdot 2^\ell$ and $\mathsf{L} = 5$ for Movielens and Netflix; for Jester dataset, we took $\nu_\ell = ||\mathbf{P}||_\infty / 6 \cdot 8^\ell$ and $\mathsf{L} = 5$. After five phases, instead of Steps 20,21 in Alg. 4, we start running standard UCB for each user with the set of active items for the remaining rounds. We also take $\lambda = 5\sqrt{\Delta_\ell / 200}$ for the convex relaxation problem in 9. In Step 14 of Alg. 4, we choose the best k using the following heuristic: we go on increasing k by 1 if the objective function of k-means reduces by a factor of at least 0.6; also if the objective is less than 50, we do not split the cluster anymore.

B Detailed comparison with Online Clustering

Gentile et al. [2014, 2017], Li et al. [2019] study the contextual version of the MAB-LC problem considered in our work. In their set up, a random user $u \in \mathcal{U}$ arrives at time t, the online algorithm is presented with an action space $\mathcal{A}_t$ where each action $a \in \mathcal{A}_t$ has a feature vector $\mathbf{x}_a$. If the action chosen is k_t, then the mean rewards obtained is $\mathbf{x}_{k_t}^T \beta_{c(u)}$ where β_c are the model parameters for cluster c. All users u such that $c(u) = c$ have an identical reward model.

We can map our problem to this problem by presenting a fixed action set in every slot, i.e. $\mathbf{x}_i = \mathbf{e}_i$ for all $i \in [\mathsf{M}]$ where $\{\mathbf{e}_i\}$ is the canonical basis in $\mathbb{R}^\mathsf{M}$, and $\mathcal{A}_t = \{\mathbf{x}\}_{i=1}^\mathsf{M}$ for all t. However, such a conversion results in highly sub-optimal regret of $\tilde{O}(\sqrt{\mathsf{M}^2 \mathsf{C} T} + \mathsf{M}^3 \mathsf{N})$. There are two main reasons for this. One is that this conversion leads to extremely high feature vector dimension of M. The other is that the algorithms in Gentile et al. [2014, 2017], Li et al. [2019] *crucially* depend on the assumption that for a fixed a at time t, the feature vector $\mathbf{x}_a$ is sampled i.i.d from a distribution on the unit sphere such that minimum singular value of $\mathbb{E}[\mathbf{x}_a \mathbf{x}_a^T]$ is at least a constant. Based on our conversion above, it is easy to see that $\mathbb{E}[\mathbf{x}_a \mathbf{x}_a^T] = \frac{1}{\mathsf{M}}$ for MAB-LC. Since M is very large, the minimum singular value in our setting is quite small which leads to poor regret. This assumption is crucial to the analysis of Gentile et al. [2014, 2017], Li et al. [2019], and removing it is non-trivial. Consider a user u in the system and the Gram matrix $S_{u,t} = \sum_{s<t:u_s=u} \mathbf{x}_{k_s} \mathbf{x}_{k_s}^T$ formed for user u based on feature vectors x_{k_s} of actions chosen at times when the user u arrived in the system. Crucial property that is needed for online clustering to proceed in the works of Gentile et al. [2014, 2017], Li et al. [2019] is that the minimum singular value of $S_{u,t}$ is $\Omega(T_{u,t})$ where $T_{u,t}$ is the number of time slots user u arrived till time t with very high probability (see for instance Claim 1 in Gentile et al. [2014], Lemma 4 in Li et al. [2019]). This is ensured through the randomness assumption for $\mathbf{x}_a$. In our case with mapping to canonical basis vectors, minimum singular value of $S_{u,t}$ will scale sub-linearly ($o(T_{u,t})$ for large $T_{u,t}$) if the algorithm is doing well on user u in terms of regret, i.e. focusing on arms close to the best arm.

Note that this can be seen as a motivation for our elimination style algorithm, since we only rely on the overlap in the set of 'good arms' of every user for clustering, while in Gentile et al. [2014, 2017], Li et al. [2019], the authors use the estimate of the entire mean reward vector to cluster. This requires the estimation error to be low in 'all directions' for the Gram matrix.

C Feasibility of Assumptions

In this section, we are going to show that Assumption 1 is feasible for a large ensemble of matrices. In particular when the entries of $\mathbf{X} \in \mathbb{R}^{\mathsf{C} \times \mathsf{M}}$ with SVD decomposition $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$ are sampled according to $\mathcal{N}(0,1)$, we prove that Assumption 1 is going to be true with high probability.

Lemma 6. *If $\mathbf{x}^\mathsf{T} \mathbf{X}_{|\mathcal{S}} \mathbf{X}_{|\mathcal{S}}^\mathsf{T} \mathbf{x} \geq \alpha\gamma \mathsf{C}\lambda_1^2/\mathsf{M}$ for a subset $\mathcal{S} \subseteq [\mathsf{M}], |\mathcal{S}| = \gamma\mathsf{C}$ for all unit vectors $\mathbf{x} \in \mathbb{R}^\mathsf{C}$, then the minimum eigenvalue of $\mathbf{V}_\mathcal{S}^\mathsf{T} \mathbf{V}_\mathcal{S} \geq \alpha\gamma\mathsf{C}/\mathsf{M}$. In other words, Subset Strong Smoothness (SSS) of $\mathbf{X}$ with SVD decomposition $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$ implies SSS of $\mathbf{V}$.*

Proof. If $\mathbf{x}^\mathsf{T} \mathbf{X}_{|\mathcal{S}} \mathbf{X}_{|\mathcal{S}}^\mathsf{T} \mathbf{x} \geq \alpha\gamma \mathsf{C}\lambda_1^2/\mathsf{M}$ for a subset $\mathcal{S} \subseteq [\mathsf{M}], |\mathcal{S}| = \gamma\mathsf{C}$ for all unit vectors $\mathbf{x} \in \mathbb{R}^\mathsf{C}$, then the minimum eigenvalue of $\mathbf{V}_\mathcal{S}^\mathsf{T} \mathbf{V}_\mathcal{S} \geq \alpha\gamma\mathsf{C}/\mathsf{M}$. To see this, note $\mathbf{X}_{|\mathcal{S}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}_\mathcal{S}^\mathsf{T}$ implying that $\mathbf{V}_\mathcal{S}^\mathsf{T} = (\mathbf{U}\mathbf{\Sigma})^{-1}\mathbf{X}_{|\mathcal{S}}$. Hence, $\mathbf{V}_\mathcal{S}^\mathsf{T} \mathbf{V}_\mathcal{S} = (\mathbf{U}\mathbf{\Sigma})^{-1}\mathbf{X}_{|\mathcal{S}}\mathbf{X}_{|\mathcal{S}}^\mathsf{T}(\mathbf{U}\mathbf{\Sigma})^{-\mathsf{T}}$ implying that $(\mathbf{V}_\mathcal{S}^\mathsf{T} \mathbf{V}_\mathcal{S})^{-1} = (\mathbf{U}\mathbf{\Sigma})^\mathsf{T}(\mathbf{X}_{|\mathcal{S}}\mathbf{X}_{|\mathcal{S}}^\mathsf{T})^{-1}(\mathbf{U}\mathbf{\Sigma})$. Taking the operator norm on both sides, we have $\lambda_{\min}(\mathbf{V}_\mathcal{S}^\mathsf{T} \mathbf{V}_\mathcal{S}) \geq \lambda_1^{-2}\lambda_{\min}(\mathbf{X}_{|\mathcal{S}}\mathbf{X}_\mathcal{S}^{|\mathsf{T}})$ implying that $\mathbf{x}^\mathsf{T}\mathbf{V}_\mathcal{S}^\mathsf{T} \mathbf{V}_\mathcal{S} \mathbf{x} \geq \alpha\gamma\mathsf{C}/\mathsf{M}$. $\square$

Lemma 7. *Suppose $\mathsf{C} \ll \mathsf{M}$ and further, the entries of $\mathbf{X}$ are generated independently according to $\mathcal{N}(0,1)$.*

Proof. We must have $\sqrt{M} - \sqrt{C} - t \leq \lambda_C \leq \lambda_1 \leq \sqrt{M} + \sqrt{C} + t$ w.p. at least $1 - 2e^{-t^2/2}$ implying that $\sqrt{M}/2 \leq \lambda_C \leq \lambda_1 \leq 2\sqrt{M}$; hence we must have $\lambda_1/\lambda_C = O(1)$ w.p. at least $1 - O(e^{-M})$. Moreover, we have $\mathbf{X}^\mathsf{T}\mathbf{X} = \mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^\mathsf{T}$. Clearly, we must have $||\mathbf{X}||_{\infty,2}\lambda_C^{-1} \leq ||\mathbf{V}||_{2,\infty} \leq ||\mathbf{X}||_{\infty,2}\lambda_1^{-1}$. For any column $\mathbf{X}_{|i}$, we have $\left\|\mathbf{X}_{|i}\right\|_2^2$ is a chi-squared random variable with C degrees of freedom. Using standard concentration inequalities for chi-squared random variables, we have $\left\|\mathbf{X}_{|i}\right\|_2 \leq 8\sqrt{C\log M}$ w.p. at least $1 - M^{-2}$. By taking a union bound over all $i \in [M]$, we have $||\mathbf{X}||_{\infty,2} \leq 8\sqrt{C\log M}$ w.p. at least $1 - M^{-1}$. Hence $||\mathbf{V}||_{2,\infty} \leq 16\sqrt{\frac{C\log M}{M}}$. Therefore, when $\mathbf{X}$ is a random gaussian matrix, the first part of Assumption 1 holds true and the second part holds true with $\mu = O(\log M)$ with high probability.

On the other hand, for a subset $\mathcal{S} \subseteq [M], |\mathcal{S}| = \gamma C$, we must have the minimum singular value of $\mathbf{X}_{|\mathcal{S}}$ to be at least $\sqrt{C}(\sqrt{\gamma} - 1) - t$ w.p. at least $1 - 2e^{-t^2/2}$. Taking $\gamma = 16\log M$ and $t = \sqrt{4C\log M}$, we must have the minimum singular value of $\mathbf{X}_{|\mathcal{S}}$ to be at least $\sqrt{C\gamma}/2$ w.p. at least $1 - 2e^{-2c\log M}$. Taking a union bound over all such subsets $\mathcal{S}$ of $[M]$ with size $|\mathcal{S}| = \gamma C$, we must have for all unit norm vectors $\mathbf{x} \in \mathbb{R}^C$

$$\mathbf{x}^\mathsf{T}\mathbf{X}_{|\mathcal{S}}\mathbf{X}_{|\mathcal{S}}^\mathsf{T}\mathbf{x} \geq \frac{C\gamma}{4} = \frac{C\gamma \cdot 4M}{16M} \geq \frac{C\gamma\lambda_1^2}{16M}$$

w.p. at least $1 - O(e^{-M})$ implying that $\alpha \geq 1/16$. Combining with Lemma 6, we can conclude that with high probability for $\gamma = O(\log M)$, the third part of Assumption 1 is satisfied for $\alpha = 1/16$ if $\mathbf{X}$ is a random gaussian matrix. These results can also be extended to matrices $\mathbf{X}$ whose columns are independent. $\square$

D Missing Details in Section 3

Low Rank Matrix Completion algorithm. Algorithm 5 describes the low-rank matrix completion algorithm we use in our work. This algorithm is adapted from Jain and Pal [2022], with minor modifications that are necessary for our setting. At its core, the algorithm solves a nuclear norm regularized convex objective to complete the matrix (equation (10)). This procedure is repeated $f = O(\log NMT)$ times, and the final estimate of the matrix is computed as the entry-wise median of the f solutions.

Algorithm 6 collects the data needed for matrix completion in Equation (10). At a high level, this subroutine randomly selects entries in the matrix (each entry is selected with probability p). It then computes an estimate of each of the selected entries by pulling the arm corresponding to the entry multiple times (b times) and taking an average of the obtained rewards. The collected data is then shared with Algorithm 5 for matrix completion. The main difficulty in implementing this algorithm is that in MAB-LC, the users arrive randomly in each iteration. Consequently, one has to wait for the required users to arrive to collect the necessary data. The question now is, how long does the algorithm wait to collect all the necessary data? By mapping this problem to the popular *Coupon Collector Problem*, it can be show that the the sample complexity of the algorithm increases atmost by $\log$ factors.

D.1 Proof of Lemma 3

Proof of Lemma 3. Suppose the reward matrix $\mathbf{P}$ has the SVD decomposition $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T}$. We are looking at a sub-matrix of $\mathbf{P}$ denoted by $\mathbf{P}_{\mathsf{sub}} \in \mathbb{R}^{N' \times M'}$ which can be represented as $\mathbf{U}_{\mathsf{sub}}\boldsymbol{\Sigma}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$ where $\mathbf{U}_{\mathsf{sub}}, \mathbf{V}_{\mathsf{sub}}$ are sub-matrices of $\mathbf{U}, \mathbf{V}$ respectively and are not necessarily orthogonal. Here, the rows in $\mathbf{P}_{\mathsf{sub}}$ corresponds to some cluster of users $C^{(a)}$ for $a \in C$ and the columns in $\mathbf{P}_{\mathsf{sub}}$ corresponds to the trimmed set of arms in $[M]$. Suppose the rows of the sub-matrix $\mathbf{P}_{\mathsf{sub}}$ correspond to the users in $C' \leq C$ clusters. Note that $\mathbf{U}_{\mathsf{sub}}$ can be further represented as $\mathbf{U}_{\mathsf{sub}}^{(1)}\mathbf{U}_{\mathsf{sub}}^{(2)}$ where $\mathbf{U}_{\mathsf{sub}}^{(1)} \in \mathbb{R}^{N' \times C'}$ is a binary matrix with orthonormal columns and 1-sparse rows (the non-zero index with value $1/\sqrt{\text{cluster size}}$ in the row indicates the cluster of the user). $\mathbf{U}_{\mathsf{sub}}^{(2)} \in \mathbb{R}^{C' \times C'}$ indicate the distinct rows of $\mathbf{U}_{\mathsf{sub}}$ corresponding to each of the C' clusters (multiplied by $\sqrt{\text{cluster size}}$).

Hence we can write (provided $\mathbf{V}_{\mathsf{sub}}$ is invertible)

$$\mathbf{P}_{\mathsf{sub}} = \mathbf{U}_{\mathsf{sub}}^{(1)}\mathbf{U}_{\mathsf{sub}}^{(2)}\boldsymbol{\Sigma}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{1/2}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T} = \mathbf{U}_{\mathsf{sub}}^{(1)}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}^\mathsf{T}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$$

where $\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}^\mathsf{T}$ is the SVD of the matrix $\mathbf{U}_{\mathsf{sub}}^{(2)}\boldsymbol{\Sigma}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{1/2}$. Since $\widehat{\mathbf{U}}$ is orthogonal, $\mathbf{U}_{\mathsf{sub}}^{(1)}\widehat{\mathbf{U}}$ is orthogonal as well. Similarly, $\widehat{\mathbf{V}}^\mathsf{T}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$ is orthogonal as well whereas $\widehat{\boldsymbol{\Sigma}}$ is diagonal. Hence $\mathbf{U}_{\mathsf{sub}}^{(1)}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}^\mathsf{T}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$ indeed corresponds to the SVD of $\mathbf{P}_{\mathsf{sub}}$ and we only need to argue about the incoherence of $\mathbf{U}_{\mathsf{sub}}^{(1)}\widehat{\mathbf{U}}$ and

---

Algorithm 5 LOW RANK MATRIX ESTIMATE (Adapted with slight modifications from Jain and Pal [2022])

Require: users $\mathcal{U} \subseteq [\mathsf{N}]$, arms $\mathcal{V} \subseteq [\mathsf{M}]$, rank r, incoherence μ of reward matrix $\mathbf{P}_{\mathcal{U},\mathcal{V}}$, total rounds T, noise variance σ^2, desired error in estimate ζ.

1: Set $d_2 = \min(|\mathcal{U}|, |\mathcal{V}|)$, sampling probability $p = C\mu^2 d_2^{-1} \log^3 d_2$, variance reduction factor $b = \left\lceil \left(\frac{c\sigma r\sqrt{\mu}}{\zeta \log d_2} \right)^2 \right\rceil$, number of repetitions of algorithm $f = O(\log \mathsf{MNT})$, regularization parameter $\lambda = C_\lambda \sigma \sqrt{d_2 p}$ for suitable constants $c, C, C_\lambda > 0$.

2: **for** $k = 0, 1, 2, \ldots, f$ **do**

3:      $\mathbf{Z} \leftarrow$ DATACOLLECTIONSUBROUTINE$(\mathcal{U}, \mathcal{V}, p, b)$

4:      Without loss of generality, assume $|\mathcal{U}| \leq |\mathcal{V}|$ (if $|\mathcal{U}| \geq |\mathcal{V}|$, we simply swap rows and columns). Randomly partition the columns into $k = \lceil |\mathcal{V}|/|\mathcal{U}| \rceil$ sets. Precisely, for each $i \in \mathcal{V}$, independently set δ_i to be a value in the set $\{1, 2 \ldots k\}$ uniformly at random. Partition indices in $\mathcal{V}$ into $\mathcal{V}^{(1)}, \mathcal{V}^{(2)}, \ldots, \mathcal{V}^{(k)}$ where $\mathcal{V}^{(q)} = \{i \in \mathcal{V} \mid \delta_i = q\}$ for each $q \in [k]$. Set $\Omega^{(q)} \leftarrow \Omega \cap (\mathcal{U} \times \mathcal{V}^{(q)})$ for all $q \in [k]$.

5:      **for** $q \in [k]$ **do**

6:          Solve convex program

$$\min_{\mathbf{Q}^{(q)} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{V}^{(q)}|}} \frac{1}{2} \sum_{(i,j) \in \Omega^{(q)}} \left(\mathbf{Q}^{(q)}_{i\pi(j)} - \mathbf{Z}_{ij} \right)^2 + \lambda \|\mathbf{Q}^{(q)}\|_\star, \qquad (10)$$

         where $\|\mathbf{Q}^{(q)}\|_\star$ denotes nuclear norm of matrix $\mathbf{Q}^{(q)}$ and $\pi(j)$ is index of j in set $\mathcal{V}^{(q)}$.

7:      **end for**

8:      Compute matrix $\widetilde{\mathbf{P}}^{(k)} \in \mathbb{R}^{\mathsf{N} \times \mathsf{M}}$ such that $\widetilde{\mathbf{P}}^{(k)}_{\mathcal{U},\mathcal{V}^{(q)}} = \mathbf{Q}^{(q)}$ for all $q \in [k]$ and for every index $(i,j) \notin \mathcal{U} \times \mathcal{V}$, $\widetilde{\mathbf{P}}^{(k)}_{ij} = 0$. *#We combine the estimates of each of the smaller matrices to form an estimate of the larger matrix i.e. without partitioning the columns. Moreover, $\{\widetilde{\mathbf{P}}^{(k)}_{\mathcal{U},\mathcal{V}}\}^f_{k=1}$ correspond to f independent estimates of $\mathbf{P}_{\mathcal{U},\mathcal{V}}$.*

9: **end for**

10: **return** entry-wise median of $\{\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}\}$.

Algorithm 6 DATACOLLECTIONSUBROUTINE

Require: users $\mathcal{U} \subseteq [\mathsf{N}]$, arms $\mathcal{V} \subseteq [\mathsf{M}]$, sampling probability p, variance reduction factor b.

1: For each tuple $(u, v) \in \mathcal{U} \times \mathcal{V}$, independently set $\delta_{uv} = 1$ with probability p and $\delta_{uv} = 0$ with probability $1 - p$. Let $\Omega \subseteq \mathcal{U} \times \mathcal{V}$ to be the set of indices with $\delta_{uv} = 1$.

2: **for** $\ell = 1, 2, \ldots, b$ **do**

3:      For all $(i, j) \in \Omega$, set $\mathsf{Mask}_{ij} = 0$.

4:      **while** there exists $(i, j) \in \Omega$ such that $\mathsf{Mask}_{ij} = 0$ **do**

5:          WAIT UNTIL next round t such that some $u(t) \in \mathcal{U}$ is sampled.

6:          For user $u(t) \in \mathcal{U}$, pull arm $\rho(t)$ in $\{j \in \mathcal{V} \mid (u(t), j) \in \Omega, \mathsf{Mask}_{u(t)j} = 0\}$ and set $\mathsf{Mask}_{u(t)\rho(t)} = 1$. If not possible then pull any arm $\rho(t)$ in $\mathcal{V}$ such that $(u(t), \rho(t)) \notin \Omega$. Observe $\mathbf{R}^{(t)}$.

7:      **end while**

8: **end for**

9: Initialize a matrix $\mathbf{Z}$ of size $|\mathcal{U}| \times |\mathcal{V}|$ to 0's.

10: For each tuple $(i, j) \in \Omega$, set $\mathbf{Z}_{ij}$ to be average of rewards collected for that entry.

11: **return** $\mathbf{Z}$

$(\widehat{\mathbf{V}}^\mathsf{T} (\mathbf{V}^\mathsf{T}_{\mathsf{sub}} \mathbf{V}_{\mathsf{sub}})^{-1/2} \mathbf{V}^\mathsf{T}_{\mathsf{sub}})^\mathsf{T}$. Notice that $\max_i \|(\mathbf{U}^{(1)}_{\mathsf{sub}} \widehat{\mathbf{U}})^\mathsf{T} \mathbf{e}_i\|_2 \leq \|(\mathbf{U}^{(1)}_{\mathsf{sub}})^\mathsf{T} \mathbf{e}_i\|_2 \leq 1/\sqrt{\mathsf{J}_{\min}}$ where $\mathsf{J}_{\min}$ is the minimum cluster size. Since $\tau = \frac{\mathsf{J}_{\max}}{\mathsf{J}_{\min}}$ and $\mathsf{CJ}_{\max} \geq \mathsf{N}'$, we must have $\mathsf{J}_{\min} \geq \frac{\mathsf{N}'}{\mathsf{C}\tau}$. Hence $\max_i \|(\mathbf{U}^{(1)}_{\mathsf{sub}})^\mathsf{T} \mathbf{e}_i\|_2 \leq \sqrt{\frac{\mathsf{C}\tau}{\mathsf{N}'}}$. Similarly, we have

$$\max_i \|\widehat{\mathbf{V}} (\mathbf{V}^\mathsf{T}_{\mathsf{sub}} \mathbf{V}_{\mathsf{sub}})^{-1/2} \mathbf{V}^\mathsf{T}_{\mathsf{sub}} \mathbf{e}_i\| \leq \max_i \|(\mathbf{V}^\mathsf{T}_{\mathsf{sub}} \mathbf{V}_{\mathsf{sub}})^{-1/2} \mathbf{V}^\mathsf{T}_{\mathsf{sub}} \mathbf{e}_i\|$$

$$\leq \frac{\|\mathbf{V}_{\mathsf{sub}}\|_{2,\infty}}{\sqrt{\lambda_{\min}(\mathbf{V}^\mathsf{T}_{\mathsf{sub}} \mathbf{V}_{\mathsf{sub}})}} \leq \frac{\|\mathbf{V}\|_{2,\infty}}{\sqrt{\lambda_{\min}(\mathbf{V}^\mathsf{T}_{\mathsf{sub}} \mathbf{V}_{\mathsf{sub}})}} \leq \sqrt{\frac{\mu\mathsf{C}}{\alpha\mathsf{M}'}}$$

where the last line follows from the fact that $\min_{\mathbf{x} \in \mathbb{R}^\mathsf{C}} \mathbf{x}^\mathsf{T} \mathbf{V}^\mathsf{T}_{\mathsf{sub}} \mathbf{V}_{\mathsf{sub}} \mathbf{x} = \min_{\mathbf{x} \in \mathbb{R}^\mathsf{C}} \sum_{i \in \mathcal{S}} \mathbf{x}^\mathsf{T} \mathbf{V}_i \mathbf{V}^\mathsf{T}_i \mathbf{x}$ where $\mathcal{S}, |\mathcal{S}| = \mathsf{M}'$ is the set of rows in $\mathbf{V}$ represented in $\mathbf{V}_{\mathsf{sub}}$. Hence we can partition the set $\mathcal{S}$ into $|\mathcal{S}|/(\gamma\mathsf{C})$ disjoint sets, each of size $\gamma\mathsf{C}$; from Assumption 1, we get that the minimum over each of those sets is $\alpha\gamma\mathsf{C}/\mathsf{M}$ and therefore, the conclusion follows.

Algorithm 7 UPPER CONFIDENCE BOUND (UCB) Bubeck et al. [2012]

Require: user $u \in [\mathsf{N}]$, set of arms $\mathcal{V} \subseteq [\mathsf{M}]$, error ζ, noise σ^2, total rounds T.
 1: WAIT UNTIL round t when user $u(t)$ is sampled.
 2: Choose arm $j_t \in \mathcal{V}$ according to $j_t = \mathrm{argmax}_{j \in \mathcal{V}} \mathsf{UCB}_j(s-1, \mathsf{T}^{-3})$ where $s-1$ is the number of rounds ran in this instantiation of the UCB algorithm so far and (suppose $t_j(s-1)$ is the number of times arm $j \in \mathcal{V}$ has been pulled by u in the previous $s-1$ rounds and $\widehat{\mu}_j(s-1)$ is the empirical mean of the arm j due to the previous $t_j(s-1)$ observations)

$$\mathsf{UCB}_j(s-1, \mathsf{T}^{-3}) = \begin{cases} \infty & \text{if arm } j \text{ has not been played before} \\ \widehat{\mu}_j(s-1) + \sigma\sqrt{\frac{6\log \mathsf{T}}{t_j(s-1)}} \end{cases}$$

$\square$

D.2 Proof of Lemma 2

Proof of Lemma 2. Suppose $\mathbf{P}_{\mathsf{sub}}$ has dimensions $m' \times n'$. We have that $\lambda_1^2 = \sup_{\mathbf{x} \in \mathbb{R}^n | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \mathbf{P}^\mathsf{T} \mathbf{P} \mathbf{x}$ and $(\lambda_1')^2 = \sup_{\mathbf{x} \in \mathbb{R}^{n'} | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \mathbf{P}_{\mathsf{sub}}^\mathsf{T} \mathbf{P}_{\mathsf{sub}} \mathbf{x}$. Similarly, we can write $\lambda_1' = \sup_{\mathbf{x} \in \mathbb{R}^{n'} | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \sum_{i \in \mathcal{S}} (\mathbf{P}_{i|\mathcal{S}'}^\mathsf{T} \mathbf{P}_{i|\mathcal{S}'}) \mathbf{x}$ where $\mathbf{P}_{i|\mathcal{S}'}$ is the ith row of the matrix $\mathbf{P}$ restricted to the column indices in $\mathcal{S}'$. Since $\mathbf{x}^\mathsf{T}(\mathbf{b}\mathbf{b}^\mathsf{T})\mathbf{x} > 0$ for any vector $\mathbf{x}$, we have

$$\mathsf{J}_{\max} \lambda_{\max}^2 \geq \mathsf{J}_{\max} \sup_{\mathbf{x} \in \mathbb{R}^n | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \sum_{i \in [\mathsf{C}]} (\mathbf{X}_i^\mathsf{T} \mathbf{X}_i) \mathbf{x} \overset{(a)}{\geq} \sup_{\mathbf{x} \in \mathbb{R}^n | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \sum_{i \in [n]} (\mathbf{P}_i^\mathsf{T} \mathbf{P}_i) \mathbf{x}$$

$$\geq \sup_{\mathbf{x} \in \mathbb{R}^n | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \sum_{i \in \mathcal{S}} (\mathbf{P}_i^\mathsf{T} \mathbf{P}_i) \mathbf{x} \geq \sup_{\mathbf{x} \in \mathbb{R}^n | ||\mathbf{x}||_2 = 1, \mathbf{x}_{|[n] \setminus \mathcal{S}'} = \mathbf{0}} \mathbf{x}^\mathsf{T} \sum_{i \in \mathcal{S}} (\mathbf{P}_i^\mathsf{T} \mathbf{P}_i) \mathbf{x} \geq \sup_{\mathbf{x} \in \mathbb{R}^{n'} | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \sum_{i \in \mathcal{S}} (\mathbf{P}_{i|\mathcal{S}'}^\mathsf{T} \mathbf{P}_{i|\mathcal{S}'}) \mathbf{x}$$

implying that $\sqrt{\mathsf{J}_{\max}} \lambda_{\max} \geq \lambda_1'$. The inequality (a) follows from the fact that $\mathsf{J}_{\max} \mathbf{X}_i^\mathsf{T} \mathbf{X}_i \geq \sum_{u \in [n] | u \in \mathcal{C}^{(i)}} \mathbf{P}_i^\mathsf{T} \mathbf{P}_i$. In order to prove the inequality on λ_r', we need to do some more work. Let us denote the nullspace of the vectors $\{\mathbf{P}_i\}_{i \in [m]}$ by $\mathcal{K}$ and the subspace orthogonal to $\mathcal{K}$ by $\mathcal{K}^c$. Let us denote the SVD of $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$ where $\mathbf{U} \in \mathbb{R}^{\mathsf{C} \times \mathsf{C}}$, $\mathbf{V}^\mathsf{T} \in \mathbb{R}^{\mathsf{C} \times \mathsf{M}}$. Consider the matrix $\mathbf{P}$ restricted to the rows in the set $\mathcal{S}$ denoted by $\mathbf{P}_{\mathcal{S}}$. Notice that $\mathbf{P}_{\mathcal{S}}^\mathsf{T} \mathbf{P}_{\mathcal{S}} = \mathbf{V}\mathbf{\Sigma}\mathbf{U}^\mathsf{T}\mathbf{D}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$ where $\mathbf{D}$ is a diagonal matrix whose i^{th} entry corresponds to the number of users in the set $\mathcal{S}$ belonging to the i^T cluster. If the set $\mathcal{S}$ is a union of clusters, then the minimum diagonal entry in $\mathbf{D}$ must be $\mathsf{J}_{\min}$. Let us denote the nullspace of the matrix $\mathbf{P}_{\mathcal{S}}$ by $\mathcal{K}$. We can write the minimum non-zero eigenvalue of $\mathbf{P}_{\mathcal{S}}^\mathsf{T} \mathbf{P}_{\mathcal{S}}$ as $\min_{\mathbf{x} \in \mathcal{K}^c | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \mathbf{V}\mathbf{\Sigma}\mathbf{U}^\mathsf{T}\mathbf{D}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}\mathbf{x}$. Let us write the vector $\mathbf{z} = \mathbf{\Sigma}\mathbf{V}^\mathsf{T}\mathbf{x}$; $\mathbf{z}$ must belong to the sub-space $\mathcal{T}$ spanned by the rows of $\mathbf{U}$ corresponding to the non-zero diagonal indices of $\mathbf{D}$ as otherwise $\mathbf{x}$ will have non-zero projection on the null-space of $\mathbf{P}_{\mathcal{S}}$. Note that the rows of $\mathbf{U}$ are orthonormal as well (since $\mathbf{U}$ is a square matrix i.e. $\mathbf{U}^\mathsf{T} = \mathbf{U}^{-1}$). Hence, $\min_{\mathbf{z} \in \mathcal{T}} \mathbf{z}^\mathsf{T} \mathbf{U}^\mathsf{T} \mathbf{D} \mathbf{U} \mathbf{z} = \mathsf{J}_{\min} ||\mathbf{z}||_2^2 \geq \mathsf{J}_{\min} \lambda_{\min}^2$. Next, we have that

$$\inf_{\mathbf{x} \in \mathcal{K}^c | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \sum_{i \in \mathcal{S}} (\mathbf{P}_i^\mathsf{T} \mathbf{P}_i) \mathbf{x}$$

$$\overset{(b)}{\leq} \inf_{\mathbf{x} \in \mathcal{K}^c | ||\mathbf{x}||_2 = 1, \mathbf{x}_{|[n] \setminus \mathcal{S}'} = \mathbf{0}} \mathbf{x}^\mathsf{T} \sum_{i \in \mathcal{S}} (\mathbf{P}_i^\mathsf{T} \mathbf{P}_i) \mathbf{x} \leq \inf_{\mathbf{x} \in (\mathcal{K}')^c | ||\mathbf{x}||_2 = 1} \mathbf{x}^\mathsf{T} \sum_{i \in \mathcal{S}} (\mathbf{P}_{i|\mathcal{S}'}^\mathsf{T} \mathbf{P}_{i|\mathcal{S}'}) \mathbf{x}.$$

where $\mathcal{K}'$ is the null-space of the matrix $\mathbf{P}$ restricted to the columns in $\mathcal{S}'$. Note that the step (b) follows as long as the there exists a vector with non-zero entries only on $\mathcal{S}'$ in the row space which implies that the sub-matrix is non-zero.

Therefore $\lambda_{|\mathsf{C}|'}' \geq \sqrt{\mathsf{J}_{\min}} \lambda_{\min}$. Using the fact that $\tau = \mathsf{J}_{\max}/\mathsf{J}_{\min}$, the lemma is proved.

$\square$

E   Proofs of Theorems 1 and 2

E.1 Proof Overview

For any phase indexed by ℓ, we are going to prove that conditioned on the events $\mathcal{E}^{(\ell)}$, the event $\mathcal{E}^{(\ell+1)}$ is also going to be true with high probability with proper choice of $\Delta_{\ell+1}, \epsilon_{\ell+1}$. First, inspired by low rank matrix completion techniques, conditioned

on $\mathcal{E}^{(\ell)}$ and by using Lemmas 2, 3 along with the fact that each set of users in $\mathcal{M}^{(\ell)}$ is *nice*, we can show with Lemma 1 that in phase ℓ, by using $m_\ell = O(\mathsf{V}/\Delta_{\ell+1}^2)$ (where $\mathsf{V} = \widetilde{O}(\sigma^2\mu^3(\mathsf{N}+\mathsf{M}))$) rounds, $\mathcal{E}_2^{(\ell)}$ is true with high probability (see Alg. 5 in Appendix D). Next, we can show the following series of lemmas (let $\mathcal{B}^{(\ell)} = \bigcup_{i\in[a_\ell]||\mathcal{N}^{(\ell,i)}|\geq\gamma\mathsf{C}} \mathcal{M}^{(\ell,i)}$ denote the set of users having more than $\gamma\mathsf{C}$ active arms) regarding the sets $\mathcal{T}_u^{(\ell)}$ of good arms obtained from the estimates of the relevant reward sub-matrices (Step 6 in Alg. 5):

Lemma 8. *Conditioned on $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, for every $u \in \mathcal{B}^{(\ell)}$, $\pi_u(1) \in \mathcal{T}_u^{(\ell)}$ and $\max_{s,s'\in\mathcal{T}_u^{(\ell)}} |\mathbf{P}_{us} - \mathbf{P}_{us'}| \leq 4\Delta_{\ell+1}$.*

Lemma 8 states that for every relevant user, the best arm always belongs to the set of good arms and characterizes how good the remaining arms are.

Lemma 9. *Fix any $i \in [a_\ell]$ such that $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma\mathsf{C}$. Conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, nodes in $\mathcal{M}^{(\ell,i)}$ corresponding to the same cluster form a clique. Also, users in each connected component of the graph $\mathcal{G}^{(\ell,i)}$ form a nice subset.*

Recall that we draw a graph $\mathcal{G}^{(\ell,i)}$ with nodes corresponding to users in $\mathcal{M}^{(\ell,i)}$ and edges drawn according to eq. (6). Lemma 9 says that users in same cluster always form a clique; however, this does not rule out inter-cluster edges. Nevertheless, if two users have an edge, then the next lemma shows that good arms for one are good for the other:

Lemma 10. *Fix any $i \in [a_\ell]$ such that $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma\mathsf{C}$. Consider two users $u, v \in \mathcal{M}^{(\ell,i)}$ having an edge in the graph $\mathcal{G}^{(\ell,i)}$. Conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, we must have*

$$\max_{x\in\mathcal{T}_u^{(\ell)}, y\in\mathcal{T}_v^{(\ell)}} |\mathbf{P}_{ux} - \mathbf{P}_{uy}| \leq 16\Delta_{\ell+1}$$

$$and \quad \max_{x\in\mathcal{T}_u^{(\ell)}, y\in\mathcal{T}_v^{(\ell)}} |\mathbf{P}_{vx} - \mathbf{P}_{vy}| \leq 16\Delta_{\ell+1}$$

We can extend Lemma 10 to the case when two users have a path joining them in the constructed graph:

Lemma 11. *Fix any $i \in [a_\ell]$ such that $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma\mathsf{C}$. Consider two users $u, v \in \mathcal{M}^{(\ell,i)}$ having a path in the graph $\mathcal{G}^{(\ell,i)}$. Conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, we must have*

$$\max_{x\in\mathcal{T}_u^{(\ell)}, y\in\mathcal{T}_v^{(\ell)}} |\mathbf{P}_{ux} - \mathbf{P}_{uy}| \leq 32\mathsf{C}\Delta_{\ell+1}$$

$$and \quad \max_{x\in\mathcal{T}_u^{(\ell)}, y\in\mathcal{T}_v^{(\ell)}} |\mathbf{P}_{vx} - \mathbf{P}_{vy}| \leq 32\mathsf{C}\Delta_{\ell+1}.$$

Thus we can create the new sets of nice users as the connected components of the graph and the corresponding set of arms that are good for all is constructed by the union of the set of good arms (eq. 7). If the number of active arms for a set of users become less than $\gamma\mathsf{C}$, then we start a UCB algorithm for each of them until the end of the number of rounds. Therefore, conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, the event $\mathcal{E}^{(\ell)}$ is also going to be true. Hence, conditioned on the event $\mathcal{E}^{(\ell)}$, we can bound the regret in each round of the ℓth phase by ϵ_ℓ; roughly speaking, the number of rounds in the ℓth phase is $1/\epsilon_\ell^2$ and therefore the regret is $1/\epsilon_\ell$. By setting Δ_ℓ as in Step 3 of Alg. 1 (and $\epsilon_\ell = \Delta_\ell/64\mathsf{C}$), we can bound the regret of LATTICE and achieve the guarantee in Theorems 1 and 2.

E.2   Detailed Proof

LATTICE is run in phases indexed by $\ell = 1, 2, \dots$. In the beginning of each phase ℓ, we have the following set of desirable properties:

1. Maintain a list of groups of users $\mathcal{M}^{(\ell)} \equiv \{\mathcal{M}^{(\ell,1)}, \mathcal{M}^{(\ell,2)}, \dots, \mathcal{M}^{(\ell,a_\ell)}\}$ and arms $\mathcal{N}^{(\ell)} \equiv \{\mathcal{N}^{(\ell,1)}, \mathcal{N}^{(\ell,2)}, \dots, \mathcal{N}^{(\ell,a_\ell)}\}$ where $a_\ell \leq \mathsf{C}$ such that $\cup_{i\in[a_\ell]}\mathcal{M}^{(\ell,i)} = [\mathsf{N}]$ and $\cup_{i\in[a_\ell]}\mathcal{N}^{(\ell,i)} \subseteq [\mathsf{M}]$.

2. Moreover, for all $i \in [a_\ell]$, we will have $\mathcal{M}^{(\ell,i)} = \bigcup_{j\in\mathcal{G}^{(\ell,i)}} \mathcal{C}^{(j)}$ where the sets $\{\mathcal{G}^{(\ell,1)}, \mathcal{G}^{(\ell,2)}, \dots, \mathcal{G}^{(\ell,a_\ell)}\}$ form a partition of the set $[\mathsf{C}]$. This implies that every set of users in the family $\mathcal{M}^{(\ell)}$ is *nice* and the sets of users in $\mathcal{M}^{(\ell)}$ form a partition of $[\mathsf{N}]$.

3. For each group $\mathcal{M}^{(\ell,i)}$ in the list $\mathcal{M}^{(\ell)}$, we will have an active set of arms denoted by $\mathcal{N}^{(\ell,i)}$ such that $\mathcal{N}^{(\ell,i)} \supseteq \{\mathrm{argmax}_j \mathbf{P}_{uj} \mid u \in \mathcal{M}^{(\ell,i)}\}$ i.e. for each user u in the set $\mathcal{M}^{(\ell,i)}$, their best arm must belong to the set $\mathcal{N}^{(\ell,i)}$.

Furthermore, for all $i \in [a_\ell]$ such that $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma\mathsf{C}$, the set $\mathcal{N}^{(\ell,i)}$ must also satisfy the following:

$$\left|\max_{j \in \mathcal{N}^{(\ell,i)}} \mathbf{P}_{uj} - \min_{j \in \mathcal{N}^{(\ell,i)}} \mathbf{P}_{uj}\right| \leq \epsilon_\ell \text{ for all } u \in \mathcal{M}^{(\ell,i)} \tag{11}$$

where ϵ_ℓ is a fixed exponentially decreasing sequence in ℓ (in particular, we choose $\epsilon_1 = ||\mathbf{P}||_\infty$ and $\epsilon_\ell = C'2^{-\ell} \min\left(||\mathbf{P}||_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\right)$ for $\ell > 1$ for some constant $C' > 0$).

4. Let $\mathcal{B}^{(\ell)} \subseteq [\mathsf{N}]$ be a subset of users satisfying $\mathcal{B}^{(\ell)} = \bigcup_{i \in [a_\ell] || \mathcal{N}^{(\ell,i)}| \geq \gamma\mathsf{C}} \mathcal{M}^{(\ell,i)}$ i.e. $\mathcal{B}^{(\ell)}$ corresponds to the set of users which belong to a group $\mathcal{M}^{(\ell,i)}$ at the beginning of the ℓth phase having more than C active arms. We will also maintain that $\mathcal{B}^{(i)} \supseteq \mathcal{B}^{(j)}$ for any phase $i \leq j$ i.e. the set of users with more than C active arms goes on shrinking.

Since LATTICE is random, we will say that our algorithm is $\epsilon_\ell-good$ at the beginning of the ℓth phase if the algorithm can maintain a list of users and arms satisfying the above properties at the start of phase ℓ. Let us also define the event $\mathcal{E}^{(\ell)}$ to be true if properties $(1-4)$ are satisfied at the beginning of phase ℓ by the phased elimination algorithm. We are going to prove inductively that the phased elimination algorithm is ϵ_ℓ-good for all phases indexed by ℓ for our choice of $\{\epsilon_\ell\}$ with high probability as long as the number of phases are small.

Base Case: For $\ell = 1$ (the first phase), we initialize $\mathcal{M}^{(1,1)} = [\mathsf{N}]$, $\mathcal{N}^{(1,1)} = [\mathsf{M}]$ and therefore, we have

$$\left|\max_{j \in \mathcal{N}^{(\ell,1)}} \mathbf{P}_{uj} - \min_{j \in \mathcal{N}^{(\ell,1)}} \mathbf{P}_{uj}\right| \leq ||\mathbf{P}||_\infty \text{ for all } u \in [\mathsf{M}].$$

Hence, we also have $\mathcal{B}^{(1)} = [\mathsf{N}]$. Moreover, the set of users $[\mathsf{N}]$ satisfies $[\mathsf{N}] = \cup_{j \in \mathcal{G}^{(1,1)}} \mathcal{C}^{(j)}$ where $\mathcal{G}^{(1,1)} = [\mathsf{C}]$ and finally for every user $u \in [\mathsf{N}]$, the best arm $\mathrm{argmax}_j \mathbf{P}_{uj}$ belongs to the entire set of arms. Thus for $\ell = 1$, our initialization makes the algorithm $||\mathbf{P}||_\infty$-good.

Inductive Argument: Suppose, at the beginning of the phase ℓ, we condition on the event $\mathcal{E}^{(\ell)}$ that Algorithm is $\epsilon_\ell-good$. Next, our goal is to run Matrix completion in order to estimate each of the sub-matrices corresponding to $\{(\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)})\}_{i \in [\mathsf{C}]}$. Fix the quantity $\Delta_{\ell+1} > 0$. We can show the following lemma:

Lemma 12. *Let us fix $\Delta_{\ell+1} > 0$ and condition on the event $\mathcal{E}^{(\ell)}$. Suppose Assumptions 1 and 2 are satisfied. In that case, in phase ℓ, by using*

$$m_\ell = O\left(\frac{\sigma^2 \mathsf{C}^2 (\mathsf{C} \vee \mu\alpha^{-1})^3 \log \mathsf{M}}{\Delta_{\ell+1}^2} \left(\mathsf{N} \vee \mathsf{MC}\right) \log^2(\mathsf{MNC}\delta^{-1})\right)$$

rounds, we can compute an estimate $\widetilde{\mathbf{P}}^{(\ell)} \in \mathbb{R}^{\mathsf{N} \times \mathsf{M}}$ such that with probability $1 - \delta$, we have for a nice subset of users $\mathcal{M}^{(\ell,i)}$

$$\left|\left|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}^{(\ell)} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\right|\right|_\infty \leq \Delta_{\ell+1} \text{ for all } i \in [a_\ell] \text{ satisfying } \left|\mathcal{N}^{(i,\ell)}\right| \geq \gamma\mathsf{C}. \tag{12}$$

Proof of Lemma 12. We are going to use Lemma 1 in order to compute an estimate $\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}^{(\ell)}$ of the sub-matrix $\mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$ satisfying $\left|\left|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}^{(\ell)} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\right|\right|_\infty \leq \Delta_{\ell+1}$. From Lemma 1, we know that by using $m_\ell = O\left(s\mathsf{N} \log^2(\mathsf{MN}\delta^{-1})(|\mathcal{N}^{(\ell,i)}|p + \sqrt{|\mathcal{N}^{(\ell,i)}|p \log \mathsf{N}\delta^{-1}})\right)$ rounds (see Lemma 1) restricted to users in $\mathcal{M}^{(\ell,i)}$ such that with probability at least $1 - \delta$,

$$\left|\left|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}^{(\ell)} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\right|\right|_\infty \leq O\left(\frac{\sigma r}{\sqrt{sd_2}} \sqrt{\frac{\widetilde{\mu}^3 \log d_2}{p}}\right).$$

where $d_2 = \min(|\mathcal{M}^{(\ell,i)}|, |\mathcal{N}^{(\ell,i)}|)$, $\widetilde{\mu}$ is the incoherence factor of the matrix $\mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$ and r is the rank of the matrix bounded from above by the number of clusters. In order for the right hand side to be less than $\Delta_{\ell+1}$, we can set $sp = O\left(\frac{\sigma^2 r^2 \widetilde{\mu}^3 \log d_2}{\Delta_{\ell+1}^2 d_2}\right)$. Since the event $\mathcal{E}^{(\ell)}$ is true, we must have that $|\mathcal{M}^{(\ell,i)}| \geq \mathsf{N}/(\kappa\mathsf{C})$ ($\kappa = O(1)$ is the ratio of the sizes of maximum cluster and minimum cluster); hence $d_2 \geq \min\left(\frac{\mathsf{N}}{\kappa\mathsf{C}}, |\mathcal{N}^{(\ell,i)}|\right)$. Therefore, we must have that

$$m_\ell = O\left(\frac{\sigma^2 \mathsf{C}^2 \widetilde{\mu}^3 \log \mathsf{M}}{\Delta_{\ell+1}^2} \max\left(\mathsf{N}, \mathsf{MC}\right) \log^2(\mathsf{MNC}\delta^{-1})\right)$$

where we take a union bound over all sets comprising the partition of the users $[\mathsf{N}]$ (at most C of them). Finally, from Lemma 3, we know that $\widetilde{\mu}$ can be bounded from above by $\max(\mathsf{C}, 2\mu/\alpha)$ which we can use to say that

$$m_\ell = O\Big(\frac{\sigma^2 \mathsf{C}^2 (\mathsf{C} \vee \mu\alpha^{-1})^3 \log \mathsf{M}}{\Delta_{\ell+1}^2}\Big(\mathsf{N} \bigvee \mathsf{MC}\Big) \log^2(\mathsf{MNC}\delta^{-1})\Big)\Big)$$

to complete the proof of the lemma. □

In the following part of the analysis, we will repeatedly condition on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ which are described/reiterated below:

1. The event $\mathcal{E}^{(\ell)}$ is true when the properties $(1-4)$ described at the beginning of Section E are satisfied by the algorithm.

2. We will denote the event described in Lemma 12 equation 21 to be $\mathcal{E}_2^{(\ell)}$. In other words, if $\mathcal{E}_2^{(\ell)}$ is true, then the algorithm has successfully computed an estimate $\widetilde{\mathbf{P}}^{(\ell)} \in \mathbb{R}^{\mathsf{N}\times\mathsf{M}}$ such that

$$\left\| \widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}}^{(\ell)} - \mathbf{P}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}} \right\|_\infty \le \Delta_{\ell+1} \text{ for all } i \in [a_\ell] \text{ satisfying } \left| \mathcal{N}^{(i,\ell)} \right| \ge \gamma\mathsf{C}.$$

Fix any $i \in [a_\ell]$. For each user $u \in \mathcal{M}^{(\ell,i)}$, let us denote a set of good arms for the user u by $\mathcal{T}_u^{(\ell)} \equiv \{j \in \mathcal{N}^{(\ell,i)} \mid \max_{j' \in \mathcal{N}^{(\ell,i)}} \widetilde{\mathbf{P}}_{uj'}^{(\ell)} - \widetilde{\mathbf{P}}_{uj}^{(\ell)} \le 2\Delta_{\ell+1}\}$. If we condition on the event $\mathcal{E}_2^{(\ell)}$, then we can show the following statements to be true:

Lemma (Restatement of Lemma 8). *Condition on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ being true. In that case, for every user $u \in \mathcal{B}^{(\ell)}$, the arm with the highest reward $\mathbf{P}_{u\pi_u(1)}$ must belong to the set $\mathcal{T}_u^{(\ell)}$. Moreover, $\max_{s,s' \in \mathcal{T}_u^{(\ell)}} |\mathbf{P}_{us} - \mathbf{P}_{us'}| \le 4\Delta_{\ell+1}$.*

Proof. Let us fix a user $u \in \mathcal{M}^{(\ell,i)}$ with active set of arms $\mathcal{N}^{(\ell,i)}$. Recall that $\pi_u(1) = \mathsf{argmax}_j \mathbf{P}_{uj}$ and let us denote $t_1 = \mathsf{argmax}_{j \in \mathcal{N}^{(\ell,i)}} \widetilde{\mathbf{P}}_{uj}^{(\ell)}$ for brevity of notation. Now, we will have

$$\widetilde{\mathbf{P}}_{ut_1} - \widetilde{\mathbf{P}}_{u\pi_u(1)}^{(\ell)} = \widetilde{\mathbf{P}}_{ut_1}^{(\ell)} - \mathbf{P}_{ut_1} + \mathbf{P}_{ut_1} - \mathbf{P}_{u\pi_u(1)} + \mathbf{P}_{u\pi_u(1)} - \widetilde{\mathbf{P}}_{u\pi_u(1)}^{(\ell)} \le 2\Delta_{\ell+1}$$

which implies that $\pi_u(1) \in \mathcal{T}_u^{(\ell)}$. Here we used the fact that $\widetilde{\mathbf{P}}_{ut_1}^{(\ell)} - \mathbf{P}_{ut_1} \le \Delta_{\ell+1}$, $\mathbf{P}_{u\pi_u(1)} - \widetilde{\mathbf{P}}_{u\pi_u(1)}^{(\ell)} \le \Delta_{\ell+1}$ and $\mathbf{P}_{ut_1} - \mathbf{P}_{u\pi_u(1)} \le 0$.

Next, notice that for any $s, s' \in \mathcal{T}_u^{(\ell)}$

$$\mathbf{P}_{us} - \mathbf{P}_{us'} = \mathbf{P}_{us} - \widetilde{\mathbf{P}}_{us}^{(\ell)} + \widetilde{\mathbf{P}}_{us}^{(\ell)} - \widetilde{\mathbf{P}}_{ut_1}^{(\ell)} + \widetilde{\mathbf{P}}_{ut_1}^{(\ell)} - \widetilde{\mathbf{P}}_{us'}^{(\ell)} + \widetilde{\mathbf{P}}_{us'}^{(\ell)} - \mathbf{P}_{us'} \le 4\Delta_{\ell+1}.$$

□

Again, fix any $i \in [a_\ell]$ such that $\left| \mathcal{N}^{(\ell,i)} \right| \ge \gamma\mathsf{C}$. Consider a graph $\mathcal{G}^{(\ell,i)}$ whose nodes are given by the users in $\mathcal{M}^{(\ell,i)}$. Now, we draw an edge between two users $u, v \in \mathcal{M}^{(\ell,i)}$ if $\mathcal{T}_u^{(\ell)} \cap \mathcal{T}_v^{(\ell)} \ne \Phi$ and $\left| \widetilde{\mathbf{P}}_{ux}^{(\ell)} - \widetilde{\mathbf{P}}_{vx}^{(\ell)} \right| \le 2\Delta_{\ell+1}$ for all arms $x \in \mathcal{N}^{(\ell,i)}$.

Lemma (Restatement of Lemma 9). *Fix any $i \in [a_\ell]$ such that $\left| \mathcal{N}^{(\ell,i)} \right| \ge \gamma\mathsf{C}$. Conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, nodes in $\mathcal{M}^{(\ell,i)}$ corresponding to the same cluster form a clique. Also, users in each connected component of the graph $\mathcal{G}^{(\ell,i)}$ form a nice subset.*

Proof. For any two users $u, v \in \mathcal{M}^{(\ell,i)}$ belonging to the same cluster, consider an arm $x \in \mathcal{N}^{(\ell,i)}$. We must have

$$\widetilde{\mathbf{P}}_{ux}^{(\ell)} - \widetilde{\mathbf{P}}_{vx}^{(\ell)} = \widetilde{\mathbf{P}}_{ux}^{(\ell)} - \mathbf{P}_{ux} + \mathbf{P}_{ux} - \mathbf{P}_{vx} + \mathbf{P}_{vx} - \widetilde{\mathbf{P}}_{vx}^{(\ell)} \le 2\Delta_{\ell+1}.$$

In order to prove the next statement, consider two users $u, v \in \mathcal{M}^{(\ell,i)}$ that belongs to different clusters $\mathcal{P}, \mathcal{Q}$ respectively. Note that since the event $\mathcal{E}^{(\ell)}$ is true, $\mathcal{M}^{(\ell,i)}$ is a union of clusters comprising $\mathcal{P}, \mathcal{Q}$. Furthermore, we have already established that nodes in $\mathcal{G}^{(\ell,i)}$ (users in $\mathcal{M}^{(\ell,i)}$) restricted to the same cluster form a clique. There every connected component of the graph $\mathcal{G}^{(\ell,i)}$ can be represented as a union of a subset of clusters.

□

Lemma (Restatement of Lemma 10). *Fix any $i \in [a_\ell]$ such that $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma\mathsf{C}$. Consider two users $u, v \in \mathcal{M}^{(\ell,i)}$ having an edge in the graph $\mathcal{G}^{(\ell,i)}$. Conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, we must have*

$$\max_{x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_v^{(\ell)}} |\mathbf{P}_{ux} - \mathbf{P}_{uy}| \leq 16\Delta_{\ell+1} \text{ and } \max_{x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_v^{(\ell)}} |\mathbf{P}_{vx} - \mathbf{P}_{vy}| \leq 16\Delta_{\ell+1}$$

Proof. From the construction of $\mathcal{G}^{(\ell,i)}$, we know that users $u, v \in \mathcal{M}^{(\ell,i)}$ have an edge if $\left|\widetilde{\mathbf{P}}_{ux}^{(\ell)} - \widetilde{\mathbf{P}}_{vx}^{(\ell)}\right| \leq 2\Delta_{\ell+1}$ (implying that $|\mathbf{P}_{ux} - \mathbf{P}_{vx}| \leq 4\Delta_{\ell+1}$) for all $x \in \mathcal{N}^{(\ell,i)}$ and $\mathcal{T}_u \cap \mathcal{T}_v \neq \Phi$. Suppose $z \in \mathcal{T}_u \cap \mathcal{T}_v$. Now, for any pair of arms $x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_v^{(\ell)}$; in that case, we have

$$\mathbf{P}_{ux} - \mathbf{P}_{uy} = \mathbf{P}_{ux} - \mathbf{P}_{uz} + \mathbf{P}_{uz} - \mathbf{P}_{vz} + \mathbf{P}_{vz} - \mathbf{P}_{vy} + \mathbf{P}_{vy} - \mathbf{P}_{uy} \leq 16\Delta_{\ell+1}$$

where we used Lemma 8.

$\square$

Lemma (Restatement of Lemma 11). *Fix any $i \in [a_\ell]$ such that $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma\mathsf{C}$. Consider two users $u, v \in \mathcal{M}^{(\ell,i)}$ having a path in the graph $\mathcal{G}^{(\ell,i)}$. Conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$, we must have*

$$\max_{x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_v^{(\ell)}} |\mathbf{P}_{ux} - \mathbf{P}_{uy}| \leq 32\mathsf{C}\Delta_{\ell+1} \text{ and } \max_{x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_v^{(\ell)}} |\mathbf{P}_{vx} - \mathbf{P}_{vy}| \leq 32\mathsf{C}\Delta_{\ell+1}.$$

Proof. Consider 3 users $u, v, w \in \mathcal{M}^{(\ell,i)}$ such that u, v have an edge and similarly, v, w have an edge. From Lemma 8, we have that

$$\mathcal{T}_u^{(\ell)} \cap \mathcal{T}_v^{(\ell)} \neq \Phi \text{ and } |\mathbf{P}_{ux} - \mathbf{P}_{vx}| \leq 4\Delta_{\ell+1} \text{ for all } x \in \mathcal{N}^{(\ell,i)}$$
$$\mathcal{T}_v^{(\ell)} \cap \mathcal{T}_w^{(\ell)} \neq \Phi \text{ and } |\mathbf{P}_{vx} - \mathbf{P}_{wx}| \leq 4\Delta_{\ell+1} \text{ for all } x \in \mathcal{N}^{(\ell,i)}$$

Let $z \in \mathcal{T}_u^{(\ell)} \cap \mathcal{T}_v^{(\ell)}$ and $z' \in \mathcal{T}_v^{(\ell)} \cap \mathcal{T}_w^{(\ell)}$. Therefore, for any $x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_z^{(\ell)}$, we must have

$$\mathbf{P}_{ux} - \mathbf{P}_{uy} = \mathbf{P}_{ux} - \mathbf{P}_{uz} + \mathbf{P}_{uz} - \mathbf{P}_{vz} + \mathbf{P}_{vz} - \mathbf{P}_{vz'} + \mathbf{P}_{vz'} - \mathbf{P}_{wz'} + \mathbf{P}_{wz'} - \mathbf{P}_{wy} + \mathbf{P}_{wy} - \mathbf{P}_{vy} + \mathbf{P}_{vy} - \mathbf{P}_{uy}$$
$$\leq 32\Delta_{\ell+1}$$

Note that the shortest path between the two users $u, v \in \mathcal{M}^{(\ell,i)}$ must be a sequence of at most $2\mathsf{C}$ nodes. Now, applying the above analysis at most $2\mathsf{C}$ times, we get statement of the Lemma. In other words, consider a path connecting two users u, v as denoted by $u = a_1, a_2, \ldots, v = a_\mathsf{L}$. Let us denote $z_i = \mathcal{T}_{a_i}^{(\ell)} \cap \mathcal{T}_{a_{i+1}}^{(\ell)}$ (definition of edge). For $x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_v^{(\ell)}$, we will have

$$\mathbf{P}_{ux} - \mathbf{P}_{uy} = \mathbf{P}_{ux} - \mathbf{P}_{uz_1} + \sum_{i=1}^{\mathsf{L}-1} \left(\mathbf{P}_{a_i z_i} - \mathbf{P}_{a_{i+1} z_i} + \mathbf{P}_{a_{i+1} z_i} - \mathbf{P}_{a_{i+1} z_{i+1}} \right) + \mathbf{P}_{v z_{\mathsf{L}-1}} - \mathbf{P}_{v z'} + \sum_{j=\mathsf{L}}^{j=2} \left(\mathbf{P}_{a_j z'} - \mathbf{P}_{a_{j-1} z'} \right)$$
$$\leq 16\mathsf{L}\Delta_{\ell+1}.$$

Since the path connecting the two users can be of length at most $2\mathsf{C} - 1$ (conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$), the proof of our lemma is complete. $\square$

For the subsequent iteration indexed by $\ell + 1$, we compute the updated groups of users $\mathcal{M}^{(\ell+1)}$ in the following way: each set corresponds to the connected components of the graphs $\{\mathcal{G}^{(\ell,i)}\}$ for those indices $i \in [a_\ell]$ where $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma\mathsf{C}$ plus the groups of users $\mathcal{M}^{(\ell,i)}$ where $\left|\mathcal{N}^{(\ell,i)}\right| \leq \mathsf{C}$. More precisely, let $\mathcal{T} \subseteq [a_\ell]$ be the subset of indices for which $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma\mathsf{C}$; $\{\mathcal{G}^{(\ell,i,j)}\}$ be the connected components of the graph $\mathcal{G}^{(\ell,i)}$ for $i \in \mathcal{T}$. In that case, $\mathcal{M}^{(\ell+1)} = \{\mathcal{G}^{(\ell,i,j)} \mid i \in \mathcal{T}, \mathcal{G}^{(\ell,i,j)} \text{ is a connected component of graph } \mathcal{G}^{(\ell,i)}\} + \{\mathcal{M}^{(\ell,i)} \mid i \in [a_\ell] \setminus \mathcal{T}\}$. Similarly, we update the family of sets of active arms as follows: for users corresponding to each connected component $\mathcal{M}^{(\ell+1,s)} = \mathcal{G}^{(\ell,i,j)}$ of some graph, we define the active set of arms $\mathcal{N}^{(\ell+1,s)}$ to be $\cup_{u \in \mathcal{G}^{(\ell,i,j)}} \mathcal{T}_u^{(\ell)}$ and for each group $\{\mathcal{M}^{(\ell,i)}\}_{i \in [a_\ell] \setminus \mathcal{T}}$, we keep the corresponding set of active arms $\{\mathcal{N}^{(\ell,i)}\}_{i \in [a_\ell] \setminus \mathcal{T}}$ same. With $a_{\ell+1} = \left|\mathcal{M}^{(\ell+1)}\right|$, we will also update $\mathcal{B}^{(\ell+1)} = \bigcup_{i \in [a_{\ell+1}] \mid |\mathcal{N}^{(\ell+1,i)}| \geq \gamma\mathsf{C}} \mathcal{M}^{(\ell+1,i)}$ to be the set of users with more than C active arms.

Lemma 13. *Condition on the events $\mathcal{E}^{(\ell)}$ being true. In that case, with probability $1 - \mathsf{T}^{-4}$, with the groups of users $\mathcal{M}^{(\ell+1)}$ and their respective group of arms given by $\mathcal{N}^{(\ell+1)}$ being updated as described above and $\Delta_{\ell+1} = \epsilon_\ell/64\mathsf{C}$, the event $\mathcal{E}^{(\ell+1)}$ is also going to be true with $\epsilon_{\ell+1} \leq \epsilon_\ell/2$.*

Proof. Conditioned on the event $\mathcal{E}^{(\ell)}$ being true, the event $\mathcal{E}_2^{(\ell)}$ holds true with probability with $1 - \mathsf{T}^{-4}$ (by substituting $\delta = \mathsf{T}^{-4}$ in Lemma 12). Now, conditioned on the event $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ being true, the properties $(1-4)$ hold true at the beginning of the $(\ell+1)^{\text{th}}$ phase as well with our construction of $\mathcal{M}^{(\ell+1)}, \mathcal{N}^{(\ell+1)}$. For the $(\ell+1)^{\text{th}}$ phase from Lemma 11, we know that for any pair of users u, v in the same cluster $\mathcal{M}^{(\ell+1,i)}$ in the updated set of clusters $\mathcal{M}^{(\ell+1)}$, we must have $\max_{x\in\mathcal{T}_u^{(\ell)}, y\in\mathcal{T}_v^{(\ell)}} |\mathbf{P}_{ux} - \mathbf{P}_{vy}| \leq 32\mathsf{C}\Delta_{\ell+1}$. From Lemma 8 we know that $\operatorname{argmax}_j \mathbf{P}_{uj} \in \mathcal{T}_u^{(\ell)} \subseteq \mathcal{N}^{(\ell,i)}$ where $\mathcal{N}^{(\ell,i)}$ is the active set of arms for users in $\mathcal{M}^{(\ell,i)}$ in the updated set $\mathcal{M}$. For $\ell > 1$, we will set $\Delta_{\ell+1} = \epsilon_\ell/64\mathsf{C}$ which would give us that $\epsilon_{\ell+1} = \epsilon_\ell/2$. Finally, also note that we maintain the set of users $\mathcal{B}^{(\ell+1)}$ as stipulated in Property 4 for the beginning of the $(\ell+1)^{\text{th}}$ phase. $\qquad\square$

Proof of Theorem 1. We condition on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ being true for all ℓ. The probability that there exists any ℓ such that the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ is false is $O(\mathsf{T}^{-4})$ (by setting $\delta = \mathsf{T}^{-4}$ in the proof of Lemma 12); hence the probability that $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ is true for all ℓ is at least $1 - O(\mathsf{T}^{-3})$ (the total number of iterations can be at most T). Let us also denote the set of rounds in phase ℓ by $\mathcal{T}_\ell \subseteq [\mathsf{T}]$ (therefore $|\mathcal{T}_\ell| = m_\ell$). Let us compute the regret $\sum_{t\in\mathcal{T}^{(\ell)}} \mathbf{P}_{u(t)\pi_{u(t)}(1)} - \sum_{t\in\mathcal{T}^{(\ell)}} \mathbf{P}_{u(t),\rho(t)}$ restricted to the rounds in $\mathcal{T}^{(\ell)}$ conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ being true for all ℓ. We can bound the regret quantity in the ℓth phase from above by the sum of two quantities: 1) the first quantity is the regret incurred by users in $\mathcal{B}^{(\ell)}$ with active arms more than C 2) the second quantity is the regret incurred by the UCB algorithm played separately for each user $u \in \mathsf{N} \setminus [\mathcal{B}^{(\ell)}]$ with active arms less than $\gamma\mathsf{C}$. Let us denote the regret incurred by such an user u in the ℓth phase by $\mathsf{Reg}_{\mathsf{UCB}}(u, \mathcal{T}_u^{(\ell)})$ where $\mathcal{T}_u^{(\ell)}$ is the number of rounds in the ℓth phase where user u pulled an arm.

The first quantity can be bounded from above by $m_\ell\epsilon_\ell$ while the second quantity can be bounded by using standard results in the literature. Substituting from Lemma 12, we have that

$$\mathbb{E}\Big(\sum_{t\in\mathcal{T}^{(\ell)}} \mathbf{P}_{u(t)\pi_{u(t)}(1)} - \sum_{t\in\mathcal{T}^{(\ell)}} \mathbf{P}_{u(t),\rho(t)} \mid \bigcap_\ell \mathcal{E}^{(\ell)} \bigcap_\ell \mathcal{E}_2^{(\ell)} \Big)$$
$$= O\Big(\frac{\epsilon_\ell\sigma^2\mathsf{C}^2(\mathsf{C} \vee \mu\alpha^{-1})^3 \log\mathsf{M}}{\Delta_{\ell+1}^2}\Big(\mathsf{N} \vee \mathsf{MC}\Big) \log^2(\mathsf{MNC}\delta^{-1})\Big) + \sum_{u\in[\mathsf{N}]\setminus[\mathcal{B}^{(\ell)}]} \mathsf{Reg}_{\mathsf{UCB}}(u, \mathcal{T}_u^{(\ell)})$$

For simplicity, let us denote $\mathsf{V} = \sigma^2\mathsf{C}^2(\mathsf{C} \vee \mu\alpha^{-1})^3\Big(\mathsf{N} \vee \mathsf{MC}\Big) \log^3(\mathsf{ABCT})$. We can now bound the regret as follows (after removing the conditioning on the events $\bigcap_\ell \mathcal{E}^{(\ell)} \bigcap_\ell \mathcal{E}_2^{(\ell)}$):

$$\mathbb{E}\Big(\sum_{t\in[\mathsf{T}]} \mathbf{P}_{u(t)\pi_{u(t)}(1)} - \sum_{t\in[\mathsf{T}]} \mathbf{P}_{u(t),\rho(t)} \Big) = \mathbb{E}\Big(\sum_\ell \Big(\sum_{t\in\mathcal{T}^{(\ell)}} \mathbf{P}_{u(t)\pi_{u(t)}(1)} - \sum_{t\in\mathcal{T}^{(\ell)}} \mathbf{P}_{u(t),\rho(t)} \Big)\Big)$$
$$= \sum_\ell O\Big(\epsilon_\ell m_\ell \mid \bigcap_\ell \mathcal{E}^{(\ell)} \bigcap_\ell \mathcal{E}_2^{(\ell)}\Big) + O(\mathsf{T}^{-3}\|\mathbf{P}\|_\infty) + \sum_\ell \sum_{u\in\mathsf{N}\setminus[\mathcal{B}^{(\ell)}]} \mathsf{Reg}_{\mathsf{UCB}}(u, \mathcal{T}_u^{(\ell)})$$

The last term of the regret can be bounded from above $\sum_{u\in[\mathsf{N}]} \mathsf{Reg}_{\mathsf{UCB}}(u, \mathsf{T}_b)$ where $\mathcal{T}_b$ is the number of rounds user u pulled an arm according to the UCB algorithm; hence, $\sum_{u\in[\mathsf{N}]} \mathsf{Reg}_{\mathsf{UCB}}(u, \mathsf{T}_b) \leq \sum_{u\in[\mathsf{N}]} \sqrt{\mathsf{T}_b\gamma\mathsf{C} \log\mathsf{T}} \cdot \sigma \leq \sqrt{\gamma\mathsf{NCT} \log\mathsf{T}} \cdot \sigma$ by using the Cauchy Schwartz inequality. Moving on, we can decompose the first term regret as follows (we use $\Delta_{\ell+1} = \epsilon_\ell/64\mathsf{C}$):

$$O\Big(\sum_{\ell:\epsilon_\ell\leq\Phi} \epsilon_\ell m_\ell \mid \mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)} \text{ is true for all } \ell \Big) + O\Big(\sum_{\ell:\epsilon_\ell>\Phi} \epsilon_\ell\mathsf{V}\Delta_{\ell+1}^{-2} \mid \mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)} \text{ is true for all } \ell \Big) = \mathsf{T}\Phi + O\Big(\sum_{\ell:\epsilon_\ell>\Phi} \mathsf{C}^2\mathsf{V}\epsilon_\ell^{-1} \Big)$$

We choose $\epsilon_\ell = C'2^{-\ell} \min\Big(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log\mathsf{N}}\Big)$ (so that the condition on $\sigma > 0$ in Lemma 1 is automatically satisfied for all ℓ) for some constant $C' > 0$, the maximum number of phases ℓ for which $\epsilon_\ell > \Phi$ can be bounded from above by

$J = O\Big(\log \Big(\frac{1}{\Phi} \min \Big(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log N} \Big) \Big) \Big)$. Moreover, the constraints on p, σ present in Lemma 1 can be satisfied with such a choice of ϵ_ℓ for all ℓ. Hence, with $V = \sigma^2 C^2 (C \vee \mu\alpha^{-1})^3 \Big(N \vee MC \Big) \log^3(ABCT)$, we have

$$\text{Reg}(T) \leq O(T^{-3}\|\mathbf{P}\|_\infty) + O(T\Phi) + O\Big(JVC^2\Phi^{-1} \Big) + O(\sqrt{NCT \log T} \cdot \sigma)$$
$$= O(T^{-3}\|\mathbf{P}\|_\infty) + O(CJ\sqrt{TV}) + O(\sqrt{NCT \log T} \cdot \sigma)$$

where we substituted $\Phi = \sqrt{VC^2 T^{-1}}$ and hence $J = O\Big(\log \Big(\frac{1}{\sqrt{VC^2 T^{-1}}} \min \Big(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log N} \Big) \Big) \Big)$ in the final step. $\qquad\square$

Proof of Theorem 2. For every cluster $c \in [C]$, let us define the subset of arms $\mathcal{G}_{c,\ell} \equiv \{ j \in [M] \mid \epsilon_\ell \leq \big| \mathbf{P}_{uj} - \mathbf{P}_{u\pi_u(1)} \big| \leq \epsilon_{\ell-1} \forall u \in \mathcal{C}^{(c)} \}$ for all $\ell > 1$ and $\mathcal{G}_{c,1} \equiv \{ j \in [M] \mid \epsilon_1 \leq \big| \mathbf{P}_{uj} - \mathbf{P}_{u\pi_u(1)} \big| \forall u \in \mathcal{C}^{(c)} \}$; $\mathcal{G}_{c,\ell}$ ($\mathcal{G}_{c,1}$) corresponds to the subset of arms having a sub-optimality gap that is between $\epsilon_{\ell-1}$ and ϵ_ℓ (greater than ϵ_1) for all users belonging to the cluster $\mathcal{C}^{(c)}$. There is no ambiguity in the definition since all users in the same cluster $\mathcal{C}^{(c)}$ have the same mean rewards over all arms.

Let us also define $\mathcal{H}_c \equiv \bigcup_{\ell > 1} \text{argmin}_{j \in \mathcal{G}_{c,\ell}} \big| \mathbf{P}_{uj} - \mathbf{P}_{u\pi_u(1)} \big|$ with the understanding that whenever $\mathcal{G}_{c,\ell} = \Phi$, there is no argmin to be counted in the set. For brevity of notation, let $\Psi_{c,a} \triangleq \mathbf{P}_{u\pi_u(1)} - \mathbf{P}_{ua}$ be the sub-optimality gap in the reward of arm a for any user u in cluster c, and $T_{c,a}^{(1)}$ be the number of times arm a has been pulled by the users in cluster c during the phases indexed by ℓ when the users in cluster c belonged to $\mathcal{B}^{(\ell)}$; $T_{u,a}^{(2)}$ is the number of times arm a has been pulled by the user u according to the UCB algorithm i.e. when users in cluster c belonged to $[N] \setminus \mathcal{B}^{(\ell)}$.

Since the length of the phases increases exponentially with ℓ, hence the total number of phases is $\widetilde{O}(1)$; hence the size of $\mathcal{H}_c$ is at most $\widetilde{O}(1)$. Again, for all users $u \in [N]$ who participated in the UCB algorithm, let us denote $\widehat{\mathcal{H}}_u$ to be the set of arms ($\big| \widehat{\mathcal{H}}_u \big| \leq C$ and $\widehat{\mathcal{H}}_u \supseteq \pi_u(1)$; recall from Lemma 8 that the best arm $\pi_u(1)$ always belongs to the active set of arms) that were used in the UCB algorithm; evidently, the regret incurred due to set $\widehat{\mathcal{H}}_u$ will be dominated by the set $\{\pi_u(i)\}_{i=1}^C$ which corresponds to the best C arms for user u. We can decompose the regret by using the standard regret decomposition i.e.

$$\text{Reg}(T) = \mathbb{E}\left[\sum_{c \in [C], a \in [M]} \Psi_{c,a} T_{c,a}^{(1)} \right] + \sum_{u \in [N], a \in \{\pi_u(s)\}_{s=1}^{\gamma|C|}} \Psi_{u,a} \mathbb{E}T_{u,a}^{(2)}$$

$$\leq \mathbb{E}\left[\sum_{c \in [C], a \in [M]} \Psi_{c,a} T_{c,a}^{(1)} \Big| \bigcap_\ell \mathcal{E}^{(\ell)} \right] + \Pr\left(\bigcup_\ell (\mathcal{E}^{(\ell)})^c \right) \mathbb{E}\left[\sum_{c \in [C], a \in [M]} \Psi_{c,a} T_{c,a}^{(1)} \Big| \bigcup_\ell (\mathcal{E}^{(\ell)})^c \right]$$

$$+ \sum_{u \in [N], a \in \{\pi_u(s)\}_{s=1}^{\gamma|C|}} \Psi_{u,a} \mathbb{E}T_{u,a}^{(2)}$$

We now show the following lemma:

Lemma 14. *Fix any $\ell > 1$ and cluster $c \in [C]$. We must have $\mathbb{E}[\sum_{a \in \mathcal{G}_{c,\ell}} T_{c,a}^{(1)} \mid \cup_\ell (\mathcal{E}^{(\ell)})^c] = C^{-1}O\Big(T^{-3} \Big)$ and $\mathbb{E}[\sum_{a \in \mathcal{G}_{c,\ell}} T_{c,a}^{(1)} \mid \cap_\ell \mathcal{E}^{(\ell)}] = C^{-1}O\Big(\frac{C^2 V}{\epsilon_{\ell-1}^2} \Big)$ provided that $\mathcal{G}_{c,\ell} \neq \Phi$.*

Proof. Fix cluster $c \in [C]$. From definition, we know that all arms a in $\mathcal{G}_{c,\ell}$ satisfy the following for all users $u \in \mathcal{C}^{(c)}$: $\epsilon_\ell \leq \Psi_{c,a} = |\mathbf{P}_{ua} - \mathbf{P}_{u\pi_u(1)}| \leq \epsilon_{\ell-1}$ for $\ell > 1$. In that case, with probability at least $1 - T^{-3}$, the event $\mathcal{E}^{(j)}$ is true for all j implying that the algorithm is ϵ_ℓ-good (see Lemma 13). Hence, we must have $\Pr(\cup_\ell (\mathcal{E}^{(\ell)})^c) \mathbb{E}[\sum_{a \in \mathcal{G}_{c,\ell}} T_{c,a}^{(1)} \mid \cup_\ell (\mathcal{E}^{(\ell)})^c] = C^{-1}O\Big(T^{-2} \Big)$. Therefore, conditioning on all $\mathcal{E}^{(j)}$ being true, by definition using property (3), at the beginning of the ℓth phase, if $u \in \mathcal{C}^{(c)} \cap \mathcal{M}^{(\ell,i)}$, it must be the case that $a \notin \mathcal{N}^{(\ell,i)}$ for all $a \in \mathcal{G}_{c,\ell}$. Hence, we must have (by plugging in the sample complexity in Lemma 12 with with $V = \sigma^2 C^2 (C \vee \mu\alpha^{-1})^3 \Big(N \vee MC \Big) \log^3(ABCT)$ and $\Delta_{\ell+1} = \epsilon_\ell / 40C$)

$$\mathbb{E}[\sum_{a\in\mathcal{G}_{c,\ell}}\mathsf{T}_{c,a}^{(1)}\mid\cap\mathcal{E}^{(\ell)}]\leq\sum_{j=1}^{\ell}\frac{m_j}{\mathsf{C}}=\frac{1}{\mathsf{C}}O\Big(\sum_{j=1}^{\ell-1}\frac{\mathsf{V}}{\Delta_{j+1}^2}\Big)=\mathsf{C}^{-1}O\Big(\sum_{j=1}^{\ell-1}\frac{\mathsf{C}^2\mathsf{V}}{\epsilon_j^2}\Big)=\mathsf{C}^{-1}O\Big(\frac{\mathsf{C}^2\mathsf{V}}{\epsilon_{\ell-1}^2}\Big)$$

where we substituted the fact that $\epsilon_\ell=C'2^{-\ell}\min\left(\|\mathbf{P}\|_\infty,\frac{\sigma\sqrt{\mu}}{\log N}\right)$ for some constant $C'>0$. $\qquad\square$

Note that we will have the following set of equations; we use from definition that for every arm $a\in\mathcal{G}_{c,\ell}$, there exists a representative arm $\hat{a}$ of $\mathcal{G}_{c,\ell}$ in $\mathcal{H}_c$ such that $\Psi_{c,\hat{a}}\leq\min(\epsilon_{\ell-1},2\Psi_{c,a})$ and $\epsilon_{\ell-1}\leq 2\Psi_{c,\hat{a}}$.

$$\Pr(\cup_\ell(\mathcal{E}^{(\ell)})^c)\Big(\sum_{c\in[\mathsf{C}]}\sum_\ell\mathbb{E}[\sum_{a\in\mathcal{G}_{c,\ell}}\Psi_{c,a}\mathsf{T}_{c,a}^{(1)}\mid\cup_\ell(\mathcal{E}^{(\ell)})^c]\Big)$$

$$\leq\Pr(\cup_\ell(\mathcal{E}^{(\ell)})^c)\Big(\sum_{c\in[\mathsf{C}]}\sum_\ell(\max_{a\in\mathcal{G}_{c,\ell}}\Psi_{c,a})\sum_{a\in\mathcal{G}_{c,\ell}}\mathbb{E}[\mathsf{T}_{c,a}^{(1)}\mid\cup_\ell(\mathcal{E}^{(\ell)})^c]\Big)\leq\sum_{c\in[\mathsf{C}],a\in\mathcal{H}_c}\Psi_{c,a}\mathsf{C}^{-1}O\Big(\mathsf{T}^{-2}\Big).$$

and similarly, we will also have

$$\sum_{c\in[\mathsf{C}]}\sum_\ell\mathbb{E}[\sum_{a\in\mathcal{G}_{c,\ell}}\Psi_{c,a}\mathsf{T}_{c,a}^{(1)}\mid\cap_\ell\mathcal{E}^{(\ell)}]\leq\sum_{c\in[\mathsf{C}]}\sum_\ell(\max_{a\in\mathcal{G}_{c,\ell}}\Psi_{c,a})\sum_{a\in\mathcal{G}_{c,\ell}}\mathbb{E}[\mathsf{T}_{c,a}^{(1)}\mid\cap_\ell\mathcal{E}^{(\ell)}]$$

$$=\sum_{c\in[\mathsf{C}]}(\max_{a\in\mathcal{G}_{c,1}}\Psi_{c,a})\sum_{a\in\mathcal{G}_{c,1}}\mathbb{E}[\mathsf{T}_{c,a}^{(1)}\mid\cap_\ell\mathcal{E}^{(\ell)}]+\sum_{c\in[\mathsf{C}]}\sum_{\ell>1}\mathbb{E}[\sum_{a\in\mathcal{G}_{c,\ell}}\Psi_{c,a}\mathsf{T}_{c,a}^{(1)}\mid\cap_\ell\mathcal{E}^{(\ell)}]$$

$$=\|\mathbf{P}\|_\infty\cdot\frac{\mathsf{V}}{\epsilon_1^2}\mathbf{1}[\mathcal{G}_{c,1}\neq\Phi]+\sum_{c\in[\mathsf{C}]}\sum_{\ell>1:\mathcal{G}_{c,\ell}\neq\Phi}\mathsf{C}^{-1}O\Big(\frac{\mathsf{C}^2\mathsf{V}}{\epsilon_\ell}\Big)\leq\sum_{c\in[\mathsf{C}],a\in\mathcal{H}_c}\mathsf{C}^{-1}O\Big(\frac{\mathsf{C}^2\mathsf{V}}{\Psi_{c,a}}\Big).$$

Similarly, from well known analysis of UCB algorithm, we know that (recall that $\{\pi_u(s)\}_{s=1}^{|\gamma\mathsf{C}|}$ are the top $\gamma\mathsf{C}$ arms for the user u).

$$\sum_{u\in[\mathsf{N}],a\in\{\pi_u(s)\}_{s=1}^{|\gamma\mathsf{C}|}}\Psi_{u,a}\mathbb{E}\mathsf{T}_{u,a}^{(2)}\leq\mathsf{N}^{-1}O\Big(\sum_{u\in[\mathsf{N}],a\in\{\pi_u(s)\}_{s=1}^{|\gamma\mathsf{C}|}}\Big(\frac{\sigma\log\mathsf{T}}{\Psi_{u,a}}+3\Psi_{u,a}\Big)\Big)$$

Therefore, we can bound the regret from above as

$$\mathsf{Reg}(\mathsf{T})=\|\mathbf{P}\|_\infty\cdot\frac{\mathsf{V}}{\epsilon_1^2}\mathbf{1}[\mathcal{G}_{c,1}\neq\Phi]+\mathsf{C}^{-1}O\Big(\sum_{c\in[\mathsf{C}],a\in\mathcal{H}_c}\Psi_{c,a}\mathsf{T}^{-2}+\frac{\mathsf{C}^2\mathsf{V}}{\Psi_{c,a}}\Big)+\mathsf{N}^{-1}O\Big(\sum_{u\in[\mathsf{N}],a\in\{\pi_u(s)\}_{s=1}^{|\gamma\mathsf{C}|}}\Big(\frac{\sigma\log\mathsf{T}}{\Psi_{u,a}}+3\Psi_{u,a}\Big)\Big).$$

Loosely speaking, this bound translates as $\mathsf{Reg}(\mathsf{T})=\widetilde{O}((\mathsf{M}+\mathsf{N})/\Psi)$ where Ψ is the minimum sub-optimality gap; $\widetilde{O}(\cdot)$ hides factors in C and other logarithmic terms. $\qquad\square$

F   Proofs of Theorems 3, 4

F.1   Proofs of Theorem 3, Corollary 1

We first derive lower bounds for the following two settings: (a) known cluster assignment, and (b) known cluster rewards. The lower bound for MAB-LC follows by taking the maximum of these two bounds.

Known Cluster Assignments.   Suppose we know the mapping between users and clusters. In this setting, it is easy to see that the optimal strategy is to treat users within a cluster as a single super-user, and reduce the problem to that of solving C different multi-armed bandit problems (each corresponding to the C clusters). One could rely on the regret lower bound of stochastic MAB [Lattimore and Szepesvári, 2020, Cesa-Bianchi and Lugosi, 2006, Bubeck and Cesa-Bianchi, 2012] and provide the following informal proof for the regret lower-bound for our problem. We have C MAB instances, where the cth instance has M arms and occurs T_c times (note that T_c is a random variable). Since the regret of the cth instance is lower bounded by $0.05\mathbb{E}\left[\min(\sqrt{\mathsf{MT}_c},\mathsf{T}_c)\right]$, the overall regret is lower bounded $0.05\sum_{c\in[\mathsf{C}]}\mathbb{E}\left[\min(\sqrt{\mathsf{MT}_c},\mathsf{T}_c)\right]$.

We now make the above argument more formal. At a high level, the proof involves designing problem instances that are hard to separate. We then argue that any algorithm should suffer large regret on at least one of the problems.

- Partition users into C clusters. Let $\mathcal{N}_c$ be the set of users in cluster c.

- Choose indices $(a_1, \ldots, a_C)$ such that $a_i \in [M]$ for all i. Note that there are M^C possible such choices. We are going to define M^C problem instances each corresponding to a choice of $(a_1, \ldots, a_C)$. In these problems, each a_i corresponds to the optimal arm in cluster i. Define the mean rewards of the jth arm in i^{th} cluster as

$$
\mathbf{X}_{ij} = \begin{cases} \frac{1-\epsilon}{2}, & \text{if } j \neq a_i, \\ \frac{1+\epsilon}{2}, & \text{otherwise.} \end{cases}
$$

Let's call this problem instance $\text{Prob}_{a_1, \ldots a_C}$.

Define problem instance $\text{Prob}_{0,c,a_1,\ldots a_C}$ as follows. It is exactly equal to $\text{Prob}_{a_1,\ldots a_C}$ except for one difference. The rewards of all the arms in cluster c are set to $\frac{1-\epsilon}{2}$.

In the proof, we first consider deterministic algorithms. Using Fubini's theorem, these results can be easily extended to randomized algorithms [Bubeck and Cesa-Bianchi, 2012]. Next, we assume T_c, the number of appearances of cluster c, is a fixed quantity. The final results can simply be obtained by taking expectation over $\{\mathsf{T}_c\}_{c\in[C]}$. Let $T(c, a)$ be the number of times arm a has been pulled during the appearances of cluster c. Then the regret of the algorithm under $\text{Prob}_{a_1,\ldots a_C}$ can be written as

$$
\text{Reg}_{a_1,\ldots a_C}(\mathsf{T}) = \sum_{c\in[C]} \sum_{a\in[M]\setminus\{a_c\}} \epsilon \mathbb{E}_{a_1 \ldots a_C}[T(c,a)] = \sum_{c\in[C]} \epsilon \left(\mathsf{T}_c - \mathbb{E}_{a_1,\ldots a_C}[T(c,a_c)] \right).
$$

Let J_c be a random variable that is drawn according to the discrete distribution $\left(\frac{T(c,0)}{\mathsf{T}_c}, \ldots \frac{T(c,\mathsf{M}-1)}{\mathsf{T}_c} \right)$. Then

$$
\text{Reg}_{a_1,\ldots a_C}(\mathsf{T}) = \sum_{c\in[C]} \epsilon \mathsf{T}_c \left(1 - \mathbb{P}_{a_1,\ldots a_C}[J_c = a_c] \right).
$$

So, we have

$$
\frac{1}{M^C} \sum_{a_1,\ldots a_C} \text{Reg}_{a_1,\ldots a_C}(\mathsf{T}) = \frac{1}{M^C} \sum_{a_1,\ldots a_C} \sum_{c\in[C]} \epsilon \mathsf{T}_c \left(1 - \mathbb{P}_{a_1,\ldots a_C}[J_c = a_c] \right). \tag{13}
$$

Next, from Pinsker's inequality, we have

$$
\mathbb{P}_{a_1,\ldots a_C}[J_c = a_c] \leq \mathbb{P}_{0,c,a_1,\ldots a_C}[J_c = a_c] + \sqrt{\frac{1}{2} KL(\mathbb{P}_{0,c,a_1,\ldots a_C}, \mathbb{P}_{a_1,\ldots a_C})}
$$

$$
\overset{(a)}{\leq} \mathbb{P}_{0,c,a_1,\ldots a_C}[J_c = a_c] + \sqrt{\frac{1}{2} KL\left(\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2} \right) \mathbb{E}_{0,c,a_1,\ldots a_C}[T(c,a_c)]}.
$$

Inequality (a) simply follows from data processing inequality and the definition of KL divergence (also see Equation 3.20 of Bubeck and Cesa-Bianchi [2012]). Next, we take the average of the LHS and the average of the RHS in the above equation over all possible values of a_c. This gives us

$$
\frac{1}{M} \sum_{a'\in[M]} \mathbb{P}_{a_1,\ldots,a_c=a',\ldots a_C}[J_c = a'] \leq \frac{1}{M} \sum_{a'\in[M]} \mathbb{P}_{0,c,a_1,\ldots a_c=a',\ldots a_C}[J_c = a']
$$

$$
+ \frac{1}{M} \sum_{a'\in[M]} \sqrt{\frac{1}{2} KL\left(\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2} \right) \mathbb{E}_{0,c,a_1,\ldots a_c=a',\ldots a_C}[T(c,a')]}
$$

Now observe that $\text{Prob}_{0,c,a_1,\ldots a_c,\ldots a_C}$ doesn't depend on a_c (that is, $\text{Prob}_{0,c,a_1,\ldots a_c,\ldots a_C}$ are the same problem instances for all values of a_c). So

$$
\sum_{a'\in[M]} \mathbb{P}_{0,c,a_1,\ldots a_c=a',\ldots a_C}[J_c = a'] = 1.
$$

The second term in the RHS of the previous inequality can be bounded using Cauchy–Schwarz inequality

$$\frac{1}{M} \sum_{a' \in [M]} \sqrt{\frac{1}{2} \mathbb{E}_{0,c,a_1,\ldots a_c=a',\ldots a_C}[T(c,a')]} \leq \sqrt{\frac{1}{2M} \sum_{a' \in [M]} \mathbb{E}_{0,c,a_1,\ldots a_c=a',\ldots a_C}[T(c,a')]}$$

$$= \sqrt{\frac{1}{2M} T_c}$$

Substituting this in the previous inequality gives us

$$\frac{1}{M} \sum_{a' \in [M]} \mathbb{P}_{a_1,\ldots,a_c=a',\ldots a_C}[J_c = a'] \leq \frac{1}{M} + \sqrt{\frac{T_c}{2M} KL\left(\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}\right)}$$

Substituting the above inequality in Equation (13), we get

$$\max_{a_1,\ldots a_C} \mathsf{Reg}_{a_1,\ldots a_C}(T) \geq \frac{1}{M^C} \sum_{a_1,\ldots a_C} \mathsf{Reg}_{a_1,\ldots a_C}(T) \geq \sum_{c \in [C]} \epsilon T_c \left(1 - \frac{1}{M} - \sqrt{\frac{\epsilon T_c}{2M} \log \frac{1+\epsilon}{1-\epsilon}}\right).$$

Finally, taking expectation over T_c, and choosing best possible ϵ, we get the following lower bound on the worst-case regret

$$\max_{a_1,\ldots a_C} \mathsf{Reg}_{a_1,\ldots a_C}(T) \geq \max_{\epsilon} \sum_{c \in [C]} \epsilon \mathbb{E}\left[T_c \left(1 - \frac{1}{M} - \sqrt{\frac{\epsilon T_c}{2M} \log \frac{1+\epsilon}{1-\epsilon}}\right)\right]$$

$$\overset{(a)}{\geq} \max_{\epsilon} \sum_{c \in [C]} \epsilon \mathbb{E}\left[T_c \left(1 - \frac{1}{M} - \sqrt{\frac{\epsilon^2 T_c}{M}}\right)\right],$$

where (a) follows from the the fact that $\log \frac{1+\epsilon}{1-\epsilon} \leq 2\epsilon$. Rewriting the RHS in the above equation, we get

$$\max_{a_1,\ldots a_C} \mathsf{Reg}_{a_1,\ldots a_C}(T) \geq \max_{\epsilon} \sum_{c \in [C]} \epsilon \left(1 - \frac{1}{M}\right) \mathbb{E}[T_c] - \frac{\epsilon^2}{M^{1/2}} \mathbb{E}[T_c^{3/2}]$$

$$= \max_{\epsilon} \epsilon \left(1 - \frac{1}{M}\right) T - \epsilon^2 \frac{\sum_{c \in [C]} \mathbb{E}[T_c^{3/2}]}{M^{1/2}}.$$

We now focus on maximizing the RHS of the above equation. Note that the objective is quadratic in ϵ. So one can obtain an exact expression for the optimal value of ϵ. If $\left(M^{1/2} - M^{-1/2}\right) < 2\frac{\sum_{c \in [C]} \mathbb{E}[T_c^{3/2}]}{T}$, then the RHS is given by

$$\left(M^{1/2} - M^{-1/2}\right) \frac{T^2}{4 \sum_{c \in [C]} \mathbb{E}[T_c^{3/2}]}.$$

On the other hand, if $\left(M^{1/2} - M^{-1/2}\right) > 2\frac{\sum_{c \in [C]} \mathbb{E}[T_c^{3/2}]}{T}$, the RHS is given by

$$\left(1 - \frac{1}{M}\right) T - \frac{\sum_{c \in [C]} \mathbb{E}[T_c^{3/2}]}{M^{1/2}}.$$

For the special case where P_N is uniform and the clusters have the same size, we have $\mathbb{E}[T_c^{3/2}] = \Theta(T^{3/2} C^{-3/2})$. Substituting this in the above bounds, we get

$$\max_{a_1,\ldots a_C} \mathsf{Reg}_{a_1,\ldots a_C}(T) \geq \begin{cases} 0.05\sqrt{MCT}, & \text{if } T > 0.5MC \\ 0.05T & \text{otherwise.} \end{cases} \tag{14}$$

Known Cluster Rewards. We now consider the setting where we know the reward distributions of arms in each cluster. In this setting, it is easy to see that the optimal strategy is to solve a C-armed MAB problem for each user, where the C arms correspond to the best arms in each of the C clusters. That is, with the knowledge of cluster rewards, we can reduce the MAB-LC problem to that of solving N multi-armed bandit problems each with C arms. Here each MAB problem corresponds to a user. For this problem, we already derived lower bounds above. In particular, in the known cluster assignment setting, we derived lower bounds for solving C MAB instances each with M arms. So we could rely on the above lower bounds to derive lower bounds for the known cluster reward setting (we just replace C with N and M with C in the bounds). When P_N is uniform and the clusters have the same size, we could rely on Equation (14) and obtain the following minimax lower bounds

$$\inf\sup \mathsf{Reg}(T) \geq \begin{cases} 0.05\sqrt{NCT}, & \text{if } T > 0.5NC \\ 0.05T & \text{otherwise.} \end{cases} \tag{15}$$

For the more general case of non-uniform P_N and uneven cluster sizes, we obtain

$$\inf\sup \mathsf{Reg}(T) \geq \begin{cases} (C^{1/2} - C^{-1/2}) \frac{T^2}{4\sum_{b\in[N]} \mathbb{E}[T_b^{3/2}]}, & \text{if } (C^{1/2} - C^{-1/2}) < 2\frac{\sum_{b\in[N]}\mathbb{E}[T_b^{3/2}]}{T} \\ \left(1 - \frac{1}{C}\right) T - \frac{\sum_{b\in[N]}\mathbb{E}[T_b^{3/2}]}{C^{1/2}} & \text{otherwise.} \end{cases} \tag{16}$$

F.2 Proof of Theorem 4

Here is a high level idea of the proof. We find two bandit instances that are close enough to each other but the behaviour of any uniformly efficient algorithm is totally different in the two instances.

Background. Similar to lower bounding techniques used in the MAB literature, our proof relies on data processing inequality [Kaufmann, 2020]. Let μ_1, μ_2 be two stochastic K-armed bandit models. Let $\mathcal{F}_t$ be the σ-algebra generated by the observations available until round t. Let's suppose τ is the stopping time, and let I_τ be the information available until round τ. Then for any event $\mathcal{E} \in \mathcal{F}_\tau$, the data processing inequality tells us

$$KL(\mathbb{P}_{\mu_1}^{I_\tau}, \mathbb{P}_{\mu_2}^{I_\tau}) \geq KL(\mathbb{P}_{\mu_1}(\mathcal{E}), \mathbb{P}_{\mu_2}(\mathcal{E}))$$

Moreover, by definition of KL divergence we have

$$KL(\mathbb{P}_{\mu_1}^{I_\tau}, \mathbb{P}_{\mu_2}^{I_\tau}) \overset{(a)}{=} \mathbb{E}_{\mu_1}[L_\tau(\mu_1, \mu_2)] \overset{(b)}{=} \sum_{k=1}^{K} \mathbb{E}_{\mu_1}[T(k)] KL(\mu_{1,a}, \mu_{2,a}).$$

L_τ in the above equation is the log-likelihood ratio, $T(k)$ is the number of pulls of arm k, and $\mu_{1,a}$ is the distribution of rewards of arm a in bandit model μ_1.

Main Proof. Let $\pi : [N] \to [C]$ be the mapping from users to clusters and let $\mathbf{X} \in \mathbb{R}^{C \times M}$ be the mean rewards of arms in the clusters. Let $(\pi_1, \mathbf{X}_1), (\pi_2, \mathbf{X}_2)$ be two MAB-LC models. Let $a_{1,c}^*$ be the optimal arm in the cth cluster of the first model, and $a_{2,c}^*$ be the optimal arm in the second model. Let $\bar{C}$ be the set of clusters for which $a_{1,c}^* \neq a_{2,c}^*$. Let $T(c,a)$ be the number of times arm a has been pulled when cluster c appeared during the course of the algorithm. Let $\mathcal{E}_T$ be the following event

$$\mathcal{E}_T = \left\{ \sum_{c\in[\bar{C}]} T(c, a_{1,c}^*) \leq T/2 \right\}$$

Intuitively, $\mathcal{E}_T$ has a small probability under $(\pi_1, \mathbf{X}_1)$ where the optimal arms should be selected a lot. Moreover, $\mathcal{E}_T$ has large probability under $(\pi_2, \mathbf{X}_2)$ as the event only contains sub-optimal arms. This can be formally proved using Markov's inequality as follows

$$\mathbb{P}_1(\mathcal{E}_T) = \mathbb{P}_1\left(\sum_{c\in[\bar{C}]} T(c, a_{1,c}^*) \leq T/2\right) = \mathbb{P}_1\left(\sum_{c\in[\bar{C}]}\sum_{a\in[M]\backslash\{a_{1,c}^*\}} T(c,a) > T/2\right) \leq \frac{2\sum_{c\in[\bar{C}]}\sum_{a\in[M]\backslash\{a_{1,c}^*\}} \mathbb{E}_1[T(c,a)]}{T}$$

$$\mathbb{P}_2(\bar{\mathcal{E}}_T) = \mathbb{P}_2\left(\sum_{c\in[\bar{C}]} T(c, a_{1,c}^*) > T/2\right) \leq \frac{2\sum_{c\in[\bar{C}]} \mathbb{E}_2[T(c,a_{1,c}^*)]}{T}.$$

Let $\zeta_1 = \sum_{c\in[\bar{C}]}\sum_{a\in[M\setminus\{a^*_{1,c}\}]}\mathbb{E}_1[T(c,a)]$ and $\zeta_2 = \sum_{c\in[\bar{C}]}\mathbb{E}_2[T(c,a^*_{1,c})]$. Based on our assumptions of uniform efficiency of the algorithm, we know that $\zeta_1, \zeta_2 = o(T^\alpha)$ for all $\alpha \in (0,1]$. Next, observe that the KL divergence between two bernoulli distributions can be lower bounded as $KL(p,q) \geq (1-p)\log(1/(1-q)) - \log 2$ [Garivier and Kaufmann, 2021]. Using this, we have

$$KL(\mathbb{P}_1(\mathcal{E}_\mathsf{T}), \mathbb{P}_2(\mathcal{E}_\mathsf{T})) \geq \left(1 - \frac{2\zeta_1}{\mathsf{T}}\right)\log\left(\frac{\mathsf{T}}{2\zeta_2}\right) - \log 2 \approx \log T.$$

This shows that

$$\lim_{T\to\infty} \frac{KL(\mathbb{P}_1^{I_\mathsf{T}}, \mathbb{P}_2^{I_\mathsf{T}})}{\log \mathsf{T}} \geq 1. \tag{17}$$

We now rely on this result to prove Theorem 4. All we need to do is construct interesting bandit models that are hard to separate. The first set of models we construct is as follows. Let $(\pi_1, \mathbf{X}_1)$ be any MAB-LC model. Construct $(\pi_2, \mathbf{X}_2)$ from $(\pi_1, \mathbf{X}_1)$ as follows: $\pi_2 = \pi_1$, $\mathbf{X}_2$ is same as $\mathbf{X}_1$ for all cluster-arm pairs, except for one location. We take a sub-optimal arm a' in cluster c and make its mean reward to be slightly larger than the mean reward of the best arm in cluster c (i.e., $\mathbf{X}_2[c, a'] = \mathbf{X}_1[c, a^*_{1,c}] + \epsilon$ for some $\epsilon \to 0$). Applying the above result on this model pair gives us

$$\lim_{\mathsf{T}\to\infty} \frac{\mathbb{E}_1[T(c,a')]\, KL(\mathbf{X}_1[c,a'], \mathbf{X}_1[c,a^*_{1,c}])}{\log \mathsf{T}} \geq 1.$$

We now rely on the following upper bound on the KL divergence between two Bernoulli distributions

$$KL(p,q) \leq \frac{(p-q)^2}{q(1-q)}.$$

Using this in the previous inequality, we get

$$\lim_{\mathsf{T}\to\infty} \frac{\mathbb{E}_1[T(c,a')](\mathbf{X}_1[c,a^*_{1,c}] - \mathbf{X}_1[c,a'])}{\log \mathsf{T}} \geq \frac{\mathbf{X}_1[c,a^*_{1,c}](1 - \mathbf{X}_1[c,a^*_{1,c}])}{(\mathbf{X}_1[c,a^*_{1,c}] - \mathbf{X}_1[c,a'])}.$$

Note that this result holds for any c, a'. Summing over all possible values of c, a' gives us the required result in Theorem 4

$$\lim_{\mathsf{T}\to\infty} \frac{\mathsf{Reg}(\mathsf{T})}{\log \mathsf{T}} \geq \sum_{c\in[C]}\sum_{a\neq a^*_c} \frac{\mathbf{X}[c,a^*_c](1 - \mathbf{X}[c,a^*_c])}{(\mathbf{X}[c,a^*_c] - \mathbf{X}[c,a'])}. \tag{18}$$

Tighter Bounds. Note that the above lower bound didn't explicitly depend on the number of users N. We now derive a different bound that depends on N. Let $(\pi_1, \mathbf{X}_1)$ be any MAB-LC model. Construct $(\pi_2, \mathbf{X}_2)$ from $(\pi_1, \mathbf{X}_1)$ as follows: $\mathbf{X}_2$ is exactly same as $\mathbf{X}_1$. Moreover, π_2 is same as π_1 for all users except for a particular user b. To be precise, π_2 places b in a cluster that is different from $\pi_1(b)$. Let's call $\pi_1(b)$ as c and $\pi_2(b)$ as c'. One can show that Equation (17) holds for this setting. The proof of this uses similar arguments as those used to prove Equation (17) (the only thing that changes is our definition of the event $\mathcal{E}_\mathsf{T}$ which now includes all the users b that have different optimal arms across the two MAB-LC models). Applying Equation (17) to this setting gives us the following

$$\lim_{\mathsf{T}\to\infty} \frac{\sum_{a\in[\mathsf{M}]}\mathbb{E}_1[T(b,a)]\, KL(\mathbf{X}_1[c,a], \mathbf{X}_1[c',a])}{\log \mathsf{T}} \geq 1.$$

Here $T(b,a)$ is the number of times arm a has been pulled for user b. From the above inequality, the regret of user b can be lower bounded as (this follows from Holder's inequality: $\sum_i |a_i b_i| \leq (\sum_i |a_i|)\max_i |b_i|$)

$$\lim_{\mathsf{T}\to\infty} \frac{\sum_{a\in[\mathsf{M}]}\mathbb{E}_1[T(b,a)](\mathbf{X}_1[c,a^*_{1,c}] - \mathbf{X}_1[c,a])}{\log \mathsf{T}} \geq \min_{a\in[\mathsf{M}]} \frac{(\mathbf{X}_1[c,a^*_{1,c}] - \mathbf{X}_1[c,a])}{KL(\mathbf{X}_1[c,a], \mathbf{X}_1[c',a])}.$$

Furthermore, optimizing the RHS over the choice of c' gives us

$$\lim_{\mathsf{T}\to\infty} \frac{\sum_{a\in[\mathsf{M}]}\mathbb{E}_1[T(b,a)](\mathbf{X}_1[c,a^*_{1,c}] - \mathbf{X}_1[c,a])}{\log \mathsf{T}} \geq \max_{c'\neq c}\min_{a\in[\mathsf{M}]} \frac{(\mathbf{X}_1[c,a^*_{1,c}] - \mathbf{X}_1[c,a])}{KL(\mathbf{X}_1[c,a], \mathbf{X}_1[c',a])}.$$

This shows that the overall regret (for all the users) can be lower bounded as

$$\lim_{\mathsf{T}\to\infty} \frac{\mathsf{Reg}(\mathsf{T})}{\log \mathsf{T}} \geq \sum_{b\in[\mathsf{N}]} \max_{c'\neq\pi(b)} \min_{a\in[\mathsf{M}]} \frac{(\mathbf{X}[\pi(b), a^*_{\pi(b)}] - \mathbf{X}[\pi(b), a])}{KL(\mathbf{X}[\pi(b), a], \mathbf{X}[c', a])}.$$

So, a tighter lower bound for the regret can be obtained by taking a maximum of this regret lower bound and the lower bound in Equation (18).

G Missing Details in Section 5

G.1 Preliminaries

Lemma 15. *(Conditional Number bounds) Let $\mathbf{P} \in \mathbb{R}^{\mathsf{N}\times\mathsf{M}}$ be a matrix with non-zero singular values $\sigma_1 > \sigma_2 \ldots \ldots > \sigma_r$ for some $\mathsf{M} \geq r > 0$. Consider a sub-matrix $\mathbf{P}_\mathcal{S}$ which is formed by taking all rows of matrix $\mathbf{P}$ and columns from a set $\mathcal{S} \subset [\mathsf{M}]$ of indices. Suppose that the row-space of $\mathbf{P}_\mathcal{S}$ is a non-trivial vector space. Then, $\frac{\lambda_{\max}(\mathbf{P}_\mathcal{S}^T\mathbf{P}_\mathcal{S})}{\lambda_{\min}(\mathbf{P}_\mathcal{S}^T\mathbf{P}_\mathcal{S})} \leq \frac{\lambda_{\max}(\mathbf{P}^T\mathbf{P})}{\lambda_{\min}(\mathbf{P}^T\mathbf{P})}$ where $\lambda_{\max}(\cdot), \lambda_{\min}(\cdot)$ corresponds to the largest and smallest non-zero eigenvalues of the corresponding matrix.*

Proof. Let $\mathbf{P}_i$ be the i-th row of $\mathbf{P}$.

$$\lambda_{\max}(\mathbf{P}^T\mathbf{P}) = \sup_{\mathbf{x}:\|\mathbf{x}\|_2=1} \mathbf{x}^T\mathbf{P}^T\mathbf{P}\mathbf{x} = \sup_{\mathbf{x}:\|\mathbf{x}\|_2=1} \sum_i \mathbf{x}^T\mathbf{P}_i^T\mathbf{P}_i\mathbf{x} \geq \sup_{\mathbf{x}:\|\mathbf{x}\|_2=1,\ \mathbf{x}_{[\mathsf{M}]-\mathcal{S}}=\mathbf{0}} \sum_i \mathbf{x}^T\mathbf{P}_i^T\mathbf{P}_i\mathbf{x}$$
$$= \sup_{\mathbf{x}:\|\mathbf{x}\|_2=1} \sum_i \mathbf{x}^T\mathbf{P}_{i|\mathcal{S}}^T\mathbf{P}_{i|\mathcal{S}}\mathbf{x} = \lambda_{\max}(\mathbf{P}_\mathcal{S}^T\mathbf{P}_\mathcal{S}). \tag{19}$$

Let $\mathcal{K}$ be the row-space of $\mathbf{P}$. Let $\mathcal{K}'$ be the row space of $\mathbf{P}_\mathcal{S}$.

$$\lambda_{\min}(\mathbf{P}^T\mathbf{P}) = \inf_{\mathbf{x}\in\mathcal{K}:\|\mathbf{x}\|_2=1} \mathbf{x}^T\mathbf{P}^T\mathbf{P}\mathbf{x} = \inf_{\mathbf{x}\in\mathcal{K}:\|\mathbf{x}\|_2=1} \sum_i \mathbf{x}^T\mathbf{P}_i^T\mathbf{P}_i\mathbf{x} \overset{(a)}{\leq} \inf_{\mathbf{x}\in\mathcal{K}:\|\mathbf{x}\|_2=1,\ \mathbf{x}_{[\mathsf{M}]-\mathcal{S}}=\mathbf{0}} \sum_i \mathbf{x}^T\mathbf{P}_i^T\mathbf{P}_i\mathbf{x}$$
$$= \inf_{\mathbf{x}\in\mathcal{K}':\|\mathbf{x}\|_2=1} \sum_i \mathbf{x}^T\mathbf{P}_{i|\mathcal{S}}^T\mathbf{P}_{i|\mathcal{S}}\mathbf{x} = \lambda_{\min}(\mathbf{P}_\mathcal{S}^T\mathbf{P}_\mathcal{S}). \tag{20}$$

(a) is due to the fact that there is at least one $\mathbf{x} \neq 0$ in the row-space of $\mathbf{P}$ with only non-zero entries in $\mathcal{S}$. $\qquad\square$

Lemma (Restatement of Lemma 4). *Suppose Assumption 3 is true. Consider a sub-matrix $\mathbf{P}_{\mathsf{sub}}$ of $\mathbf{P}$ having non-zero singular values $\lambda'_1 > \cdots > \lambda'_{\mathsf{C}'}$ (for $\mathsf{C}' \leq \mathsf{C}$). Then, provided $\mathbf{P}_{\mathsf{sub}}$ is non-zero, we have $\frac{\lambda'_1}{\lambda'_{\mathsf{C}'}} \leq \frac{\lambda_1}{\lambda_\mathsf{C}}$.*

Proof. Let $\mathcal{S}' \subset [\mathsf{N}]$, $\mathcal{S} \subset [\mathsf{M}]$. Let $\mathbf{P}_{\mathcal{S}',\mathcal{S}}$ be the sub-matrix formed by choosing column indices $\mathcal{S}'$ and $\mathcal{S}$ from matrix $\mathbf{P} \in \mathbb{R}^{\mathsf{N}\times\mathsf{M}}$. Let $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ be the largest and smallest non-zero singular values. Suppose, that $\dim(\text{rowspace}(\mathbf{P}_{[\mathsf{N}],\mathcal{S}})) \geq 1$. Let $\dim(\text{colspace}(\mathbf{P}_{\mathcal{S}',\mathcal{S}})) \geq 1$. Applying Lemma 15 to $\mathbf{P}_{:,\mathcal{S}}$ and $\mathbf{P}$ using the fact that rowspace of $\mathbf{P}_{:,\mathcal{S}}$ is a non-trivial vector space, we get: $\frac{\lambda_{\max}(\mathbf{P}_{:,\mathcal{S}}^T\mathbf{P}_{:,\mathcal{S}})}{\lambda_{\min}(\mathbf{P}_{:,\mathcal{S}}^T\mathbf{P}_{:,\mathcal{S}})} \leq \frac{\lambda_{\max}(\mathbf{P}^T\mathbf{P})}{\lambda_{\min}(\mathbf{P}^T\mathbf{P})}$.

Again applying Lemma 15 to $\mathbf{P}_{\mathcal{S}',\mathcal{S}}^T$ and $\mathbf{P}_{:,\mathcal{S}}^T$ and using the fact that rowspace of $\mathbf{P}_{\mathcal{S}',\mathcal{S}}^T$ is a non-trivial vector space, we get: $\frac{\lambda_{\max}(\mathbf{P}_{\mathcal{S}',\mathcal{S}}\mathbf{P}_{\mathcal{S}',\mathcal{S}}^T)}{\lambda_{\min}(\mathbf{P}_{\mathcal{S}',\mathcal{S}}\mathbf{P}_{\mathcal{S}',\mathcal{S}}^T)} \leq \frac{\lambda_{\max}(\mathbf{P}_{:,\mathcal{S}}\mathbf{P}_{:,\mathcal{S}}^T)}{\lambda_{\min}(\mathbf{P}_{:,\mathcal{S}}\mathbf{P}_{:,\mathcal{S}}^T)}$. We observe that $\frac{\lambda_{\max}(\mathbf{P}_{:,\mathcal{S}}\mathbf{P}_{:,\mathcal{S}}^T)}{\lambda_{\min}(\mathbf{P}_{:,\mathcal{S}}\mathbf{P}_{:,\mathcal{S}}^T)} = \frac{\lambda_{\max}(\mathbf{P}_{:,\mathcal{S}}^T\mathbf{P}_{:,\mathcal{S}})}{\lambda_{\min}(\mathbf{P}_{:,\mathcal{S}}^T\mathbf{P}_{:,\mathcal{S}})}$. This implies: $\frac{\lambda_{\max}(\mathbf{P}_{\mathcal{S}',\mathcal{S}}\mathbf{P}_{\mathcal{S}',\mathcal{S}}^T)}{\lambda_{\min}(\mathbf{P}_{\mathcal{S}',\mathcal{S}}\mathbf{P}_{\mathcal{S}',\mathcal{S}}^T)} \leq \frac{\lambda_{\max}(\mathbf{P}^T\mathbf{P})}{\lambda_{\min}(\mathbf{P}^T\mathbf{P})}$. Taking square root on both sides yields the result. $\qquad\square$

Lemma (Restatement of Lemma 5). *Suppose Assumption 3 is true. Consider a sub-matrix $\mathbf{P}_{\mathsf{sub}} \in \mathbb{R}^{\mathsf{B}'\times\mathsf{A}'}$ (with SVD decomposition $\mathbf{P}_{\mathsf{sub}} = \widetilde{\mathbf{U}}\widetilde{\mathbf{\Sigma}}\widetilde{\mathbf{V}}$) of $\mathbf{P}$ whose rows correspond to a nice subset of users. Then, $\left\|\widetilde{\mathbf{U}}\right\|_{2,\infty} \leq \sqrt{\frac{\mu\mathsf{C}}{\beta\mathsf{N}'}}$ and $\left\|\widetilde{\mathbf{V}}\right\|_{2,\infty} \leq \sqrt{\frac{\mu\mathsf{C}}{\alpha\mathsf{M}'}}$.*

Proof of Lemma 5. Suppose the reward matrix $\mathbf{P}$ has the SVD decomposition $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. We are looking at a sub-matrix of $\mathbf{P}$ denoted by $\mathbf{P}_{\mathsf{sub}} \in \mathbb{R}^{\mathsf{N}'\times\mathsf{M}'}$ which can be represented as $\mathbf{U}_{\mathsf{sub}}\mathbf{\Sigma}\mathbf{V}_{\mathsf{sub}}^T$ where $\mathbf{U}_{\mathsf{sub}}, \mathbf{V}_{\mathsf{sub}}$ are sub-matrices of $\mathbf{U}, \mathbf{V}$ respectively and are not necessarily orthogonal. Here, the rows in $\mathbf{P}_{\mathsf{sub}}$ corresponds to a union of cluster of users $\cup_{j\in\mathcal{A}}\mathsf{C}^{(j)}$

for $\mathcal{A} \in [\mathsf{C}]$ and the columns in $\mathbf{P}_{\mathsf{sub}}$ corresponds to the trimmed set of arms in $[\mathsf{M}]$. Hence we can write (provided $\mathbf{U}_{\mathsf{sub}}, \mathbf{V}_{\mathsf{sub}}$ are invertible)

$$\mathbf{P}_{\mathsf{sub}} = \mathbf{U}_{\mathsf{sub}}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})^{-1/2}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})^{1/2}\boldsymbol{\Sigma}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{1/2}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$$
$$= \mathbf{U}_{\mathsf{sub}}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})^{-1/2}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}^\mathsf{T}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$$

where $\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}^\mathsf{T}$ is the SVD of the matrix $(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})^{1/2}\boldsymbol{\Sigma}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{1/2}$. Since $\widehat{\mathbf{U}}$ is orthogonal, $\mathbf{U}_{\mathsf{sub}}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})^{-1/2}\widehat{\mathbf{U}}$ is orthogonal. Similarly, $\widehat{\mathbf{V}}^\mathsf{T}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$ is orthogonal as well whereas $\widehat{\boldsymbol{\Sigma}}$ is diagonal. Hence $\mathbf{U}_{\mathsf{sub}}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})^{-1/2}\widehat{\mathbf{U}}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{V}}^\mathsf{T}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$ indeed corresponds to the SVD of $\mathbf{P}_{\mathsf{sub}}$ and we only need to argue about the incoherence of the matrices $\mathbf{U}_{\mathsf{sub}}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})^{-1/2}\widehat{\mathbf{U}}$ and $(\widehat{\mathbf{V}}^\mathsf{T}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T})^\mathsf{T}$. We have

$$\max_i \|\widehat{\mathbf{U}}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})^{-1/2}\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{e}_i\| \leq \max_i \|(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})^{-1/2}\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{e}_i\|$$

$$\leq \frac{\|\mathbf{U}_{\mathsf{sub}}\|_{2,\infty}}{\sqrt{\lambda_{\min}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})}} \leq \frac{\|\mathbf{U}\|_{2,\infty}}{\sqrt{\lambda_{\min}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})}} \leq \sqrt{\frac{\mu\mathsf{C}}{\mathsf{N}}} \cdot \frac{1}{\sqrt{\lambda_{\min}(\mathbf{U}_{\mathsf{sub}}^\mathsf{T}\mathbf{U}_{\mathsf{sub}})}} \leq \sqrt{\frac{\mu\mathsf{C}^2}{\mathsf{N}\beta\tau|\mathcal{A}|}} \leq \sqrt{\frac{\mu\mathsf{C}}{\beta\mathsf{N}'}}$$

where we used the fact that $\mathsf{NC}^{-1}|\mathcal{A}| \geq \mathsf{N}'\tau^{-1}$ (NC^{-1} is the average cluster size; $\mathsf{N}'/|\mathcal{A}|$ is the average cluster size among the N' users and τ^{-1} is the ratio of the sizes of smallest cluster and largest cluster). Similarly, we will also have

$$\max_i \|\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{e}_i\| \leq \max_i \|(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{e}_i\|$$

$$\leq \frac{\|\mathbf{V}_{\mathsf{sub}}\|_{2,\infty}}{\sqrt{\lambda_{\min}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})}} \leq \frac{\|\mathbf{V}\|_{2,\infty}}{\sqrt{\lambda_{\min}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})}} \leq \sqrt{\frac{\mu\mathsf{C}}{\alpha\mathsf{M}'}}$$

where the last line follows from the fact that $\min_{\mathbf{x}\in\mathbb{R}^\mathsf{C}} \mathbf{x}^\mathsf{T}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}}\mathbf{x} = \min_{\mathbf{x}\in\mathbb{R}^\mathsf{C}} \sum_{i\in\mathcal{S}} \mathbf{x}^\mathsf{T}\mathbf{V}_i^\mathsf{T}\mathbf{V}_i\mathbf{x}$ where $\mathcal{S}, |\mathcal{S}| = \mathsf{M}'$ is the set of rows in $\mathbf{V}$ represented in $\mathbf{V}_{\mathsf{sub}}$. Hence we can partition the set $\mathcal{S}$ into $|\mathcal{S}|/\mathsf{C}$ disjoint sets, each of size C; from Assumption 3, we get that the minimum over each of those sets is $\alpha\mathsf{C}/\mathsf{M}$ and therefore, the conclusion follows.

Lemma 16. *Let us fix $\Delta_{\ell+1} > 0$ and condition on the event $\mathcal{E}^{(\ell)}$. Suppose Assumptions 3 and 2 are satisfied. In that case, in phase ℓ, by using*

$$m_\ell = O\Big(\frac{\sigma^2\mathsf{C}^2(\mu\beta^{-1}\bigvee\mu\alpha^{-1})^3\log\mathsf{M}}{\Delta_{\ell+1}^2}\Big(\mathsf{N}\bigvee\mathsf{MC}\Big)\log^2(\mathsf{MNC}\delta^{-1})\Big)\Big)$$

rounds, we can compute an estimate $\widetilde{\mathbf{P}}^{(\ell)} \in \mathbb{R}^{\mathsf{N}\times\mathsf{M}}$ such that with probability $1 - \delta$, we have

$$\left\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}^{(\ell)} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\right\|_\infty \leq \Delta_{\ell+1} \text{ for all } i \in [a_\ell] \text{ satisfying } \left|\mathcal{N}^{(i,\ell)}\right| \geq \gamma\mathsf{C}. \tag{21}$$

Proof of Lemma 16. We are going to use Lemma 1 in order to compute an estimate $\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}^{(\ell)}$ of the sub-matrix $\mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$ satisfying $\left\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}^{(\ell)} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\right\|_\infty \leq \Delta_{\ell+1}$. From Lemma 1, we know that by using $m_\ell = O\Big(s\mathsf{N}\log^2(\mathsf{MN}\delta^{-1})(|\mathcal{N}^{(\ell,i)}|p + \sqrt{|\mathcal{N}^{(\ell,i)}|p\log\mathsf{N}\delta^{-1}})\Big)$ rounds (see Lemma 1) restricted to users in $\mathcal{M}^{(\ell,i)}$ such that with probability at least $1 - \delta$,

$$\left\|\widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}^{(\ell)} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\right\|_\infty \leq O\Big(\frac{\sigma r}{\sqrt{sd_2}}\sqrt{\frac{\widetilde{\mu}^3\log d_2}{p}}\Big).$$

where $d_2 = \min(|\mathcal{M}^{(\ell,i)}|, |\mathcal{N}^{(\ell,i)}|)$ and $\widetilde{\mu}$ is the incoherence factor of the matrix $\mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$. In order for the right hand side to be less than $\Delta_{\ell+1}$, we can set $sp = O\Big(\frac{\sigma^2 r^2\widetilde{\mu}^3\log d_2}{\Delta_{\ell+1}^2 d_2}\Big)$. Since the event $\mathcal{E}^{(\ell)}$ is true, we must have that $|\mathcal{M}^{(\ell,i)}| \geq \mathsf{N}/\mathsf{C}$; hence $d_2 \geq \min\Big(\frac{\mathsf{N}}{\mathsf{C}}, |\mathcal{N}^{(\ell,i)}|\Big)$. Therefore, we must have that

$$m_\ell = O\Big(\frac{\sigma^2\mathsf{C}^2\widetilde{\mu}^3\log\mathsf{M}}{\Delta_{\ell+1}^2}\max\Big(\mathsf{N}, \mathsf{MC}\Big)\log^2(\mathsf{MNC}\delta^{-1})\Big)$$

where we take a union bound over all sets comprising the partition of the users $[N]$ (at most C of them). Finally, from Lemma 5, we know that $\widetilde{\mu}$ can be bounded from above by $\max(\mu/\beta, \mu/\alpha)$ which we can use to say that

$$m_\ell = O\Big(\frac{\sigma^2 C^2 (\mu\beta^{-1} \vee \mu\alpha^{-1})^3 \log M}{\Delta_{\ell+1}^2}\Big(N \bigvee MC\Big) \log^2(MNC\delta^{-1})\Big)$$

to complete the proof of the lemma. $\qquad\square$

$\qquad\square$

G.2   Joint Arm Elimination:

The first part of the algorithm is run in phases indexed by $\ell = 1, 2, \ldots$ in a similar way as for the CS problem. As before, in the beginning of each phase ℓ, our goal is to maintain properties $(1 - 4)$ proposed at the beginning of Section E. Similarly, for a fixed ϵ_ℓ that will be determined later, we will say that our phased elimination algorithm is $\epsilon_\ell-good$ at the beginning of the ℓth phase if the randomized algorithm LATTICE (for RCS) can maintain a list of users and arms satisfying the properties $(1 - 4)$ at the start of phase ℓ. Let us also define the event $\mathcal{E}^{(\ell)}$ to be true if properties $(1 - 4)$ are satisfied at the beginning of phase ℓ by LATTICE for RCS.

Let us fix $\Delta_{\ell+1} > 0$ and condition on the event $\mathcal{E}^{(\ell)}$ being true at the beginning of phase ℓ. Suppose Assumptions 3 and 2 are satisfied. In that case, in phase ℓ, recall from Lemma 16 that by using

$$m_\ell = O\Big(\frac{\sigma^2 C^2 (\mu\beta^{-1} \vee \mu\alpha^{-1})^3 \log M}{\Delta_{\ell+1}^2}\Big(N \bigvee MC\Big) \log^2(ABCT)\Big)$$

rounds, we can compute an estimate $\widetilde{\mathbf{P}}^{(\ell)} \in \mathbb{R}^{N \times M}$ such that with probability $1 - T^{-4}$, we have

$$\left\|\widetilde{\mathbf{P}}^{(\ell)}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}} - \mathbf{P}_{\mathcal{M}^{(\ell,i)}, \mathcal{N}^{(\ell,i)}}\right\|_\infty \leq \Delta_{\ell+1} \text{ for all } i \in [a_\ell] \text{ satisfying } \left|\mathcal{N}^{(i,\ell)}\right| \geq \gamma C.$$

Furthermore, we denoted the aforementioned event in phase ℓ by $\mathcal{E}^{(\ell)}_2$. Recall that in the ℓth phase, we have $\mathcal{T}^{(\ell)}_u \equiv \{j \in \mathcal{N}^{(\ell,i)} \mid \max_{j' \in \mathcal{N}^{(\ell,i)}} \widetilde{\mathbf{P}}^{(\ell)}_{uj'} - \widetilde{\mathbf{P}}^{(\ell)}_{uj} \leq 2\Delta_{\ell+1}\}$. For every $i \in [a_\ell]$ such that $\left|\mathcal{N}^{(\ell,i)}\right| \geq \gamma C$, we consider a graph $\mathcal{G}^{(\ell,i)}$ whose nodes are given by the users in $\mathcal{M}^{(\ell,i)}$, an edge is drawn between two users $u, v \in \mathcal{M}^{(\ell,i)}$ if $\mathcal{T}^{(\ell)}_u \cap \mathcal{T}^{(\ell)}_v \neq \Phi$ and $\max_{x \in \mathcal{N}^{(\ell,i)}} \left|\widetilde{\mathbf{P}}^{(\ell)}_{ux} - \widetilde{\mathbf{P}}^{(\ell)}_{vx}\right| \leq 3\Delta_{\ell+1}$. Note that Lemma 8 holds true implying that conditioned on the event $\mathcal{E}^{(\ell)}_2$, $\pi_u(1) \in \mathcal{T}^{(\ell)}_u$ for all users $u \in [N]$. Next, we show the following:

Lemma 17. *Let $\Delta_{\ell+1} \geq \nu$. Fix any $i \in [a_\ell]$. Conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}^{(\ell)}_2$, nodes in $\mathcal{M}^{(\ell,i)}$ corresponding to the same cluster form a clique.*

Proof. For any two users $u, v \in \mathcal{M}^{(\ell,i)}$ belonging to the same cluster, consider arms $x \in \mathcal{T}^{(\ell)}_u, y \in \mathcal{T}^{(\ell)}_v$; in that case, we have

$$\widetilde{\mathbf{P}}^{(\ell)}_{ux} - \widetilde{\mathbf{P}}^{(\ell)}_{vx} = \widetilde{\mathbf{P}}^{(\ell)}_{ux} - \mathbf{P}_{ux} + \mathbf{P}_{ux} - \mathbf{P}_{vx} + \mathbf{P}_{vx} - \widetilde{\mathbf{P}}^{(\ell)}_{vx} \leq 2\Delta_{\ell+1} + \nu \leq 3\Delta_{\ell+1}.$$

it is clear that the nodes in $\mathcal{M}^{(\ell,i)}$ corresponding to the same cluster form a clique. $\qquad\square$

Corollary 2. *Condition on the events $\mathcal{E}^{(\ell)}, \mathcal{E}^{(\ell)}_2$ and suppose $\Delta_{\ell+1} \geq \nu$. 1) Each connected component of the graph $\mathcal{G}^{(\ell,i)}$ can be represented as a nice set of users. 2) Consider two users $u, v \in \mathcal{M}^{(\ell,i)}$ having an edge in the graph $\mathcal{G}^{(\ell,i)}$. We must have $\max_{x \in \mathcal{T}^{(\ell)}_u, y \in \mathcal{T}^{(\ell)}_v} |\mathbf{P}_{ux} - \mathbf{P}_{vy}| \leq 20\Delta_{\ell+1}$. 3) Consider two users $u, v \in \mathcal{M}^{(\ell,i)}$ having a path in the graph $\mathcal{G}^{(\ell,i)}$. Then, we must have $\max_{x \in \mathcal{T}^{(\ell)}_u, y \in \mathcal{T}^{(\ell)}_v} |\mathbf{P}_{ux} - \mathbf{P}_{vy}| \leq 40C\Delta_{\ell+1}$.*

Proof.   1. This follows from the statement that users in the same set belonging to the same cluster form a clique (see Lemma 17).

2. From the construction of $\mathcal{G}^{(\ell,i)}$, we know that users $u, v \in \mathcal{M}^{(\ell,i)}$ have an edge if $\left|\widetilde{\mathbf{P}}^{(\ell)}_{ux} - \widetilde{\mathbf{P}}^{(\ell)}_{vx}\right| \leq 3\Delta_{\ell+1}$ (implying that $|\mathbf{P}_{ux} - \mathbf{P}_{vx}| \leq 6\Delta_{\ell+1}$) for all $x \in \mathcal{N}^{(\ell,i)}$ and $\mathcal{T}_u \cap \mathcal{T}_v \neq \Phi$. Suppose $z \in \mathcal{T}_u \cap \mathcal{T}_v$. Now, for any pair of arms $x \in \mathcal{T}^{(\ell)}_u, y \in \mathcal{T}^{(\ell)}_v$; in that case, we have

$$\mathbf{P}_{ux} - \mathbf{P}_{uy} = \mathbf{P}_{ux} - \mathbf{P}_{uz} + \mathbf{P}_{uz} - \mathbf{P}_{vz} + \mathbf{P}_{vz} - \mathbf{P}_{vy} + \mathbf{P}_{vy} - \mathbf{P}_{uy} \leq 20\Delta_{\ell+1}$$

3. The proof follows in exactly the same way as in Lemma 11 with minor modification in constants.

$\square$

Lemma 18. *Let $\Delta_{\ell+1} \leq 2\nu$. Conditioned on the events $\mathcal{E}^{(\ell)}$ and $\mathcal{E}_2^{(\ell)}$, for any two users u, v belonging to two different clusters, we will have*

$$\left| \widetilde{\mathbf{P}}_{ux}^{(\ell)} - \widetilde{\mathbf{P}}_{vy}^{(\ell)} \right| \geq 8\Delta_{\ell+1} \text{ or } \left| \widetilde{\mathbf{P}}_{ux}^{(\ell)} - \widetilde{\mathbf{P}}_{vy}^{(\ell)} \right| \geq 8\Delta_{\ell+1}.$$

Proof. For any two users $u, v \in \mathcal{M}^{(\ell,i)}$ belonging to the same cluster, consider arms $\pi_u(1) \in \mathcal{T}_u^{(\ell)}, \pi_v(1) \in \mathcal{T}_v^{(\ell)}$ (we have proved in Lemma 8 that $\pi_u(1) \in \mathcal{T}_u^{(\ell)}, \pi_v(1) \in \mathcal{T}_v^{(\ell)}$ conditioned on the event $\mathcal{E}_2^{(\ell)}$); in that case, recall that we have either $\left| \mathbf{P}_{u\pi_u(1)} - \mathbf{P}_{v\pi_u(1)} \right| \geq 20\nu$ or $\left| \mathbf{P}_{u\pi_v(1)} - \mathbf{P}_{v\pi_v(1)} \right| \geq 20\nu$ (without loss of generality suppose the former is true). We will have

$$\widetilde{\mathbf{P}}_{u\pi_u(1)}^{(\ell)} - \widetilde{\mathbf{P}}_{v\pi_u(1)}^{(\ell)} = \widetilde{\mathbf{P}}_{u\pi_u(1)}^{(\ell)} - \mathbf{P}_{u\pi_u(1)} + \mathbf{P}_{u\pi_u(1)} - \mathbf{P}_{v\pi_u(1)} + \mathbf{P}_{v\pi_u(1)} - \widetilde{\mathbf{P}}_{v\pi_u(1)}^{(\ell)} \geq 20\nu - 2\Delta_{\ell+1} \geq 16\nu \geq 8\Delta_{\ell+1}.$$

it is clear that the nodes in $\mathcal{M}^{(\ell,i)}$ corresponding to the same cluster form a clique.

$\square$

Lemma 19. *Let $\nu \leq \Delta_{\ell+1} \leq 2\nu$. Conditioned on the events $\mathcal{E}^{(\ell)}$ and $\mathcal{E}_2^{(\ell)}$, the union of the graphs $\mathcal{G}^{(\ell,i)}$ can be represented as C connected components if $\mathcal{B}^{(\ell)} = \Phi$.*

Proof. The proof follows from Lemma 17 and Lemma 18.

$\square$

For the subsequent iteration indexed by $\ell + 1$, we compute the updated groups of users $\mathcal{M}^{(\ell+1)}$ in the following way: each set corresponds to the connected components of the graphs $\{\mathcal{G}^{(\ell,i)}\}$ for those indices $i \in [a_\ell]$ where $\left| \mathcal{N}^{(\ell,i)} \right| \geq \gamma\mathsf{C}$ plus the groups of users $\mathcal{M}^{(\ell,i)}$ where $\left| \mathcal{N}^{(\ell,i)} \right| \leq \mathsf{C}$. More precisely, let $\mathcal{T} \subseteq [a_\ell]$ be the subset of indices for which $\left| \mathcal{N}^{(\ell,i)} \right| \geq \gamma\mathsf{C}$; $\{\mathcal{G}^{(\ell,i,j)}\}$ be the connected components of the graph $\mathcal{G}^{(\ell,i)}$ for $i \in \mathcal{T}$. In that case, $\mathcal{M}^{(\ell+1)} = \{\mathcal{G}^{(\ell,i,j)} \mid i \in \mathcal{T}, \mathcal{G}^{(\ell,i,j)}$ is a connected component of graph $\mathcal{G}^{(\ell,i)}\} + \{\mathcal{M}^{(\ell,i)} \mid i \in [a_\ell] \setminus \mathcal{T}\}$. Similarly, we update the family of sets of active arms as follows: for each connected component $\mathcal{M}^{(\ell+1,s)} = \mathcal{G}^{(\ell,i,j)}$ of some graph, we define the active set of arms $\mathcal{N}^{(\ell+1,s)}$ to be $\cup_{u \in \mathcal{G}^{(\ell,i,j)}} \mathcal{T}_u^{(\ell)}$ and for each group $\{\mathcal{M}^{(\ell,i)}\}_{i \in [a_\ell] \setminus \mathcal{T}}$, we keep the corresponding set of active arms $\{\mathcal{N}^{(\ell,i)}\}_{i \in [a_\ell] \setminus \mathcal{T}}$ same. With $a_{\ell+1} = \left| \mathcal{M}^{(\ell+1)} \right|$, we will also update $\mathcal{B}^{(\ell+1)} = \bigcup_{i \in [a_{\ell+1}] \mid |\mathcal{N}^{(\ell+1,i)}| \geq \gamma\mathsf{C}} \mathcal{M}^{(\ell+1,i)}$ to be the set of users with more than C active arms.

Lemma 20. *Condition on the events $\mathcal{E}^{(\ell)}$ being true. In that case, with probability $1 - \mathsf{T}^{-4}$, with the groups of users $\mathcal{M}^{(\ell+1)}$ and their respective group of arms given by $\mathcal{N}^{(\ell+1)}$ being updated as described above and $\Delta_{\ell+1} = \epsilon_\ell/80\mathsf{C}$, the event $\mathcal{E}^{(\ell+1)}$ is also going to be true with $\epsilon_{\ell+1} \leq \epsilon_\ell/2$.*

Proof. Conditioned on the event $\mathcal{E}^{(\ell)}$ being true, the event $\mathcal{E}_2^{(\ell)}$ holds true with probability with $1 - \mathsf{T}^{-4}$ (by substituting $\delta = \mathsf{T}^{-4}$ in Lemma 12). Now, conditioned on the event $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ being true, the properties $(1-4)$ hold true at the beginning of the $(\ell + 1)^{\text{th}}$ phase as well with our construction of $\mathcal{M}^{(\ell+1)}, \mathcal{N}^{(\ell+1)}$. For the $(\ell + 1)^{\text{th}}$ phase from Lemma 11, we know that for any pair of users u, v in the same cluster $\mathcal{M}^{(\ell+1,i)}$ in the updated set of clusters $\mathcal{M}^{(\ell+1)}$, we must have $\max_{x \in \mathcal{T}_u^{(\ell)}, y \in \mathcal{T}_v^{(\ell)}} |\mathbf{P}_{ux} - \mathbf{P}_{vy}| \leq 40\mathsf{C}\Delta_{\ell+1}$. From Lemma 8 we know that $\text{argmax}_j \mathbf{P}_{uj} \in \mathcal{T}_u^{(\ell)} \subseteq \mathcal{N}^{(\ell,i)}$ where $\mathcal{N}^{(\ell,i)}$ is the active set of arms for users in $\mathcal{M}^{(\ell,i)}$ in the updated set $\mathcal{M}$. For $\ell > 1$, we will set $\Delta_{\ell+1} = \epsilon_\ell/80\mathsf{C}$ which would give us that $\epsilon_{\ell+1} = \epsilon_\ell/2$. Finally, also note that we maintain the set of users $\mathcal{B}^{(\ell+1)}$ as stipulated in Property 4 for the beginning of the $(\ell + 1)^{\text{th}}$ phase.

$\square$

We stop the joint phased elimination algorithm at the end of the phase $\ell^\star$ when we have C connected components. From Lemma 19, we know that users in the same connected component must belong to the same cluster. We then move on to the second part where we run cluster-wise phased elimination:

G.3 Cluster-wise Arm Elimination:

Suppose the total number of phases that the joint phased elimination algorithm was ran for is $\ell^\star$ ($\nu \leq \Delta_{\ell^\star+1} \leq 2\nu$). We condition on the event that the joint phased elimination algorithm is ϵ_ℓ-good for all $\ell \in [\ell^\star]$ that it ran for (in other words the event $\mathcal{E}^{(\ell^\star)}$ is true); note that the failure probability for this event is at most T^{-3} (see Lemma 20). At the end of the joint phased elimination algorithm, we will have C connected components denoted by $\mathsf{N}^{(\ell^\star)} \equiv \{\mathcal{M}^{(\ell^\star,1)}, \mathcal{M}^{(\ell^\star,2)}, \dots, \mathcal{M}^{(\ell^\star,\mathsf{C})}\}$ and arms $\mathcal{N}^{\ell^\star} \equiv \{\mathcal{N}^{(\ell^\star,1)}, \mathcal{N}^{(\ell^\star,2)}, \dots, \mathcal{N}^{(\ell^\star,\mathsf{C})}\}$ such that $\mathcal{N}^{(\ell^\star,i)} \supseteq \{\operatorname{argmax}_j \mathbf{P}_{uj} \mid u \in \mathcal{M}^{(\ell^\star,i)}\}$ i.e. for each user u in the set $\mathcal{M}^{(\ell^\star,i)}$, their best arm must belong to the set $\mathcal{N}^{(\ell^\star,i)}$. Furthermore, the set $\mathcal{N}^{(\ell^\star,i)}$ must also satisfy the following:

$$\left| \max_{j \in \mathcal{N}^{(\ell^\star,i)}} \mathbf{P}_{uj} - \min_{j \in \mathcal{N}^{(\ell^\star,i)}} \mathbf{P}_{uj} \right| \leq \epsilon_{\ell^\star+1} \text{ for all } u \in \mathcal{M}^{(\ell^\star,i)} \tag{22}$$

This part of the algorithm is again run in phases but this time, we do not cluster anymore since the clustering operation is complete in the joint phased elimination. Hence, we can continue to define the events $\mathcal{E}^{(\ell')}, \mathcal{E}_2^{(\ell')}$ for all phases ℓ' for which the Cluster-wise phased elimination algorithm is run indexed by $\ell' = \ell^\star + 1, \ell^\star + 2, \dots$ for continuity. Let us condition on the event $\mathcal{E}^{(\ell')}$ which means that at the beginning of the ℓ'^{th} phase, our algorithm is $\epsilon_{\ell'}$-good. We apply Lemma 12 at the beginning of the ℓ'^{th} phase such that with probability at least $1 - \mathsf{T}^{-4}$, we have $\left\| \widetilde{\mathbf{P}}_{\mathcal{M}^{(\ell',i)}, \mathcal{N}^{(\ell',i)}}^{(\ell')} - \mathbf{P}_{\mathcal{M}^{(\ell',i)}, \mathcal{N}^{(\ell',i)}} \right\|_\infty \leq \Delta_{\ell'+1}$ for all $i \in [\mathsf{C}]$ by using ($\mu^\star = \max(\mu\alpha^{-1}, \mu\beta^{-1})$)

$$O\left(\frac{\sigma^2 \mathsf{C}^2 (\mu^\star)^3 \log \mathsf{M}}{\Delta_{\ell'+1}^2} \left(\mathsf{N} \bigvee \mathsf{MC} \right) \log^2(\mathsf{ABCT}) \right)$$

rounds. Note that Lemma 8 still holds true for the ℓ'^{th} phase. For each $i \in [\mathsf{C}]$, we only update $\mathcal{N}^{(\ell,i)} = \cap_{u \in \mathcal{M}^{(\ell,i)}} \mathcal{T}_u$. We again choose $\Delta_{\ell'+1} = \epsilon_{\ell'}/32\mathsf{C}$ implying that $\epsilon_{\ell'+1} = \epsilon_{\ell'}/2$. Since Lemma 8 holds true, $\operatorname{argmax}_j \mathbf{P}_{uj}$ will belong to the updated $\mathcal{N}^{(\ell,i)}$ for every $u \in \mathcal{M}^{(\ell,i)}$ and furthermore,

$$\left| \max_{j \in \mathcal{N}^{(\ell,i)}} \mathbf{P}_{uj} - \min_{j \in \mathcal{N}^{(\ell,i)}} \mathbf{P}_{uj} \right| \leq 4\Delta_{\ell'+1} \leq \epsilon_{\ell'+1} \text{ for all } u \in \mathcal{M}^{(\ell,i)}$$

.

Now, we are ready to prove the main theorem

Proof of Theorem 5. We condition on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ being true for all ℓ (including the joint phased elimination and the cluster-wise phased elimination algorithm). The probability that there exists any ℓ such that the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ is false is $O(\mathsf{T}^{-4})$ (by setting $\delta = \mathsf{T}^{-4}$ in the proof of Lemma 12); hence the probability that $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ is true for all ℓ is at least $1 - O(\mathsf{T}^{-3})$ (the total number of iterations can be at most T). Let us also denote the set of rounds in phase ℓ by $\mathcal{T}_\ell \subseteq [\mathsf{T}]$ (therefore $|\mathcal{T}_\ell| = m_\ell$). Let us compute the regret $\sum_{t \in \mathcal{T}^{(\ell)}} \mathbf{P}_{u(t)\pi_{u(t)}(1)} - \sum_{t \in \mathcal{T}^{(\ell)}} \mathbf{P}_{u(t),\rho(t)}$ restricted to the rounds in $\mathcal{T}^{(\ell)}$ conditioned on the events $\mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)}$ being true for all ℓ. We can bound the regret quantity in the ℓth phase from above by $m_\ell \epsilon_\ell$. Substituting from Lemma 12 and using the fact that $\Delta_{\ell+1}^2 = \epsilon_\ell^2/1024\mathsf{C}^2$, we have that

$$\sum_{t \in \mathcal{T}^{(\ell)}} \mathbf{P}_{u(t)\pi_{u(t)}(1)} - \sum_{t \in \mathcal{T}^{(\ell)}} \mathbf{P}_{u(t),\rho(t)} = O\left(\frac{\sigma^2 \mathsf{C}^2 (\mu^\star)^3 \log \mathsf{M}}{\Delta_{\ell+1}^2} \left(\mathsf{N} \bigvee \mathsf{MC} \right) \log^2(\mathsf{MNC}\delta^{-1}) \right)$$

We can now bound the regret as follows (after removing the conditioning on the events $\bigcap_\ell \mathcal{E}^{(\ell)} \bigcap_\ell \mathcal{E}_2^{(\ell)}$):

$$\mathbb{E}\left(\sum_{t \in [\mathsf{T}]} \mathbf{P}_{u(t)\pi_{u(t)}(1)} - \sum_{t \in [\mathsf{T}]} \mathbf{P}_{u(t),\rho(t)} \right) = \mathbb{E}\left(\sum_\ell \left(\sum_{t \in \mathcal{T}^{(\ell)}} \mathbf{P}_{u(t)\pi_{u(t)}(1)} - \sum_{t \in \mathcal{T}^{(\ell)}} \mathbf{P}_{u(t),\rho(t)} \right) \right)$$

$$= \sum_\ell O\left(\epsilon_\ell m_\ell \mid \bigcap_\ell \mathcal{E}^{(\ell)} \bigcap_\ell \mathcal{E}_2^{(\ell)} \right) + O(\mathsf{T}^{-3} \|\mathbf{P}\|_\infty) + \sum_\ell \sum_{u \in \mathsf{N} \setminus [\mathcal{B}^{(\ell)}]} \mathsf{Reg}_{\mathsf{UCB}}(u, \mathcal{T}_u^{(\ell)})$$

At this point, the analysis is similar to the proof of Theorem 1. The contribution of the regret from the final term (analysis of the UCB algorithm for each user with at most $\gamma\mathsf{C}$ arms) is a strictly lower order term as demonstrated in the proof of

Theorem 5 and we ignore it below. Moving on, we can decompose the regret as follows:

$$\text{Reg}(\mathsf{T}) \leq O(\mathsf{T}^{-3}\|\mathbf{P}\|_\infty) + O\Big(\sum_{\ell:\epsilon_\ell \leq \Phi} \epsilon_\ell m_\ell \mid \mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)} \text{ is true for all } \ell\Big) + O\Big(\sum_{\ell:\Delta_\ell > \Phi} \epsilon_\ell \mathsf{V}\Delta_{\ell+1}^{-2} \mid \mathcal{E}^{(\ell)}, \mathcal{E}_2^{(\ell)} \text{ is true for all } \ell\Big)$$

$$\leq O(\mathsf{T}^{-3}\|\mathbf{P}\|_\infty) + \mathsf{T}\Phi + O\Big(\sum_{\ell:\epsilon_\ell > \Phi} \mathsf{C}^2\mathsf{V}\epsilon_\ell^{-1}\Big)$$

Since we chose $\epsilon_\ell = C'2^{-\ell}\min\left(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\right)$ for some constant $C' > 0$, the maximum number of phases ℓ for which $\epsilon_\ell > \Phi$ can be bounded from above by $\mathsf{J} = O\Big(\log\Big(\frac{1}{\Phi}\min\Big(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\Big)\Big)\Big)$. Hence, with $\mathsf{V} = \sigma^2\mathsf{C}^2(\mu^\star)^3\log \mathsf{M}\max\Big(\mathsf{N}, \mathsf{MC}\Big)\log^2(\mathsf{ABCT})$, we have

$$\text{Reg}(\mathsf{T}) \leq O(\mathsf{T}^{-3}\|\mathbf{P}\|_\infty) + O(\mathsf{T}\Phi) + O\Big(\mathsf{JVC}^2\Phi^{-1}\Big)$$

$$= O(\mathsf{T}^{-3}\|\mathbf{P}\|_\infty) + O(\mathsf{CJ}\sqrt{\mathsf{TV}})$$

where we substituted $\Phi = \sqrt{\mathsf{VC}^2\mathsf{T}^{-1}}$ and hence $\mathsf{J} = O\Big(\log\Big(\frac{1}{\sqrt{\mathsf{VT}^{-1}}}\min\Big(\|\mathbf{P}\|_\infty, \frac{\sigma\sqrt{\mu}}{\log \mathsf{N}}\Big)\Big)\Big)$ in the final step.

□

H Proof of a general version of Lemma 1

For the sake of completeness, we reproduce the proof of a more general version of Lemma 1 here (Note that Lemma 1 has been explicitly proved in Jain and Pal [2022]). We start with the following corollary:

Lemma 21 (Theorem 2 in Chen et al. [2019]). *Let* $\mathbf{P} = \bar{\mathbf{U}}\Sigma\bar{\mathbf{V}}^\mathsf{T} \in \mathbb{R}^{d\times d}$ *such that* $\bar{\mathbf{U}} \in \mathbb{R}^{d\times r}, \bar{\mathbf{V}} \in \mathbb{R}^{d\times r}$ *and* $\Sigma \triangleq \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r) \in \mathbb{R}^{r\times r}$ *with* $\bar{\mathbf{U}}^\mathsf{T}\bar{\mathbf{U}} = \bar{\mathbf{V}}^\mathsf{T}\bar{\mathbf{V}} = \mathbf{I}$ *and* $\|\bar{\mathbf{U}}\|_{2,\infty} \leq \sqrt{\mu r/d}, \|\bar{\mathbf{V}}\|_{2,\infty} \leq \sqrt{\mu r/d}$. *Let* $1 \geq p \geq C\kappa^4\mu^2 d^{-1}\log^3 d$ *for some sufficiently large constant* $C > 0$, $\sigma = O\Big(\sqrt{\frac{p}{d\kappa^4\mu r\log d}}\min_i \lambda_i\Big)$, *rank* r *and condition number* $\kappa \triangleq \frac{\max_i \lambda_i}{\min_i \lambda_i}$. *Then, with probability exceeding* $1 - O(d^{-3})$, *we can recover a matrix* $\widehat{\mathbf{P}}$ *s.t.,*

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty \leq O\Big(\frac{\sigma}{\min_i \lambda_i} \cdot \sqrt{\frac{\kappa^3\mu rd\log d}{p}}\|\mathbf{P}\|_\infty\Big). \tag{23}$$

Lemma 22. *Let* $\mathbf{P} = \bar{\mathbf{U}}\Sigma\bar{\mathbf{V}}^\mathsf{T} \in \mathbb{R}^{\mathsf{N}\times \mathsf{M}}$ *such that* $\bar{\mathbf{U}} \in \mathbb{R}^{\mathsf{N}\times r}, \bar{\mathbf{V}} \in \mathbb{R}^{\mathsf{M}\times r}$ *and* $\Sigma \triangleq \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r) \in \mathbb{R}^{r\times r}$ *with* $\bar{\mathbf{U}}^\mathsf{T}\bar{\mathbf{V}} = \bar{\mathbf{V}}^\mathsf{T}\bar{\mathbf{V}} = \mathbf{I}$ *and* $\|\bar{\mathbf{U}}\|_{2,\infty} \leq \sqrt{\mu r/\mathsf{N}}, \|\bar{\mathbf{V}}\|_{2,\infty} \leq \sqrt{\mu r/\mathsf{M}}$. *Let* $d_1 = \max(\mathsf{N}, \mathsf{M})$ *and* $d_2 = \min(\mathsf{N}, \mathsf{M})$. *Let* $1 \geq p \geq C\kappa^4\mu^2 d_1 d_2^{-2}\log^3 d_1$ *for some sufficiently large constant* $C > 0$, $\sigma = O\Big(\sqrt{\frac{p}{d\kappa^4\mu r\log d}}\min_i \lambda_i\Big)$, *rank* $r = O(1)$ *and condition number* $\kappa \triangleq \frac{\max_i \sigma_i}{\min_i \sigma_i} = O(1)$. *Then, with probability exceeding* $1 - O(d_1^{-3})$, *we can recover a matrix* $\widehat{\mathbf{P}}$ *s.t.,*

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty = O\Big(\frac{\sigma}{\sqrt{d_2}}\Big(\frac{d_1}{d_2}\Big)^{1/2}\sqrt{\frac{\kappa^5\mu^3 r^3\log d_1}{p}}\Big). \tag{24}$$

Proof of Lemma 22. Without loss of generality, let us assume that the matrix $\mathbf{P}$ is tall i.e. $\mathsf{N} \geq \mathsf{M}$. Now, let us construct the matrix

$$\mathbf{Q} = \begin{bmatrix} \mathbf{P} & \mathbf{0}_{\mathsf{N}\times \mathsf{B}-\mathsf{A}} \end{bmatrix} = \bar{\mathbf{U}}\Sigma[\bar{\mathbf{V}}^\mathsf{T} \ \mathbf{0}_{\mathsf{A}-\mathsf{B}}^\mathsf{T}]$$

where $\mathbf{Q} \in \mathbb{R}^{\mathsf{N}\times \mathsf{N}}$. Clearly, the decomposition $\mathbf{Q} = \bar{\mathbf{U}}\Sigma[\bar{\mathbf{V}}^\mathsf{T} \ \mathbf{0}_{\mathsf{B}-\mathsf{A}}^\mathsf{T}]$ also coincides with the SVD of $\mathbf{Q}$ since both $\bar{\mathbf{U}}$ and $[\bar{\mathbf{V}}^\mathsf{T} \ \mathbf{0}_{\mathsf{A}-\mathsf{B}}^\mathsf{T}]^\mathsf{T}$ are orthonormal matrices while Σ remains unchanged. In case when $\mathsf{M} > \mathsf{N}$, we can construct $\mathbf{Q}$ similarly by vertically stacking $\mathbf{P}$ with a zero matrix of dimensions $\mathsf{A} - \mathsf{B} \times \mathsf{N}$. Hence, generally speaking, let us denote $d_1 = \max(\mathsf{N}, \mathsf{M})$ and $d_2 = \min(\mathsf{N}, \mathsf{M})$.

The matrix $\mathbf{Q}$ is $\bar{\mu}$-incoherent where $\bar{\mu}r(d_1)^{-1} = \mu r d_2^{-1}$ implying that $\bar{\mu} = \mu d_1/d_2$. Moreover, we also have $\|\mathbf{Q}\|_\infty = \|\mathbf{P}\|_\infty$ implying that $\max_{ij} |\mathbf{P}_{ij}| = \max_{ij} |\mathbf{Q}_{ij}|$. Therefore, by invoking Lemma 21, the sample size must obey

$$p \geq C\kappa^4 \mu^2 d_1 d_2^{-2} \log^3 d_1 \quad \text{and} \quad \sigma = O\Big(\sqrt{\frac{p}{d\kappa^4 \mu r \log d}} \min_i \lambda_i\Big)$$

Then with probability at least $O(d_1^{-3})$, we can recover a matrix $\widehat{\mathbf{Q}}$ such that

$$\|\widehat{\mathbf{Q}} - \mathbf{Q}\|_\infty \leq O\Big(\frac{\sigma}{\min_i \lambda_i} \cdot \sqrt{\frac{\kappa^3 \mu \frac{d_1}{d_2} r d_1 \log d_1}{p}} \|\mathbf{P}\|_\infty\Big).$$

Using the fact that $\|\mathbf{P}\|_\infty \leq \max_i \lambda_i \|\bar{\mathbf{U}}\|_{2,\infty} \|\bar{\mathbf{V}}\|_{2,\infty} = \max_i \lambda_i \mu r/\sqrt{d_1 d_2} \implies \|\mathbf{P}\|_\infty / \min_i \lambda_i = \kappa \cdot \mu r/\sqrt{d_1 d_2}$, we obtain a matrix $\widehat{\mathbf{P}}$ such that

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty \leq O\Big(\frac{\sigma \mu r}{\sqrt{d_1 d_2}} \Big(\frac{d_1}{d_2}\Big)^{1/2} \sqrt{\frac{\kappa^5 \mu r d_1 \log d_1}{p}}\Big) = O\Big(\frac{\sigma}{\sqrt{d_2}} \Big(\frac{d_1}{d_2}\Big)^{1/2} \sqrt{\frac{\kappa^5 \mu^3 r^3 \log d_1}{p}}\Big).$$

$\square$

Lemma 23. *Let the matrix $\mathbf{P} \in \mathbb{R}^{N \times M}$ satisfy the conditions as stated in Lemma 22. Let $d_1 = \max(N, M)$, $d_2 = \min(N, M)$, $1 \geq p \geq C\kappa^4 \mu^2 d_1 d_2^{-2} \log^3 d_1$ for some sufficiently large constant $C > 0$, $\sigma = O\Big(\sqrt{\frac{p}{d\kappa^4 \mu r \log d}} \min_i \lambda_i\Big)$. For any pair of indices $(i, j) \in [N] \times [M]$,*

$$|\widehat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}| \leq O\Big(\frac{\sigma}{\sqrt{d_2}} \Big(\frac{d_1}{d_2}\Big)^{1/2} \sqrt{\frac{\kappa^5 \mu^3 r^3 \log d_1}{p}}\Big) \tag{25}$$

Proof. Let us assume that the matrix $\mathbf{P}$ is tall i.e. $N \geq M$. Now, let us partition the set of rows into $\frac{N}{M}$ groups by assigning each group uniformly at random to each row. Notice that the expected number of rows in each group is M and by using Chernoff bound, the number of rows in each group lies in the interval $[\frac{M}{2}, \frac{3M}{2}]$ with probability at least $1 - 2\exp(-M/12)$. When $N \leq M$, we partition the set of columns in a similar manner into M/N groups so the number of columns in each group lies in the interval $[\frac{N}{2}, \frac{3N}{2}]$ with probability at least $1 - 2\exp(-N/12)$. Hence, generally speaking, let us denote $d_1 = \max(N, M)$ and $d_2 = \min(N, M)$; we constructed d_1/d_2 sub-matrices of $\mathbf{P}$ denoted by $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \ldots, \mathbf{P}^{(d_1/d_2)}$. Let us analyze the guarantees on estimating $\mathbf{P}^{(1)}$. The analysis for other matrices follow along similar lines. Note that $\mathbf{P}^{(1)} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$ where $\mathbf{V}_{\mathsf{sub}}$ denotes the $N' \times r$ matrix where the rows in $\mathbf{V}_{\mathsf{sub}}$ corresponds to the N' rows in $\mathbf{V}$ assigned to $\mathbf{P}^{(1)}$.

First we will bound from below the minimum eigenvalue of the matrix $\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}}$. Note that every row of $\mathbf{V}$ is independently sampled with probability $p \triangleq d_2/d_1$ for the matrix $\mathbf{P}^{(1)}$. Hence, we have

$$\frac{1}{p}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}} = \frac{1}{p}\sum_{i \in [d_1]} \delta_i \mathbf{V}_i \mathbf{V}_i^\mathsf{T} = \sum_{i \in [d_1]} \mathbf{W}^{(i)}$$

where δ_i denotes the indicator random variable which is true when $\mathbf{V}_i$ (the ith row of $\mathbf{V}$) is chosen for $\mathbf{P}^{(1)}$ and $\mathbf{W}^{(i)} = \frac{1}{p}\delta_i \mathbf{V}_i \mathbf{V}_i^\mathsf{T}$. Notice that the random matrices $\mathbf{W}^{(i)}$ are independent with $\mathbb{E}\mathbf{W}^{(i)} = \mathbf{V}_i \mathbf{V}_i^\mathsf{T}$. Hence we define $\mathbf{Z}^{(i)} = \mathbf{W}^{(i)} - \mathbb{E}\mathbf{W}^{(i)}$ satisfying $\mathbb{E}\mathbf{Z}^{(i)} = 0$. Moreover, for all $i \in [d_1]$, we have $\|\mathbf{Z}^{(i)}\|_2 \leq \Big(1 + \frac{1}{p}\Big)\|\mathbf{V}_i \mathbf{V}_i^\mathsf{T}\|_2 \leq \Big(1 + \frac{1}{p}\Big)\max_i \|\mathbf{V}_i\|_2^2 \leq \frac{2\mu r}{p d_1}$. Next, we can show the following:

$$\|\sum_{i \in [d_1]} \mathbf{Z}^{(i)}(\mathbf{Z}^{(i)})^\mathsf{T}\|_2 \leq \|\Big(\frac{1}{p} - 1\Big)\sum_{i \in [d_1]} (\mathbf{V}_i \mathbf{V}_i^\mathsf{T})(\mathbf{V}_i \mathbf{V}_i^\mathsf{T})^\mathsf{T}\|_2 \leq \|\Big(\frac{1}{p} - 1\Big)\|\mathbf{V}_i\|^2 \sum_{i \in [d_1]} (\mathbf{V}_i \mathbf{V}_i^\mathsf{T})\|_2$$
$$\leq \frac{\mu r}{p d_1} \lambda_{\max}(\mathbf{V}^\mathsf{T}\mathbf{V}).$$

Similarly, we will also have

$$\|\sum_{i \in [d_1]} (\mathbf{Z}^{(i)})^\mathsf{T}\mathbf{Z}^{(i)}\|_2 \leq \frac{\mu r}{p d_1} \lambda_{\max}(\mathbf{V}^\mathsf{T}\mathbf{V}) \leq \frac{\mu r}{p d_1}$$

where we used that $\mathbf{V}^\mathsf{T}\mathbf{V}$ is orthogonal. Therefore, by using Bernstein's inequality for matrices, we have with probability at least $1 - \delta$,

$$\|\frac{1}{p}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}} - \mathbf{V}^\mathsf{T}\mathbf{V}\| \leq \frac{2\mu r}{3pd_1}\log\frac{2r}{\delta} + \sqrt{\frac{\mu r}{pd_1}\log\frac{2r}{\delta}}.$$

Hence, by using Weyl's inequality, we will have with probability $1 - \delta$

$$\lambda_{\min}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}}) \geq p - \frac{2\mu r}{3d_1}\log\frac{2r}{\delta} - \sqrt{\frac{p\mu r}{d_1}\log\frac{2r}{\delta}}.$$

Hence, with probability at least $1 - d_1^{-10}$, if $d_2 = \Omega(\mu r \log r d_1)$, then $\lambda_{\min}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}}) \geq p/2$ implying that $\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}}$ is invertible. Also, under the same condition, note that we can show similarly that $\lambda_{\max}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}}) \leq 3p/2$ implying that the condition number of each sub-matrix also stays $\kappa \cdot O(1)$ with high probability.

Clearly, $\mathbf{V}_{\mathsf{sub}}$ is not orthogonal and therefore, we have

$$\mathbf{P}^{(1)} = \mathbf{U}\mathbf{\Sigma}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{1/2}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T} = \mathbf{U}\widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$$

where $\widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}$ is the SVD of the matrix $\mathbf{\Sigma}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{1/2}$. Since $\widehat{\mathbf{U}}$ is orthogonal, $\widehat{\mathbf{U}}\mathbf{U}$ is orthogonal as well. Similarly, $(\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T})^\mathsf{T}$ is orthogonal as well whereas $\widehat{\mathbf{\Sigma}}$ is diagonal. Hence $\mathbf{U}\widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$ indeed corresponds to the SVD of $\mathbf{P}^{(1)}$ and we only need to argue about the incoherence of $\widehat{\mathbf{U}}\mathbf{U}$ and $\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}$. Notice that $\max_i \|(\mathbf{U}\widehat{\mathbf{U}})^\mathsf{T}\mathbf{e}_i\| \leq \|\mathbf{U}^\mathsf{T}\mathbf{e}_i\| \leq \sqrt{\frac{\mu r}{d_2}}$. On the other hand,

$$\max_i \|\widehat{\mathbf{V}}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{e}_i\| \leq \max_i \|(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})^{-1/2}\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{e}_i\|$$

$$\leq \frac{\|\mathbf{V}_{\mathsf{sub}}\|_{2,\infty}}{\sqrt{\lambda_{\min}(\mathbf{V}_{\mathsf{sub}}^\mathsf{T}\mathbf{V}_{\mathsf{sub}})}} \leq \frac{\|\mathbf{V}\|_{2,\infty}}{\sqrt{p/2}} \leq \sqrt{\frac{2\mu r}{d_2}}.$$

Hence, with probability $1 - 2\exp(-d_2/12) - O(d_1^{-3})$, we can recover an estimate $\widehat{\mathbf{P}}^{(1)}$ and apply Lemma 22 to conclude that

$$\|\mathbf{P}^{(1)} - \widehat{\mathbf{P}}^{(1)}\|_\infty = O\Big(\frac{\sigma}{\sqrt{d_2}}\Big(\frac{d_1}{d_2}\Big)^{1/2}\sqrt{\frac{\kappa^5\mu^3 r^3 \log d_1}{p}}\Big).$$

as long as the conditions stated in the Lemma are satisfied. Therefore, we can compute estimates of all the sub-matrices $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots, \mathbf{P}^{(d_1/d_2)}$ that have similar guarantees as above with probability at least $1 - 2d_1 d_2^{-1}\exp(-d_2/12) - O(d_1 d_2^{-4})$. Hence, by combining all the estimates, we can obtain a final estimate $\widehat{\mathbf{P}}$ of the matrix $\mathbf{P}$ that satisfies

$$\|\mathbf{P} - \widehat{\mathbf{P}}\|_\infty \leq O\Big(\frac{\sigma}{\sqrt{d_2}}\Big(\frac{d_1}{d_2}\Big)^{1/2}\sqrt{\frac{\kappa^5\mu^3 r^3 \log d_1}{p}}\Big).$$

$\square$

Corollary 3. *Consider an algorithm $\mathcal{A}$ that recommends random arms to users in each round (according to eq. 1) without recommending the same arm more than once to each user. Suppose the reward matrix $\mathbf{P}$ and parameters p, σ satisfies the conditions stated in Lemma 23. In that case, using $m = O\Big(\mathsf{M}p + \sqrt{\mathsf{M}p \log \mathsf{N}\delta^{-1}}\Big)$ recommendations per user, $\mathcal{A}$ is able to recover a matrix $\widehat{\mathbf{P}}$ such that for any $(i, j) \in [\mathsf{N}] \times [\mathsf{M}]$, we have equation 25 with probability exceeding $1 - \delta - O(d_2^{-3})$.*

Proof of Corollary 3. Recall that $d_1 = \max(\mathsf{N}, \mathsf{M})$ and $d_2 = \min(\mathsf{N}, \mathsf{M})$. Suppose we sample the elements in $[\mathsf{N}] \times [\mathsf{M}]$ each with some parameter $p = \Omega(\mu^2 d_2^{-1}\log^3 d_2)$ independently in order to sample a set $\Omega \in [\mathsf{N}] \times [\mathsf{M}]$ of indices. Let us define the event $\mathcal{F}_1$ which is true when the maximum number of indices observed in some row is more than $m = O\Big(\mathsf{M}p + \sqrt{\mathsf{M}p \log \mathsf{N}\delta^{-1}}\Big)$ i.e. $\max_{i \in [\mathsf{N}]} |(i, j) \in \Omega \mid j \in [\mathsf{M}]| \geq m$. We will bound the probability of the event $\mathcal{F}_1$ from above by using Chernoff bound. Let us denote the number of arms observed for user $i \in [\mathsf{N}]$ to be Y_i i.e. $Y_i = |(i, j) \in \Omega \mid j \in [\mathsf{M}]|$. Algorithm $\mathcal{A}$ can then obtain the noisy entries of $\mathbf{P}$ corresponding to the set Ω by doing the

following: in each round, conditioned on the event that the user $i \in [N]$ is sampled, if there is an unobserved tuple of indices $(i, j) \in \Omega$, then $\mathcal{A}$ will recommend j to user i and obtain an noisy observation $\mathbf{P}_{ij} + \mathbf{E}_{ij}$; on the other hand, if there no unobserved entry, then $\mathcal{A}$ will simply recommend a random arm j such that $(i, j) \notin \Omega$. Notice that each of the random variables $Y_1, Y_2, \ldots, Y_N \sim \mathrm{Binomial}(M, p)$ and are independent. By using Chernoff bound, we have that for each $i \in [N]$,

$$\Pr\left(\cup_{i \in [N]} |Y_i - Mp| \geq Mp\epsilon \right) \leq 2N \exp\left(-\frac{\epsilon^2 Mp}{3} \right))$$

$$\implies Y_i \leq Mp + O\left(\sqrt{Mp \log N\delta^{-1}} \right) \text{ for all } i \in [N]$$

with probability $1 - \delta$ implying that $\Pr(\mathcal{F}_1) \leq \delta$. Let $\mathcal{F}_2$ be the event when the recovered matrix $\widehat{\mathbf{P}}$ does not satisfy the guarantee on $\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty$ as stated in Lemma 23 equation 25 given a set of observed indices Ω sampled according to the aforementioned process. From Lemma 23, we know that $\Pr(\mathcal{F}_2) = O(d_2^{-3})$ where $d_1 = \max(N, M)$. Hence we can conclude $\Pr(\mathcal{F}_1 \cup \mathcal{F}_2) \leq \Pr(\mathcal{F}_1) + \Pr(\mathcal{F}_2) = \delta + O(d_2^{-3})$. This completes the proof of the corollary. □

Remark 8. *Note that the failure probability in Corollary 3, Lemmas 23, 22 and 21 is $O(d_2^{-3})$. However the constant 3 can be replaced by any arbitrary constant c for example $c = 100$ without any change in the guarantees on $\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty$. Hence, the guarantees presented in Lemma 23 and Corollary 3 hold with probability at least $1 - O(d_2^{-c})$ for any arbitrary constant c.*

Lemma 24 (Generalized Restatement of Lemma 1). *Let $\mathbf{P} = \bar{\mathbf{U}} \Sigma \bar{\mathbf{V}}^\mathsf{T} \in \mathbb{R}^{N \times M}$ such that $\bar{\mathbf{U}} \in \mathbb{R}^{N \times r}, \bar{\mathbf{V}} \in \mathbb{R}^{M \times r}$ and $\Sigma \triangleq \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r) \in \mathbb{R}^{r \times r}$ with $\bar{\mathbf{U}}^\mathsf{T} \bar{\mathbf{U}} = \bar{\mathbf{V}}^\mathsf{T} \bar{\mathbf{V}} = \mathbf{I}$ and $\|\bar{\mathbf{U}}\|_{2,\infty} \leq \sqrt{\mu r/N}, \|\bar{\mathbf{V}}\|_{2,\infty} \leq \sqrt{\mu r/M}$. Let $d_1 = \max(N, M)$ and $d_2 = \min(N, M)$ such that $1 \geq p \geq C\kappa^4 \mu^2 d_1 d_2^{-2} \log^3 d_1$ for some sufficiently large constant $C > 0$. In that case, for any positive integer $s > 0$, there exists an algorithm $\mathcal{A}$ that uses $m = O\left(s \log(MN\delta^{-1})(Mp + \sqrt{Mp \log N\delta^{-1}}) \right)$ recommendations per user such that $\mathcal{A}$ is able to recover a matrix $\widehat{\mathbf{P}}$ satisfying*

$$\|\mathbf{P} - \widehat{\mathbf{P}}\|_\infty = O\left(\frac{\sigma}{\sqrt{d_2}} \left(\frac{d_1}{d_2} \right)^{1/2} \sqrt{\frac{\kappa^5 \mu^3 r^3 \log d_1}{sp}} \right). \tag{26}$$

with probability exceeding $1 - O(\delta \log(MN\delta^{-1}))$ where $\frac{\sigma}{\sqrt{s}} = O\left(\sqrt{\frac{p}{d\kappa^4 \mu r \log d}} \min_i \lambda_i \right)$.

Proof. Consider the proof of Lemma 23 where we sampled a set of indices $\Omega \in [N] \times [M]$ and observed $\mathbf{P}_{ij} + \mathbf{E}_{ij}$ corresponding to the indices $(i, j) \in \Omega$ where $\mathbf{E}_{ij}$'s are independent zero mean sub-gaussian random variables and have a variance proxy of σ^2 (along with other entries of the reward matrix $\mathbf{P}$). From Corollary 3, we know that $m = O(Mp + \sqrt{Mp \log N\delta^{-1}})$ recommendations per user are sufficient to obtain the guarantees in Lemma 23 equation 25 with probability $1 - \delta$ if in each round t, we observe the reward corresponding to one arm for the sampled user $u(t)$ (see eq. 1). Thus, with m recommendations per user, we observe the noisy entries of $\mathbf{P}$ corresponding to the indices in a superset $\Omega' \supseteq \Omega$.

In our problem, an algorithm has the flexibility of recommending an arm more than once to the same user. Therefore, consider an algorithm $\mathcal{A}$ that uses ms recommendations per user to obtain noisy observations corresponding to each index in Ω' s times and uses the mean of observations corresponding to each index. In that case, the algorithm $\mathcal{A}$ observes $\mathbf{P}_{ij} + \widetilde{\mathbf{E}}_{ij}$ for all $(i, j) \in \Omega' \supseteq \Omega$ where $\widetilde{\mathbf{E}}_{ij}$ are independent zero mean sub-gaussian random variables with zero mean and variance proxy σ^2/s. Hence the effective variance σ^2/s should satisfy the upper bound on the noise variance implying that $\frac{\sigma}{\sqrt{s}} = O\left(\sqrt{\frac{pd_2}{\mu^3 \log d_2}} \|\mathbf{P}\|_\infty \right)$.

At this point, algorithm $\mathcal{A}$ can use Lemma 23 to recover an estimate $\widehat{\mathbf{P}}$ of the matrix $\mathbf{P}$ satisfying the following: for any $(i, j) \in [N] \times [M]$, we have from equation 25

$$|\widehat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}| \leq O\left(\frac{\sigma}{\sqrt{d_2}} \left(\frac{d_1}{d_2} \right)^{1/2} \sqrt{\frac{\kappa^5 \mu^3 r^3 \log d_1}{sp}} \right). \tag{27}$$

with probability exceeding $1 - \delta - O(d_2^{-c})$ for any arbitrary constant $c > 0$. We can set c such that the failure probability $\delta + O(d_2^{-c}) < 1/10$.

In order to boost the probability of success, the algorithm $\mathcal{A}$ can repeat the entire process $f = \log(\mathsf{MN}\delta^{-1})$ times in order to obtain f estimates $\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}$. The total number of rounds needed to compute these estimates if we use the observation model in equation 1 is at most mf with probability at least $1 - \delta f$. Furthermore, these estimates are independent and therefore, we can compute a final estimate $\widehat{\mathbf{P}}$ by computing the entry-wise median of the matrix estimates $\widehat{\mathbf{P}}^{(1)}, \widehat{\mathbf{P}}^{(2)}, \ldots, \widehat{\mathbf{P}}^{(f)}$ i.e. for all $(i,j) \in [\mathsf{N}] \times \mathsf{M}$, we compute $\widehat{\mathbf{P}}_{ij} = \mathsf{median}(\widehat{\mathbf{P}}_{ij}^{(1)}, \ldots, \widehat{\mathbf{P}}_{ij}^{(f)})$. Since, with probability $9/10$, each estimates satisfy the guarantee in 27, we can again apply Chernoff bound that $\mathbf{P}$ satisfies the guarantee in 27 with probability $1 - \delta/\mathsf{MN}$. Now taking a union bound over all indices, we must have that

$$\|\widehat{\mathbf{P}} - \mathbf{P}\|_\infty \leq O\Big(\frac{\sigma}{\sqrt{d_2}}\Big(\frac{d_1}{d_2}\Big)^{1/2}\sqrt{\frac{\kappa^5\mu^3 r^3 \log d_1}{sp}}\Big) \tag{28}$$

with probability at least $1 - \delta$. Therefore the total failure probability is $1 - O(\delta + \delta\log(\mathsf{MN}\delta^{-1}))$.

$\square$

We obtain the statement of Lemma 1 by substituting $r = O(1), \kappa = O(1)$. Moreover, by another simple application of the Chernoff bound again, we can ensure with probability $1 - \widetilde{O}(\delta)$ that the total number of recommendations (so that the number of recommendations per user is m as described in Lemma 24) is $\widetilde{O}(sp\mathsf{MN})$ for which the guarantees in eq. 26 is satisfied. Now we characterize m_ℓ in this general setting:

Lemma 25. *Let us fix $\Delta_{\ell+1} > 0$ and condition on the event $\mathcal{E}^{(\ell)}$. Suppose Assumptions 1 i satisfied. In that case, in phase ℓ, by using*

$$m_\ell = \widetilde{O}\Big(\frac{\sigma^2 \mathsf{C}^2(\mathsf{C} \bigvee \mu\alpha^{-1})^3 r^3 \kappa^5 \tau^{9/2}}{\Delta_{\ell+1}^2}\Big(\mathsf{N} \bigvee \mathsf{AC}\tau\Big)\Big)$$

rounds, we can compute an estimate $\widetilde{\mathbf{P}}^{(\ell)} \in \mathbb{R}^{\mathsf{N} \times \mathsf{M}}$ such that with probability $1 - \delta$, we have

$$\Big\|\widetilde{\mathbf{P}}^{(\ell)}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\Big\|_\infty \leq \Delta_{\ell+1} \textit{ for all } i \in [a_\ell] \textit{ satisfying } \Big|\mathcal{N}^{(i,\ell)}\Big| \geq \gamma\mathsf{C}. \tag{29}$$

Proof of Lemma 25. We are going to use Lemma 1 in order to compute an estimate $\widetilde{\mathbf{P}}^{(\ell)}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$ of the sub-matrix $\mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$ satisfying $\Big\|\widetilde{\mathbf{P}}^{(\ell)}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\Big\|_\infty \leq \Delta_{\ell+1}$. From Lemma 1, we know that by using $\widetilde{O}(sp\mathsf{MN})$ rounds restricted to users in $\mathcal{M}^{(\ell,i)}$ such that with probability at least $1 - \delta$,

$$\Big\|\widetilde{\mathbf{P}}^{(\ell)}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}} - \mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}\Big\|_\infty = O\Big(\frac{\sigma}{\sqrt{d_2}}\Big(\frac{d_1}{d_2}\Big)^{1/2}\sqrt{\frac{\widetilde{\kappa}^5\widetilde{\mu}^3 r^3 \log d_1}{sp}}\Big).$$

where $d_2 = \min(|\mathcal{M}^{(\ell,i)}|, |\mathcal{N}^{(\ell,i)}|)$ and $\widetilde{\mu}, \widetilde{\kappa}$ is the incoherence factor and condition number of the matrix $\mathbf{P}_{\mathcal{M}^{(\ell,i)},\mathcal{N}^{(\ell,i)}}$. In order for the right hand side to be less than $\Delta_{\ell+1}$, we can set $sp = O\Big(\frac{\sigma^2\widetilde{\kappa}^5\widetilde{\mu}^3 r^3 \log d_1}{\Delta_{\ell+1}^2 d_2}\Big)$. Since the event $\mathcal{E}^{(\ell)}$ is true, we must have that $\Big|\mathcal{M}^{(\ell,i)}\Big| \geq \mathsf{N}/(\mathsf{C}\tau)$; hence $d_2 \geq \min\Big(\frac{\mathsf{N}}{\mathsf{C}\tau}, \Big|\mathcal{N}^{(\ell,i)}\Big|\Big)$. Therefore, we must have that

$$m_\ell = O\Big(\frac{\sigma^2\mathsf{C}^2\tau^2\widetilde{\kappa}^5\widetilde{\mu}^3 r^3 \log \mathsf{M}}{\Delta_{\ell+1}^2}\max\Big(\mathsf{N}, \mathsf{MC}\tau\Big)\log^2(\mathsf{MNC}\delta^{-1})\Big)$$

where we take a union bound over all sets comprising the partition of the users $[\mathsf{N}]$ (at most C of them). Finally, from Lemma 3, we know that $\widetilde{\mu}$ can be bounded from above by $\max(\mathsf{C}, 2\mu/\alpha)$; from lemma 2, we can say that $\widetilde{\kappa}$ can be bounded from above by $\kappa\sqrt{\tau}$ which we can use to say that

$$m_\ell = \widetilde{O}\Big(\frac{\sigma^2\mathsf{C}^2(\mathsf{C} \bigvee \mu\alpha^{-1})^3 r^3 \kappa^5 \tau^{9/2}}{\Delta_{\ell+1}^2}\Big(\mathsf{N} \bigvee \mathsf{AC}\tau\Big)\Big)$$

to complete the proof of the lemma. $\square$

Plugging this m_ℓ in the proof of Theorem 1, we can show that the Algorithm 1 guarantees the following regret

$$\mathsf{Reg}(\mathsf{T}) = \widetilde{O}(\sqrt{\mathsf{TV}}) + \sqrt{\mathsf{BT}}\sigma) \tag{30}$$

where $\mathsf{V} = \widetilde{O}\Big(\sigma^2\mathsf{C}^2(\mathsf{C} \bigvee \mu)^3 r^3 \kappa^5 \tau^{9/2}\Big(\mathsf{N} + \mathsf{M}\Big)\Big)$. This result follows by proceeding through the exact same steps as in the proof of Theorem 1.