FREQUENCY-AGNOSTIC ONLINE CLUSTERING OF BANDITS VIA LOCAL CERTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

For a linear bandit problem, in each sequential decision round, an agent picks an action vector and receives a stochastic reward whose expectation is the inner product of the action vector and an unknown preference vector. Online clustering of linear bandits improves sample efficiency by grouping agents with common preferences into clusters and pooling their observations. However, prior analyses often tie the benefits of collaboration to the identification of the full cluster structure, so rarely observed agents can delay these benefits even for agents whose preferences are already estimated accurately. To address this limitation, we introduce *Local Certification for Clustering of Bandits* (LC-CLUB), a frequency-agnostic collaboration framework that enables certified sharing before global cluster identification and requires no knowledge of arrival frequencies. Each agent first learns from its own observations and, once its estimate is accurate enough to be certified, begins sharing with certified agents in the same cluster. Our regret analysis separates each agent's pre-certification interactions from its subsequent cluster-level learning, so a rarely observed agent contributes only its own few interactions to the clustering term. With this new algorithm design and analysis, LC-CLUB eliminates the uniform-arrival assumption imposed in prior analyses, has no inverse dependence on the minimum agent-arrival probability, and requires no within-cluster frequency alignment, substantially broadening its applicability to heterogeneous-arrival settings with highly unequal agent arrival rates. It also removes an explicit factor of the linear-bandit dimension d from the coefficient of $\log T$ in the clustering term, leaving a horizon-independent remainder. We further show that, under Gaussian-perturbed adversarial contexts, predictable convex action-selection scores, including LinUCB, accumulate information in every direction at a linear rate with high probability, so local certification needs no separate uniform-exploration phase. Experimentally, LC-CLUB-L reduces cumulative regret by 12.8% on average and up to 38% relative to the strongest baseline.

1 INTRODUCTION

Linear bandits provide a fundamental framework for sequential decision-making, where the learner selects an action and observes feedback only for that action (Li et al., 2010; Chu et al., 2011; Abbasi-Yadkori et al., 2011). In systems serving multiple agents, such as users of a recommendation platform, learning an independent linear bandit model for each agent can be sample-inefficient when preferences are shared (Cherkaoui et al., 2025). Online clustering of bandits exploits this shared structure by *pooling* observations among agents inferred to have a common preference parameter, i.e., in the same cluster (Gentile et al., 2014; Li et al., 2019; 2025), with applications in recommendation and dynamic pricing (Li et al., 2016; Yang et al., 2020; Miao et al., 2022). If the true pooling were known, the learner would need to learn one model per cluster rather than one model per agent. The central difficulty is deciding *when pooling is statistically justified*: pooling observations from agents with different clusters can bias the shared estimate, whereas delaying collaboration leaves useful observations unshared.

Many existing clustering-of-bandits algorithms combine upper confidence bound (UCB)-based arm selection with adaptive inference of the underlying agent clusters (Gentile et al., 2014; Li & Zhang, 2018). Graph-based methods initially share observations across all agents and progressively separate them as evidence of preference differences accumulates. A common two-stage strategy in their

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

regret analyses is to first establish high-probability identification of the entire cluster partition and then derive cluster-level regret guarantees for subsequent rounds. For example, a recent work, Uni-CLUB (Li et al., 2025), introduces explicit uniform exploration to collect sufficiently informative observations, but its regret analysis likewise relies on a global cluster-identification guarantee.

This two-stage analytical framework becomes particularly restrictive under heterogeneous agent arrivals. Some agents with high arrival rates may accumulate enough observations for reliable cluster identification very early, while poorly observed agents prevent cluster identification and thus delay the reach of the second stage. For example, Wang et al. (2023a) and Yang et al. (2024) state their main results under uniform arrivals and discuss extensions to nonuniform stochastic arrivals with a positive minimum arrival probability $p_{\min} > 0$. These extensions still rely on global cluster identification arguments, and consequently, their regret bounds retain an inverse dependence on $p_{\min}$. SCLUB (Li et al., 2019) avoids an explicit dependence on $1/p_{\min}$ by incorporating arrival frequencies into its clustering structure. However, its guarantee assumes that agents sharing the same preference parameter also have identical arrival frequencies. This coupling excludes, for example, two agents with the same preference vector but substantially different activity levels on a recommendation platform.

These limitations reveal a mismatch between global cluster-identification and regret minimization: a rarely observed agent participates in only a few decision rounds, yet the cluster-identification cost incurred by its unresolved cluster membership can dominate the overall identification cost. Ideally, the regret incurred before identifying an agent's cluster should scale with its own interaction count, and agents that have enough data should not wait for the others. This motivates our first question:

> *Can we obtain cluster-level regret guarantees without requiring global cluster identification under heterogeneous agent arrivals?*

In this paper, we show that global cluster identification is not a prerequisite for cluster-level regret guarantees. Our key insight is that reliable collaboration requires identifying relationships only among agents that have accumulated sufficient observations, rather than recovering the cluster membership of every agent. This shifts the focus from the time needed for global cluster identification to the evidence accumulated by individual agents. Building on this insight, we introduce LC-CLUB-U (Local Certification for Clustering of Bandits with Uniform Exploration), which enables agents to begin sharing based on their own accumulated observations (Section 4). Given a valid lower bound $\widetilde{\gamma} > 0$ on the separation between distinct preference parameters, each agent initially explores uniformly to estimate its preference parameter from its own observations and is admitted to a graph of certified agents only after its local estimate is certified as sufficiently accurate. We prove that, with high probability, this graph recovers the true cluster structure restricted to the agents admitted so far. Certified agents therefore use pooled LinUCB within their respective clusters, while agents with insufficient observations remain outside the shared models and do not delay others' admission.

Our regret analysis separates each agent's pre-certification interactions from its subsequent cluster-level learning: an agent contributes to the identification term only through its own interactions before certification, and once certified it immediately joins cluster-level learning without waiting for the remaining agents. The guarantee holds under stochastic arrivals and arbitrary oblivious arrival sequences fixed before interaction, without a positive minimum arrival probability or within-cluster frequency alignment. We call this property *frequency-agnostic*: the bound depends on realized arrival counts but requires no knowledge of, or condition on, arrival frequencies.

Local certification removes the need for global cluster identification, but LC-CLUB-U still relies on an initial uniform-exploration phase to obtain sufficiently accurate individual estimates. By selecting arms uniformly regardless of accumulated reward information, this phase can incur substantial early regret even in favorable environments where reward-oriented actions themselves provide sufficiently informative observations. This raises a second question:

> *when can reward-oriented arm selection itself collect enough information for certification, without the uniform-exploration phase?*

The challenge is that local certification requires information in every parameter direction, while reward-oriented arm selection need not preserve the diversity of the candidate contexts. Although Wang et al. (2023a) establish linear information growth without the restrictive numerical tail condition, the information-growth coefficient in their theoretical guarantee decays exponentially with

the number of candidate arms, leading to an exponentially worse regret term for cluster identification. We therefore seek conditions on context generation under which the actions selected to pursue rewards also ensure sufficiently rapid information growth in every direction.

We show that a separate uniform-exploration phase is unnecessary under the Gaussian-perturbed adversarial context model. For finite-valued convex scoring functions determined before the current candidate set is revealed, the selected arm's conditional second-moment matrix preserves information in every direction. This yields high-probability linear growth of the minimum eigenvalue of the selected-arm design matrix after a finite initial period. The result applies to the full LinUCB index even when its current parameter estimate is inaccurate. It leads to LC-CLUB-L, which uses personal LinUCB before certification and pooled LinUCB afterwards (Section 5).

Main contributions. Our contributions are summarized as follows.

- *Local-certification algorithms for early collaboration.* We introduce an agent-wise local-certification framework and instantiate it in graph-based clustering algorithms. In the uniform-exploration instantiation LC-CLUB-U, each agent initially collects its own observations and becomes eligible for collaboration once its local estimate is sufficiently accurate. Certified agents then pool observations within their identified clusters, while agents with insufficient evidence remain outside the shared models.
- *Regret guarantees with no need for global cluster identification.* We develop an expanding partial-oracle analysis that tracks the true clustering restricted to the agents certified so far and yields regret bounds of the form

$$\mathrm{Reg}(T) = \widetilde{\mathcal{O}}\left(\sum_{i=1}^{u} \min\{T_i, n_{\mathrm{cert}}\} + d\sqrt{mT}\right),$$

  where $\widetilde{\mathcal{O}}$ hides logarithmic factors and problem-dependent constants independent of T, T_i is agent i's realized arrival count, and n_{cert} is the common uniform-exploration threshold, calibrated using the known separation lower bound $\widetilde{\gamma}$. The two terms correspond to individual pre-certification interactions and subsequent cluster-level learning. Unlike prior analyses, the bound has no $1/p_{\min}$ factor, needs no within-cluster frequency alignment, and holds for any oblivious arrival sequence.
- *Quantitative information growth without separate uniform exploration.* Under Gaussian-perturbed adversarial contexts, predictable, finite-valued convex scores preserve an $\Omega(1/\log(2K))$ fraction of the perturbation covariance in the selected arm's conditional second moment, where K is the number of candidate arms. This yields high-probability linear growth of the minimum eigenvalue of the selected-arm design matrix after a finite initial period, providing the personal evidence needed for certification and a regret guarantee for LC-CLUB-L (Theorems 2 and 3).

2 Related Work

Online clustering of bandits. CLUB (Gentile et al., 2014) introduced online clustering of linear bandits by learning a graph over agents and sharing observations within connected components. SCLUB (Li et al., 2019) replaces the graph with a dynamically maintained set partition using split and merge operations. Subsequent work considers context-dependent or local neighborhoods (Gentile et al., 2017; Ban & He, 2021), cascading feedback (Li & Zhang, 2018), federated learning (Liu et al., 2022; Yang et al., 2024), misspecified or corrupted agents (Wang et al., 2023a;b), dueling feedback (Wang et al., 2025), and offline data (Liu et al., 2025), collaborative filtering (Li et al., 2016), conversational recommendation (Wu et al., 2021), and neural bandits (Ban et al., 2024). UniCLUB and UniSCLUB (Li et al., 2025) introduce an explicit uniform-exploration prefix to guarantee sufficiently informative agent-level designs under weaker context assumptions. These works develop increasingly flexible mechanisms for discovering agent structure and sharing observations; our focus is complementary: we ask when an individual agent has accumulated enough reliable evidence to safely participate in collaboration, without waiting for population-wide recovery.

Clustering under heterogeneous agent arrivals. Table 1 summarizes the arrival assumptions underlying the main regret guarantees of representative online clustering bandit algorithms. Most existing regret analyses for online clustering of bandits rely on stochastic agent-arrival models.

Table 1: Comparison with existing algorithms under different agent-arrival and context-diversity assumptions. Since the cluster-level regret terms are similar across methods, we report only the regret incurred by cluster identification.

Algorithm	**Arrival assumption**	**Diversity assumption**	**Regret incurred by clustering**
CLUB (Gentile et al., 2014), CLUB-cascade (Li & Zhang, 2018), FCLUB (Liu et al., 2022)	Uniform arrivals	$\lambda_{\min}(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top]) = \lambda_x$, $(\boldsymbol{z}^\top \boldsymbol{X})^2$ is σ^2-sub-Gaussian, $\sigma^2 \le \frac{\lambda_x^2}{8\log(4K)}$.	$\widetilde{O}\Big(u\Big(\frac{d}{\gamma^2\lambda_x} + \frac{1}{\lambda_x^2}\Big)\Big)$
SCLUB (Li et al., 2019)	Equal within-cluster frequencies; gap $\gamma_p > 0$ between unequal frequencies.	$\lambda_{\min}(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top]) = \lambda_x$, $(\boldsymbol{z}^\top \boldsymbol{X})^2$ is σ^2-sub-Gaussian, $\sigma^2 \le \frac{\lambda_x^2}{8\log(4K)}$.	$\widetilde{O}\Big(\frac{1}{\gamma_p^2} + u\Big(\frac{d}{\gamma^2\lambda_x} + \frac{1}{\lambda_x^2}\Big)\Big)$
RCLUMB (Wang et al., 2023a), RCLUB-WCU (Wang et al., 2023b), FedC^{3}UCB-H (Yang et al., 2024)	Uniform arrivals	$\lambda_{\min}(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top]) = \lambda_x$, $(\boldsymbol{z}^\top \boldsymbol{X})^2$ is σ^2-sub-Gaussian.	$\widetilde{O}\Big(u\Big(\frac{d}{\widetilde{\gamma}^2\lambda_x} + \frac{1}{\widetilde{\lambda}_x^2}\Big)\Big)$
UniCLUB, UniSCLUB† (Li et al., 2025)	Uniform arrivals	$\lambda_{\min}(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top]) = \lambda_x$.	$\widetilde{O}\Big(\frac{ud}{\widetilde{\gamma}^2\lambda_x}\Big)$
LC-CLUB (ours)†	Stochastic or oblivious arrivals	$\lambda_{\min}(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top]) = \lambda_x$.	$\widetilde{O}\Big(\frac{u}{\widetilde{\gamma}^2\lambda_x}\Big)$

† Requires a supplied valid lower bound on the cluster gap.
K, u, and d denote the number of arms, the number of agents, and the dimension, respectively. $\boldsymbol{z} \in \mathbb{R}^d$ denotes an arbitrary unit vector. $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue. The cluster-separation parameter γ and its valid lower bound $\widetilde{\gamma}$ are defined in Assumption 1. The random context $\boldsymbol{X}$ and the context-diversity parameter λ_x are defined in Assumption 2. $\lambda_x = O(\frac{1}{d})$, $\widetilde{\lambda}_x = O(\frac{1}{d^{2K+1}})$.

CLUB (Gentile et al., 2014) assumes uniform arrivals. SCLUB (Li et al., 2019) extends the model to a fixed nonuniform arrival distribution and obtains a regret guarantee without an explicit $1/p_{\min}$ dependence, but couples arrival frequencies with the preference clusters: agents sharing the same preference parameter are required to have the same, or sufficiently close, arrival frequencies. Hence, its $p_{\min}$-free guarantee does not cover general heterogeneous frequencies within a preference cluster. Later variants largely retain uniform arrivals in their main guarantees. In particular, UniCLUB/UniSCLUB (Li et al., 2025) and the robust clustering methods of Wang et al. (2023b) assume uniform arrivals. The misspecified-agent-model analysis of Wang et al. (2023a) and the asynchronous federated clustering framework of Yang et al. (2024) likewise state their main results under uniform agent arrivals. Both works note that their analyses can be extended to general nonuniform stochastic arrivals when every agent has a positive minimum arrival probability $p_{\min} > 0$. Thus, under general heterogeneous stochastic arrivals, existing guarantees either impose additional structure on arrival frequencies or retain dependence on the least-frequently arriving agent. Our analysis removes the arrival-distribution assumption and accommodates arbitrary oblivious arrival sequences, while requiring neither a minimum arrival probability nor within-cluster frequency alignment.

Context diversity and information growth. A complementary line of work studies when reward-driven action selection itself generates sufficiently informative designs, including greedy policies under Gaussian or local anti-concentration conditions (Han et al., 2020; Kim & Oh, 2024). This issue is particularly important in clustering bandits because the collected observations must support not only low-regret decision making but also accurate structural identification. UniCLUB and UniSCLUB (Li et al., 2025) guarantee such information through explicit uniform exploration.

3 Model and Assumption

Notation. We use bold lowercase letters for vectors and bold uppercase letters for matrices. For any positive integer n, let $[n] \coloneqq \{1, \dots, n\}$, and let $|\mathcal{A}|$ denote the cardinality of a set $\mathcal{A}$. For vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, we write $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y}$ and use $\|\boldsymbol{x}\|$ to denote the Euclidean norm. For a positive semidefinite matrix $\boldsymbol{V}$, we define $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\boldsymbol{V}} \coloneqq \boldsymbol{x}^\top \boldsymbol{V} \boldsymbol{y}$ and $\|\boldsymbol{x}\|_{\boldsymbol{V}} \coloneqq \sqrt{\boldsymbol{x}^\top \boldsymbol{V} \boldsymbol{x}}$. We use $\lambda_{\min}(\boldsymbol{V})$ and $\lambda_{\max}(\boldsymbol{V})$ to denote the minimum and maximum eigenvalues of $\boldsymbol{V}$, respectively.

We study u agents partitioned into m unknown clusters $\mathcal{I}_1, \ldots, \mathcal{I}_m$. Agent i belongs to cluster $j(i)$ and has parameter $\boldsymbol{\theta}_i = \boldsymbol{\theta}^{j(i)} \in \mathbb{R}^d$, with $\|\boldsymbol{\theta}_i\|_2 \le S$ for a known $S > 0$. At each round t, agent i_t arrives and the learner is presented with a finite set of K available arms $\mathcal{A}_t$, with $|\mathcal{A}_t| = K$. Each arm $a \in \mathcal{A}_t$ is associated with a context vector $\boldsymbol{x}_{t,a} \in \mathbb{R}^d$, and we denote the corresponding context set by $\mathcal{D}_t = \{\boldsymbol{x}_{t,a} : a \in \mathcal{A}_t\}$. We assume $\|\boldsymbol{x}_{t,a}\|_2 \le L$ for every t and $a \in \mathcal{A}_t$. The learner selects an arm $a_t \in \mathcal{A}_t$ for agent i_t, writes $\boldsymbol{x}_t := \boldsymbol{x}_{t,a_t}$, and observes reward $r_t = \boldsymbol{x}_t^\top \boldsymbol{\theta}_{i_t} + \eta_t$, where η_t is conditionally zero-mean and R-sub-Gaussian given the interaction history, for a known $R > 0$. Let $\mathcal{F}_{t-1}$ contain the history through round $t-1$, and let $\mathcal{H}_t$ additionally contain the agent i_t, the context set $\mathcal{D}_t$, and the chosen action, before observing r_t. The noise obeys

$$\mathbb{E}[\eta_t \mid \mathcal{H}_t] = 0, \qquad \mathbb{E}[e^{s\eta_t} \mid \mathcal{H}_t] \le e^{s^2R^2/2} \quad (s \in \mathbb{R}).$$

The pseudo-regret is

$$\mathrm{Reg}(T) = \sum_{t=1}^{T} \left[\max_{a \in \mathcal{A}_t} \boldsymbol{x}_{t,a}^\top \boldsymbol{\theta}_{i_t} - \boldsymbol{x}_t^\top \boldsymbol{\theta}_{i_t} \right], \tag{1}$$

with per-round bound

$$0 \le \max_{a \in \mathcal{A}_t} \boldsymbol{x}_{t,a}^\top \boldsymbol{\theta}_{i_t} - \boldsymbol{x}_t^\top \boldsymbol{\theta}_{i_t} \le 2LS =: b_{\max}. \tag{2}$$

We assume the agents and clusters satisfy the assumptions as follows:

Assumption 1 (Well-separated clusters)**.** All agents within the same cluster $\mathcal{I}_j$ share a common preference vector $\boldsymbol{\theta}^j$. The preference vectors of distinct clusters are separated by a fixed but unknown minimum gap $\gamma := \min_{j \ne j'} \|\boldsymbol{\theta}^j - \boldsymbol{\theta}^{j'}\|_2 > 0$. Following UniCLUB (Li et al., 2025), the learner is given a valid lower bound $\widetilde{\gamma}$ satisfying $0 < \widetilde{\gamma} \le \gamma$. The algorithms use only $\widetilde{\gamma}$ and do not require knowledge of the exact separation γ.

Assumption 2 (Stochastic contexts)**.** At each round t, after agent i_t is revealed, the K candidate contexts $\boldsymbol{x}_{t,1}, \ldots, \boldsymbol{x}_{t,K} \in \mathbb{R}^d$ are drawn independently from a fixed distribution $\mathcal{P}_x$, conditional on $\mathcal{F}_{t-1}$ and i_t. For $\boldsymbol{x} \sim \mathcal{P}_x$, assume $\|\boldsymbol{x}\|_2 \le L$ almost surely and define

$$\Sigma_x := \mathbb{E}_{\boldsymbol{x} \sim \mathcal{P}_x}[\boldsymbol{x}\boldsymbol{x}^\top] \succ 0, \qquad \lambda_x := \lambda_{\min}(\Sigma_x) > 0.$$

Remark 1 (Approximate clusters)**.** A relaxed model allows $\|\boldsymbol{\theta}_i - \boldsymbol{\theta}^{j(i)}\|_2 \le \epsilon$, so agents may differ slightly from their cluster representative. Appendix A.4 shows how to adjust the certification threshold when this variation is sufficiently small relative to the representative separation.

Remark 2 (Arrival models)**.** We consider two arrival models. Under *stochastic arrivals*, the agents are generated according to a possibly nonstationary distribution, including i.i.d. draws from $\boldsymbol{p} = (p_1, \ldots, p_u)$ or a pre-specified schedule $\boldsymbol{p}_t$. Under *oblivious arrivals*, the entire sequence $i_{1:T}$ is fixed independently of the interaction history and can be arbitrary or highly imbalanced. Adaptive arrivals depending on observed contexts, actions, or rewards are not considered. For stochastic arrivals, we assume that, conditional on almost every complete arrival sequence, the context-generation and reward-noise conditions remain valid. No minimum arrival probability or relationship among arrival frequencies within the same cluster is assumed. Thus, the method is *frequency-agnostic*.

4 LC-CLUB-U: A Locally Certified Graph Algorithm

In this section, we present *Local Certification for Clustering of Bandits with Uniform Exploration* (LC-CLUB-U), which certifies each agent individually before enabling information sharing. It maintains a *certified graph* $G = (\mathcal{C}, E)$, initially empty, whose vertices are certified agents and whose edges join pairs declared to share a preference cluster.

Personal exploration and certification. Each agent maintains personal statistics $(\boldsymbol{S}_i, \boldsymbol{b}_i)$, count n_i, and estimate $\widehat{\boldsymbol{\theta}}_i$, updated after every own observation, including after certification (line 10). An uncertified agent selects arms uniformly for its first n_{cert} own interactions (line 4), where n_{cert} is fixed in advance and counted on its local clock. This yields a well-conditioned personal design without synchronizing agents. Immediately after the n_{cert}-th observation (line 11), its latest personal estimate is compared with the latest estimate of every certified agent; an edge is added when their distance is at most $\widetilde{\gamma}/2$ (line 12), and the agent joins $\mathcal{C}$ (line 13). Edges are never revised, so each pair is compared at most once. If all personal estimates are within $\widetilde{\gamma}/4$ of their parameters, the test recovers exactly the edges between certified cluster mates.

Algorithm 1: LC−CLUB−U: Local Certification with Uniform Exploration

Input: horizon T; confidence level $\delta \in (0,1)$; regularization parameter $\lambda > 0$; bounds L, R, S; diversity parameter λ_x (or a lower bound); gap lower bound $\widetilde{\gamma}$.

Initialize: β_t by (3) and n_{cert} by (4); $G = (\mathcal{C}, E) = (\emptyset, \emptyset)$; for every i, $n_i = 0$, $(\boldsymbol{S}_i, \boldsymbol{b}_i, \widehat{\boldsymbol{\theta}}_i) = (\mathbf{0}, \mathbf{0}, \mathbf{0})$.

1 **for** $t = 1, \ldots, T$ **do**
2 Receive agent i_t (set $i \leftarrow i_t$) and contexts $\{\boldsymbol{x}_{t,a} : a \in \mathcal{A}_t\}$
3 **if** $i \notin \mathcal{C}$ **then** *// uncertified agent*
4 Draw a_t uniformly from $\mathcal{A}_t$
5 **else**
6 $\mathcal{N}[i] \leftarrow \{i\} \cup \{j \in \mathcal{C} : \{i, j\} \in E\}$ *// closed certified neighborhood*
7 $\boldsymbol{V}_i \leftarrow \lambda \boldsymbol{I} + \sum_{j \in \mathcal{N}[i]} \boldsymbol{S}_j$, $\widetilde{\boldsymbol{\theta}}_i \leftarrow \boldsymbol{V}_i^{-1} \sum_{j \in \mathcal{N}[i]} \boldsymbol{b}_j$ *// pooled estimate*
8 $a_t \leftarrow \arg\max_{a \in \mathcal{A}_t} \{\boldsymbol{x}_{t,a}^\top \widetilde{\boldsymbol{\theta}}_i + \beta_t \sqrt{\boldsymbol{x}_{t,a}^\top \boldsymbol{V}_i^{-1} \boldsymbol{x}_{t,a}}\}$, breaking ties by smallest index
9 Observe r_t and set $\boldsymbol{x}_t \leftarrow \boldsymbol{x}_{t,a_t}$
10 $\boldsymbol{S}_i \leftarrow \boldsymbol{S}_i + \boldsymbol{x}_t \boldsymbol{x}_t^\top$, $\boldsymbol{b}_i \leftarrow \boldsymbol{b}_i + r_t \boldsymbol{x}_t$, $n_i \leftarrow n_i + 1$, $\widehat{\boldsymbol{\theta}}_i \leftarrow (\lambda \boldsymbol{I} + \boldsymbol{S}_i)^{-1} \boldsymbol{b}_i$
11 **if** $i \notin \mathcal{C}$ *and* $n_i = n_{\mathrm{cert}}$ **then**
12 $E \leftarrow E \cup \{\{i, \ell\} : \ell \in \mathcal{C},\ \|\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_\ell\|_2 \le \widetilde{\gamma}/2\}$
13 $\mathcal{C} \leftarrow \mathcal{C} \cup \{i\}$

Certified collaboration. A certified agent i pools the statistics of itself and its direct certified neighbors $\mathcal{N}[i]$ (lines 6–7), then selects an arm by LinUCB (line 8). Adding an edge directly changes only its endpoints' pools; connected-component pooling can instead join entire components (Gentile et al., 2014; Li et al., 2025).

Certification threshold. From its inputs, Algorithm 1 computes

$$\beta_t := R\sqrt{d \log\Big(1 + \frac{tL^2}{d\lambda}\Big) + 2\log\frac{2u}{\delta}} + \sqrt{\lambda} S, \tag{3}$$

$$n_{\mathrm{cert}} := \min\left\{\left\lceil \frac{8}{\lambda_x} \max\left\{L^2 \log\frac{2ud}{\delta},\ \frac{4\beta_T^2}{\widetilde{\gamma}^2}\right\}\right\rceil,\ n_{\mathrm{ref}}\right\}. \tag{4}$$

Here n_{ref} is the refined threshold (29) of Appendix A.2. Each of the two terms alone ensures that, with probability at least $1 - \delta$, every certified agent's personal estimate stays within $\widetilde{\gamma}/4$ of its parameter through T (Lemmas 2 and 6). The minimum equals one of them and is computed from known inputs before interaction, so the lemma for that term applies. For $\delta = 1/T$ and $u \le T$, the first term is $O(d \log T/(\widetilde{\gamma}^2 \lambda_x))$; the second also counts information collected after certification and is $O(\log T/(\widetilde{\gamma}^2 \lambda_x)) + C_\gamma$, where the horizon-independent remainder C_γ (35) depends on d and K.

Theorem 1 (Regret of LC-CLUB-U)**.** *Under Assumptions 1–2 and the noise and arrival models of Section 3, LC-CLUB-U with $\lambda \ge L^2$ and $\delta = 1/T$ satisfies,*

$$\mathbb{E}[\mathrm{Reg}(T)] = O\left(\frac{u}{\widetilde{\gamma}^2 \lambda_x} \log T + d\sqrt{mT} \log T\right). \tag{5}$$

Comparison with prior guarantees. Table 1 separates arrival and diversity requirements. Like UniCLUB, it uses only bounded full-rank contexts, avoiding the quantitative tail restriction of CLUB/SCLUB and the potentially smaller diversity coefficient $\widetilde{\lambda}_x$ in RCLUMB-type bounds (Li et al., 2025, Table 1). For expected regret, UniCLUB's identification term is $O(ud \log T/(\widetilde{\gamma}^2 \lambda_x))$ (Li et al., 2025, Theorem 1). Our refined calibration removes the explicit d from its $\log T$ coefficient while retaining the $d\sqrt{mT} \log T$ learning term. This improvement is asymptotic: the remainder uC_γ grows with $dK \log(L^2/\lambda_x)$, and at the parameters of every setting in Section 6, the minimum in (4) selects its first term.

Proof sketch. The proof has four steps. (i) With probability at least $1 - \delta$, every personal estimate stays within $\widetilde{\gamma}/4$ of its parameter from certification through T (Lemmas 2 and 6). (ii) The triangle inequality then makes each admission test exact (Lemma 3), so certified neighborhoods form an expanding partial oracle with at most m pooled designs:

$$\mathcal{N}_t[i] = \mathcal{C}_t \cap \mathcal{I}_{j(i)} \qquad \text{for every certified agent } i \text{ and round } t. \tag{6}$$

(iii) Conditioning on the arrival path fixes all certification times, so the at most u certified cluster slices are predictable data streams and self-normalized bounds give pooled confidence (Lemma 4). (iv) Admitted histories enlarge the elliptical potential without adding certified actions (Lemma 5; $\lambda \geq L^2$ as in Abbasi-Yadkori et al. (2011, Lemma 11)). Hence, writing $T_i := \sum_{t=1}^{T} \mathbf{1}\{i_t = i\}$, with probability at least $1 - 2\delta$,

$$\mathrm{Reg}(T) \leq b_{\max} \sum_{i=1}^{u} \min\{T_i, n_{\mathrm{cert}}\} + 2\beta_T \sqrt{2dmT \log\big(1 + TL^2/(d\lambda)\big)}. \tag{7}$$

Setting $\delta = 1/T$ gives (5); details are in Appendix A.2.

Remark 3 (Frequency-agnostic regret)**.** Pre-certification regret is at most $b_{\max} \sum_i \min\{T_i, n_{\mathrm{cert}}\}$ for realized arrival counts T_i: a rarely arriving agent contributes only its own few interactions, while certified agents already learn at the cluster level. The guarantee holds for every oblivious arrival path; conditioning on the path extends it to stochastic arrivals as in Remark 2. In particular, for i.i.d. arrivals with probabilities p_i, $\mathbb{E}[T_i] = p_i T$, and the identification term in the expected-regret bound (26) is $b_{\max} \sum_i \min\{p_i T, n_{\mathrm{cert}}\}$.

Remark 4 (Unknown horizon)**.** For unknown T, we use the doubling idea also employed by Phase-UniCLUB (Li et al., 2025, Algorithm 5). Restarting with $\delta_s = \delta/2^{s+1}$ preserves the learning term and adds at most a logarithmic factor to the identification term, as shown in Appendix A.1.

5 Local Certification Without Uniform Exploration

Local certification requires reliable personal estimates, but need not require uniform exploration. Under Gaussian-perturbed contexts, personal LinUCB collects information while pursuing rewards. For predictable convex scores, the following bound improves the exponentially small all-selector rate derived in Appendix A.2 under Assumption 2 to a loss only logarithmic in K. In this section, the following model replaces Assumption 2 and the bounded-context condition.

Assumption 3 (Gaussian-perturbed contexts)**.** After i_t arrives and before the contexts are revealed, an adversary chooses centers $\boldsymbol{\mu}_{t,a}$ and a common covariance $\boldsymbol{\Sigma}_t$ based on the past, with $\|\boldsymbol{\mu}_{t,a}\| \leq M_0$ and $\sigma_-^2 \boldsymbol{I} \preceq \boldsymbol{\Sigma}_t \preceq \sigma_+^2 \boldsymbol{I}$ for known $M_0 \geq 0$ and $0 < \sigma_- \leq \sigma_+$. Conditional on this information $\mathcal{P}_t$, including i_t,

$$\boldsymbol{x}_{t,a} = \boldsymbol{\mu}_{t,a} + \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{Z}_{t,a}, \qquad \boldsymbol{Z}_{t,a} \overset{\mathrm{iid}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I}). \tag{8}$$

The perturbations are not truncated, so contexts are unbounded.

Theorem 2 (Information under smoothed contexts)**.** *Under Assumption 3, let a_t maximize $s_t(\boldsymbol{x}_{t,a})$, with smallest-index ties, for a finite convex score s_t chosen from $\mathcal{P}_t$ (jointly measurably). With $\alpha_K = [256 \log(2eK)]^{-1}$,*

$$\mathrm{Cov}(\boldsymbol{x}_t \mid \mathcal{P}_t) \succeq \alpha_K \boldsymbol{\Sigma}_t. \tag{9}$$

For any stream selected before its contexts are revealed, with probability at least $1 - \delta$, its n-round Gram matrix $\boldsymbol{S}_n$ satisfies

$$\lambda_{\min}(\boldsymbol{S}_n) \geq \tfrac{1}{4} \alpha_K n \sigma_-^2 \quad \textit{for all attained } n \geq n_{\mathrm{SA}}(\delta), \tag{10}$$

with the burn-in $n_{\mathrm{SA}}(\delta)$ in (41)*.*

LinUCB in place of uniform exploration. The LinUCB index $\boldsymbol{x}^\top \widetilde{\boldsymbol{\theta}} + \beta \|\boldsymbol{x}\|_{\boldsymbol{V}^{-1}}$ is convex, so the information bound holds even for inaccurate estimates. LC-CLUB-L (Local Certification for Clustering of Bandits with LinUCB) replaces Algorithm 1's uniform draw with personal LinUCB and own-count confidence. It retains the latest personal estimates, certification at a fixed own count $n_L = \widetilde{O}(\log(uT)/(\alpha_K \sigma_-^2 \widetilde{\gamma}^2))$ given in (42), and direct-neighbor Full-history pooling, preserving Section 4's conditional argument under stochastic or oblivious arrivals.

Theorem 3 (Regret of LC-CLUB-L)**.** *Under Assumptions 1 and 3 and the noise and arrival models of Section 3, LC-CLUB-L with $\delta = T^{-2}$ and ridge parameter $\lambda = L_T^2$, where L_T is a high-probability bound on all context norms (Appendix A.5), satisfies,*

$$\mathbb{E}[\mathrm{Reg}(T)] = O\left(\frac{u\sqrt{d} \log T}{\sqrt{\alpha_K}\, \sigma_- \widetilde{\gamma}} + d\sqrt{mT} \log T \right). \tag{11}$$

For simplicity, we omit logarithmic factors in K, $1/\sigma_-$, and $1/\widetilde{\gamma}$.

The finite-sample bounds and proof are given by Theorem 5 and Corollary 2 in Appendix A.5. The high-probability bound controls the personal contribution through $\sum_i \sqrt{\min\{T_i, n_L\}}$: rarely observed agents contribute only their reached prefixes, while certified agents learn by cluster. This square-root prefix dependence yields the displayed $1/\widetilde{\gamma}$ factor; pre-certification regret remains.

6 Experiments

6.1 Experiment Setup

We simulate $u = 1000$ agents evenly divided into $m = 10$ clusters, with $d = 10$, $K = 40$ arms per round, and $T = 10^8$ rounds, over five seeds. Within each seed, all methods share the arrival path, contexts, and reward randomness. Writing $\mathbb{R}^{10} = \mathcal{H}_2 \oplus \mathcal{A}_4 \oplus \mathcal{B}_4$, the ten cluster preference vectors coincide on $\mathcal{H}_2 \oplus \mathcal{A}_4$ and differ only on $\mathcal{B}_4$, where they are proportional to the five vertices of a regular simplex and their negatives. For each seed, we apply a single random orthogonal rotation to the $\mathcal{B}_4$ components of all cluster preference vectors, preserving all pairwise distances.

Arm spaces. On the *raw sphere*, arms are uniform on $\mathbb{S}^9$ ($\lambda_x = 1/10$) and rewards are $\boldsymbol{x}^\top \boldsymbol{\theta}_i + \eta$ with $\eta \sim \mathcal{N}(0, 1)$. In the *constructed space*, each arm is

$$\boldsymbol{x} = \left(\tfrac{1}{\sqrt{2}}, \tfrac{\varepsilon}{\sqrt{18}}, \tfrac{2}{3}\boldsymbol{v}\,\mathbf{1}_{\{J=\mathcal{A}\}}, \tfrac{2}{3}\boldsymbol{v}\,\mathbf{1}_{\{J=\mathcal{B}\}} \right)$$

where ε is a random sign, $\boldsymbol{v}$ is uniform on $\mathbb{S}^3$, and J is uniform on $\{\mathcal{A}, \mathcal{B}\}$. Thus, each arm probes the cluster-distinguishing block $\mathcal{B}_4$ with probability $1/2$ ($\lambda_x = 1/18$). Rewards are Bernoulli with mean $\boldsymbol{x}^\top \boldsymbol{\theta}_i \in [0, 1]$. Appendix C gives the exact construction and a third, augmented-sphere space.

Arrivals. Arrival paths are sampled before interaction, independently of contexts, rewards, and policies, from one of two models. *(i) Zipf arrivals* use $p_i \propto r_j^{-\alpha} s_{i|j}^{-\alpha}$, where r_j and $s_{i|j}$ are random cluster and within-cluster ranks; $\alpha = 0$ gives uniform arrivals, while $\alpha = 1$ gives heterogeneous frequencies. *(ii) Sine arrivals* use $p_i(t) = [1 + \sin(2\pi t/10^7 + \phi_i)]/u$ with seed-permuted phases $\phi_i = 2\pi r_i/u$, a period of 10^7 rounds. The main text reports Zipf $\alpha = 0, 1$ and Sine on both arm spaces; Appendix C adds $\alpha = 0.5, 1.5$ and the augmented sphere.

Methods and certification. We compare LC-CLUB-U and LC-CLUB-L with LinUCB (Abbasi-Yadkori et al., 2011), which learns one model per agent, CLUB (Gentile et al., 2014), SCLUB (Li et al., 2019), and UniCLUB (Li et al., 2025); an *Oracle* that runs LinUCB on the true clusters serves as a reference. All methods share one UCB coefficient schedule. Because the constants in our certification threshold are conservative, both LC variants use the empirical radius

$$\rho(n) = \frac{R}{2}\sqrt{\frac{d}{\hat{\lambda}} \cdot \frac{1 + \log(1+n)}{1+n}} + \frac{\lambda S}{\lambda + n\hat{\lambda}}, \qquad \hat{\lambda} = \lambda_x/2, \tag{12}$$

which follows the shape of the personal radius (14), with $\lambda = S = 1$, $R = 1$ for Gaussian and $R = 1/2$ for Bernoulli rewards. An agent is admitted at the first own count with $\rho(n) < \widetilde{\gamma}/4$; LC-CLUB-L evaluates ρ at $n/\log K$.

Metrics. Curves and tables show five-seed means, with standard-error bands on the curves; tables are rounded to four significant digits and report the final cumulative regret R_T, the last-1M increment $\Delta = R(100\text{M}) - R(99\text{M})$, and the final pair F1 against the oracle graph, which links exactly the agents of the same true cluster. Appendix C gives standard errors for all 15 settings.

6.2 Results

Regret. Tables 2 and 3 also report UniCLUB, LC-CLUB-U and the Oracle. UniCLUB explores uniformly for all 10^8 rounds, since its prescribed exploration length exceeds T, and LC-CLUB-U explores uniformly before each agent is certified, so their regret is dominated by exploration. The Oracle knows the true clusters and serves only as a reference. We therefore compare LC-CLUB-L with LinUCB, CLUB, and SCLUB. Figure 1 shows that LC-CLUB-L has the lowest cumulative regret among these methods in all six displayed settings and in every seed; relative to the best of them, it reduces final regret by 2–6% on the raw sphere and by 24–38% in the constructed space.

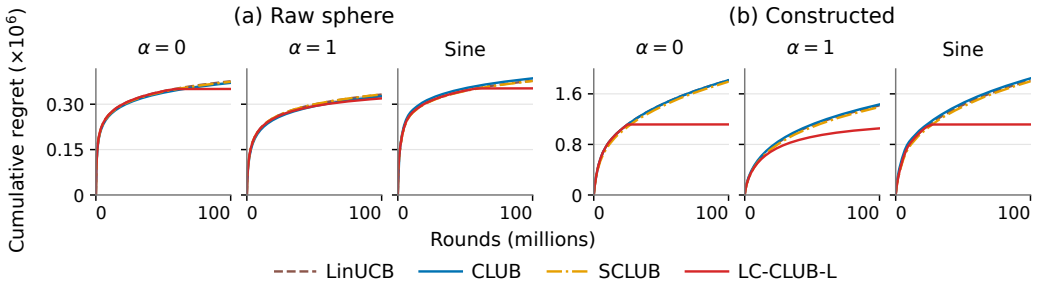

Figure 1: Cumulative regret on the raw sphere (left) and constructed space (right): five-seed means with standard-error bands.

Table 2: Raw unit sphere with Gaussian reward noise $\mathcal{N}(0,1)$: final results at $T = 100$M. Bold marks the best value in each column other than the Oracle; ties are all bold.

Algorithm	Regret R_T ($\times 10^3$)			Last-1M regret Δ			Final pair F1 (%)		
	$\alpha = 0$	$\alpha = 1$	Sine	$\alpha = 0$	$\alpha = 1$	Sine	$\alpha = 0$	$\alpha = 1$	Sine
UniCLUB	60900	60900	60900	0.6089×10^6	0.6086×10^6	0.6087×10^6	**100.0**	18.03	**100.0**
LC-CLUB-U	9553	8884	9552	**3.917**	17140	**4.328**	**100.0**	**76.36**	**100.0**
LinUCB	375.8	332.1	377.7	471.7	475.1	474.6	0	0	0
CLUB	370.5	326.2	385.4	473.6	476.8	474.5	0	0	0
SCLUB	374.1	332.4	376.5	471.7	474.7	475.0	0.008080	0.004040	0.003232
LC-CLUB-L	**350.4**	**319.2**	**352.3**	4.513	**330.9**	4.633	**100.0**	28.69	**100.0**
Oracle	5.702	5.554	5.958	4.485	4.837	4.834	100.0	100.0	100.0

Table 3: Constructed arm space with Bernoulli rewards: final results at $T = 100$M. Bold marks the best value in each column other than the Oracle; ties are all bold.

Algorithm	Regret R_T ($\times 10^3$)			Last-1M regret Δ			Final pair F1 (%)		
	$\alpha = 0$	$\alpha = 1$	Sine	$\alpha = 0$	$\alpha = 1$	Sine	$\alpha = 0$	$\alpha = 1$	Sine
UniCLUB	33370	33370	33370	0.3335×10^6	0.3335×10^6	0.3337×10^6	**100.0**	18.03	**100.0**
LC-CLUB-U	2270	2272	2270	39.89	**115.7**	40.23	**100.0**	**99.66**	**100.0**
LinUCB	1816	1418	1815	6157	4963	6145	0	0	0
CLUB	1816	1433	1845	6139	4982	6135	0	0	0
SCLUB	1793	1392	1799	6440	5021	6220	0.1784	0.2631	0.1462
LC-CLUB-L	**1116**	**1054**	**1116**	**16.03**	1555	**17.05**	**100.0**	57.41	**100.0**
Oracle	76.57	71.77	76.24	255.1	243.8	254.3	100.0	100.0	100.0

Regret growth and graph recovery. Across all settings in Tables 2 and 3, LC-CLUB-L outperforms LinUCB, CLUB, and SCLUB on all three metrics, and by a wide margin in late regret growth and graph recovery: its regret over the last 1M rounds is much smaller, and it recovers a large part of the cluster graph, whereas CLUB and SCLUB end with almost no pooled pairs. LC-CLUB-U explores longer before certification, so its late regret growth and pair F1 are sometimes even better than those of LC-CLUB-L, at the price of a much larger cumulative regret. Neither LC variant ever certifies a cross-cluster pair, and their pair F1 increases with T. For LC-CLUB-L on the raw sphere at $\alpha = 1$, pair F1 rises from 1.3% at 2×10^7 to 12.9% at 6×10^7 and 28.7% at 10^8; under uniform and Sine arrivals, it reaches 100% once all agents have certified. Table 13 in Appendix C reports F1 at ten rounds for every setting.

7 Conclusion

We introduced LC-CLUB, a frequency-agnostic framework for online clustering of bandits in which each agent joins collaboration once its own estimate is certified. Direct-neighbor pooling yields an expanding partial oracle that separates pre-certification interactions from cluster-level learning under stochastic or oblivious arrivals, and under Gaussian-perturbed contexts LinUCB itself supplies the information needed for certification. Appendix B discusses further directions.

AI USE STATEMENT

We used LLM assistants (OpenAI Codex and Claude Code) to help organize the proofs, draft the experiment code, and polish the writing. The research problem, ideas, and results are the authors' own. The authors verified all proofs, code, and text, and take full responsibility for the content of this paper.

ETHICS STATEMENT

This work studies online clustering of linear bandits under heterogeneous agent arrivals. All experiments use synthetic data: agent preferences, contexts, arrivals, and Gaussian or Bernoulli rewards are generated from the distributions specified in Section 6. No human participants, personal data, or real user data are involved.

REPRODUCIBILITY STATEMENT

The model and assumptions are specified in Section 3, LC-CLUB-U in Algorithm 1 (Section 4), and LC-CLUB-L in Section 5. Complete proofs of all theoretical results are given in Appendix A. Section 6 and Appendix C describe the arm spaces, arrival processes, empirical certification radius, baselines, and metrics. The supplementary material contains the code to run all experiments in Section 6 and Appendix C: implementations of all seven methods, configuration files for all 15 settings, and scripts that compute the regret summaries and curves from the resulting trajectories.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Yikun Ban and Jingrui He. Local clustering in contextual multi-armed bandits. In *Proceedings of the Web Conference 2021*, pp. 2335–2346, 2021.

Yikun Ban, Yunzhe Qi, Tianxin Wei, Lihui Liu, and Jingrui He. Meta clustering of neural bandits. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 95–106, 2024.

Hamza Cherkaoui, Merwan Barlier, and Igor Colin. Adaptive sample sharing for multi agent linear bandits. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 10216–10248. PMLR, 2025. URL https://proceedings.mlr.press/v267/cherkaoui25a.html. Earlier version circulated as "Clustered Multi-Agent Linear Bandits", arXiv:2309.08710.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 208–214. JMLR Workshop and Conference Proceedings, 2011.

Claudio Gentile, Shuai Li, and Giovanni Zappella. Online clustering of bandits. In *International conference on machine learning*, pp. 757–765. PMLR, 2014.

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*, pp. 1253–1262. PMLR, 2017.

Yanjun Han, Zhengqing Zhou, Zhengyuan Zhou, Jose H. Blanchet, Peter W. Glynn, and Yinyu Ye. Sequential batch learning in finite-action linear contextual bandits, 2020. URL https://arxiv.org/abs/2004.06321.

Seok-Jin Kim and Min-hwan Oh. Local anti-concentration class: Logarithmic regret for greedy linear contextual bandit. In *Advances in Neural Information Processing Systems*, volume 37, pp. 77525–77592, 2024. doi: 10.52202/079017-2465. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/8dbd2780192078711c0f31e10a819031-Abstract-Conference.html.

Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, 2010.

Shuai Li and Shengyu Zhang. Online clustering of contextual cascading bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11763. URL https://ojs.aaai.org/index.php/AAAI/article/view/11763.

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*, pp. 539–548, 2016.

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

Zhuohua Li, Maoli Liu, Xiangxiang Dai, and John Lui. Demystifying online clustering of bandits: Enhanced exploration under stochastic and smoothed adversarial contexts. *arXiv preprint arXiv:2501.00891*, 2025.

Xutong Liu, Haoru Zhao, Tong Yu, Shuai Li, and John CS Lui. Federated online clustering of bandits. In *Uncertainty in Artificial Intelligence*, pp. 1221–1231. PMLR, 2022.

Xutong Liu, Xiangxiang Dai, Jinhang Zuo, Siwei Wang, Carlee Joe-Wong, John Lui, and Wei Chen. Offline learning for combinatorial multi-armed bandits. *arXiv preprint arXiv:2501.19300*, 2025.

Sentao Miao, Xi Chen, Xiuli Chao, Jiaxi Liu, and Yidong Zhang. Context-based dynamic pricing with online clustering. *Production and Operations Management*, 31(9):3559–3575, 2022.

Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.

Zhiyong Wang, Jize Xie, Xutong Liu, Shuai Li, and John Lui. Online clustering of bandits with misspecified user models. *Advances in Neural Information Processing Systems*, 36:3785–3818, 2023a.

Zhiyong Wang, Jize Xie, Tong Yu, Shuai Li, and John Lui. Online corrupted user detection and regret minimization. *Advances in Neural Information Processing Systems*, 36:33262–33287, 2023b.

Zhiyong Wang, Jiahang Sun, Mingze Kong, Jize Xie, Qinghua Hu, John C. S. Lui, and Zhongxiang Dai. Online clustering of dueling bandits. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 64403–64431. PMLR, 2025. URL https://proceedings.mlr.press/v267/wang25cs.html.

Junda Wu, Canzhe Zhao, Tong Yu, Jingyang Li, and Shuai Li. Clustering of conversational bandits for user preference learning and elicitation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 2129–2139, 2021.

Hantao Yang, Xutong Liu, Zhiyong Wang, Hong Xie, John C. S. Lui, Defu Lian, and Enhong Chen. Federated contextual cascading bandits with asynchronous communication and heterogeneous users. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 20596–20603, 2024. doi: 10.1609/aaai.v38i18.30045. URL https://ojs.aaai.org/index.php/AAAI/article/view/30045.

Liu Yang, Bo Liu, Leyu Lin, Feng Xia, Kai Chen, and Qiang Yang. Exploring clustering of bandits for online recommendation system. In *Proceedings of the 14th ACM Conference on Recommender Systems*, pp. 120–129, 2020.

A PROOFS

Throughout the proofs, we assume $T \geq 2$ and $u \leq T$.

A.1 Theoretical Analysis of LC-CLUB-U

Following the proof order of Li et al. (2025), we first establish personal estimation and design bounds, then identify the certified neighborhoods, control shared confidence, and prove the regret theorem. The first two lemmas use personal observation counts without conditioning on the arrival path; Full-history confidence is then established conditional on that path. Throughout this subsection, β_t is given by (3) and n_{cert} by (4). We first prove persistent accuracy for the simple sufficient threshold (17). Appendix A.2 proves the alternative sufficient threshold and justifies the deterministic choice between them. All subsequent arguments use only the resulting persistent-accuracy event. Figure 2 compares this local regret decomposition with an all-agent recovery argument.

For agent i, index its own observations by n and write

$$\boldsymbol{S}_{i,n} = \sum_{k=1}^{n} \boldsymbol{x}_{i,k}\boldsymbol{x}_{i,k}^{\top}, \qquad \boldsymbol{V}_{i,n} = \lambda \boldsymbol{I} + \boldsymbol{S}_{i,n}, \qquad \widehat{\boldsymbol{\theta}}_{i,n} = \boldsymbol{V}_{i,n}^{-1} \sum_{k=1}^{n} r_{i,k}\boldsymbol{x}_{i,k}.$$

These are the cumulative owner statistics after the n-th own observation, including observations collected after certification. At calendar time t, the algorithm's $(\boldsymbol{S}_{i,t}, \boldsymbol{b}_{i,t})$ and $\widehat{\boldsymbol{\theta}}_{i,t}$ are these same quantities at $n = n_i(t)$. For a certified agent, $\widetilde{\boldsymbol{V}}_{i,t}$ and $\widetilde{\boldsymbol{\theta}}_{i,t}$ denote instead the pooled design and estimate of line 7, formed from the statistics of its closed neighborhood $\mathcal{N}_t[i]$ after round t.

Lemma 1 (Personal estimation error)**.** *Under the bounded-context and chronological noise conditions of Section 3, for every $\delta \in (0,1)$, with probability at least $1-\delta$, simultaneously for every agent i and every own count n reached by T, writing $\mu_{i,n} = \lambda_{\min}(\boldsymbol{V}_{i,n}) \geq \lambda$,*

$$\|\widehat{\boldsymbol{\theta}}_{i,n} - \boldsymbol{\theta}_i\|_2 \leq \rho_{i,n}(\delta) := \frac{R\sqrt{2\log(u/\delta) + \log(\det \boldsymbol{V}_{i,n}/\lambda^d)}}{\sqrt{\mu_{i,n}}} + \frac{\lambda S}{\mu_{i,n}}, \tag{13}$$

$$\leq \frac{R\sqrt{2\log(u/\delta) + d\log(1 + nL^2/(d\lambda))}}{\sqrt{\mu_{i,n}}} + \frac{\lambda S}{\mu_{i,n}}. \tag{14}$$

Proof. We follow the personal-ridge argument of Li et al. (2025, Lemma 1) with general scales R, S, retaining the actual design determinant. The ridge identity gives

$$\widehat{\boldsymbol{\theta}}_{i,n} - \boldsymbol{\theta}_i = \boldsymbol{V}_{i,n}^{-1}\left(\sum_{k=1}^{n} \boldsymbol{x}_{i,k}\eta_{i,k} - \lambda\boldsymbol{\theta}_i\right).$$

Apply Lemma 7 to the chronological feature $\mathbf{1}\{i_t = i\}\boldsymbol{x}_t$, which is $\mathcal{H}_t$-measurable, at level δ/u, and take a union bound over agents. This gives

$$\|\widehat{\boldsymbol{\theta}}_{i,n} - \boldsymbol{\theta}_i\|_{\boldsymbol{V}_{i,n}} \leq R\sqrt{2\log(u/\delta) + \log(\det \boldsymbol{V}_{i,n}/\lambda^d)} + \sqrt{\lambda}S. \tag{15}$$

For the Euclidean bound, apply $\|\boldsymbol{V}_{i,n}^{-1}\boldsymbol{M}\|_2 \leq \|\boldsymbol{M}\|_{\boldsymbol{V}_{i,n}^{-1}}/\sqrt{\lambda_{\min}(\boldsymbol{V}_{i,n})}$ to the noise sum and $\|\lambda\boldsymbol{V}_{i,n}^{-1}\boldsymbol{\theta}_i\|_2 \leq \lambda S/\lambda_{\min}(\boldsymbol{V}_{i,n})$ to the ridge bias separately. This proves (13) with its separate bias term. The trace–determinant inequality $\log(\det \boldsymbol{V}_{i,n}/\lambda^d) \leq d\log(1 + nL^2/(d\lambda))$ proves (14). In particular, later adaptive arm choices are covered by the same event. □

Write $\alpha_t := \beta_t - \sqrt{\lambda}S$ for the noise part of the coefficient in (3). At failure level $\delta/2$, the direct personal radius obeys, for every reached $n \leq T$,

$$\rho_{i,n}(\delta/2) \leq \frac{\alpha_T}{\sqrt{\mu_{i,n}}} + \frac{\lambda S}{\mu_{i,n}} \leq \frac{\beta_T}{\sqrt{\mu_{i,n}}}. \tag{16}$$

The last step uses $\mu_{i,n} \geq \lambda$. The bound holds at every reached own count, including after certification.

Lemma 2 (Uniform-prefix design and persistent accuracy)**.** *Under Assumptions 1 and 2, suppose each agent explores uniformly for its first q own interactions, where*

$$q := \left\lceil \frac{8}{\lambda_x} \max\left\{ L^2 \log\frac{2ud}{\delta}, \frac{4\beta_T^2}{\widetilde{\gamma}^2} \right\} \right\rceil. \tag{17}$$

With probability at least $1-\delta/2$, simultaneously for every agent and every reached own count $n \geq q$,

$$\boldsymbol{V}_{i,n} \succeq \boldsymbol{V}_{i,q} \succeq (\lambda + q\lambda_x/2)\boldsymbol{I}. \tag{18}$$

Consequently, under the noise condition of Section 3, with probability at least $1-\delta$, simultaneously for every agent and every reached count $q \leq n \leq T$,

$$\|\widehat{\boldsymbol{\theta}}_{i,n} - \boldsymbol{\theta}_i\|_2 \leq \rho_{i,n}(\delta/2) \leq \frac{\beta_T}{\sqrt{\lambda + q\lambda_x/2}} < \frac{\widetilde{\gamma}}{4}. \tag{19}$$

Proof. We use the uniform-exploration argument of Li et al. (2025, Lemma 2) at the fixed local count q. By freshness, generate in advance, for each agent i, q independent potential candidate sets and independent uniform arm choices. Let $\widetilde{\boldsymbol{x}}_{i,1}, \ldots, \widetilde{\boldsymbol{x}}_{i,q}$ be the selected contexts. Each has law $\mathcal{P}_x$; reached pre-certification observations reveal successive draws, and unused draws remain latent. The matrices $\widetilde{\boldsymbol{x}}_{i,k}\widetilde{\boldsymbol{x}}_{i,k}^\top$ are independent, bounded by $L^2\boldsymbol{I}$, and have expectation $\Sigma_x \succeq \lambda_x \boldsymbol{I}$. Matrix Chernoff (Tropp, 2012, Corollary 5.2) gives

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{k=1}^{q} \widetilde{\boldsymbol{x}}_{i,k}\widetilde{\boldsymbol{x}}_{i,k}^\top\right) < q\lambda_x/2\right) \leq d\exp\left(-\frac{q\lambda_x}{8L^2}\right) \leq \frac{\delta}{2u},$$

where the last step uses the first term in (17). A union bound over agents gives the initial-design event with failure at most $\delta/2$. Whenever i reaches q, its actual prefix equals this potential prefix. Later updates are positive semidefinite, proving (18) without conditioning on reaching q. The argument uses personal counts without conditioning on the arrival path.

Intersect this event with Lemma 1 at level $\delta/2$ and call the resulting event $\mathcal{E}_{\mathrm{loc}}$; it has probability at least $1-\delta$. On $\mathcal{E}_{\mathrm{loc}}$, for $q \leq n \leq T$, (16) and (18) give $\|\widehat{\boldsymbol{\theta}}_{i,n} - \boldsymbol{\theta}_i\|_2 \leq \rho_{i,n}(\delta/2) \leq \beta_T/\sqrt{\lambda + q\lambda_x/2}$. The second term in (17) gives $\lambda + q\lambda_x/2 > 16\beta_T^2/\widetilde{\gamma}^2$, hence the strict quarter-gap bound. Later adaptive actions only add positive semidefinite terms; the proof does not claim that they keep increasing the smallest eigenvalue. □

The quarter-gap separation argument of Li et al. (2025, Lemma 3) now applies at each admission, using the latest personal estimates of the new agent and all agents certified earlier.

Lemma 3 (Exact certified neighborhoods)**.** *Let $G^\star = ([u], E^\star)$ be the oracle graph with $E^\star = \{\{i,\ell\} : i \neq \ell,\ j(i) = j(\ell)\}$. Under Assumptions 1 and 2 and the noise model of Section 3, use the threshold* (4)*. With probability at least $1-\delta$, simultaneously for all rounds $t \leq T$,*

$$G_t = G^\star[\mathcal{C}_t], \tag{20}$$

where the right-hand side is the induced subgraph on certified agents. Consequently, for every $i \in \mathcal{C}_t$,

$$\mathcal{N}_t[i] = \mathcal{C}_t \cap \mathcal{I}_{j(i)}. \tag{21}$$

Proof. Work on $\mathcal{E}_{\mathrm{loc}}$ from Lemma 2 or 6, according to (4). Consider the admission of agent i at round t and any previously certified agent ℓ. If $j(i) = j(\ell)$, the common parameter and triangle inequality give

$$\begin{aligned}\|\widehat{\boldsymbol{\theta}}_{i,t} - \widehat{\boldsymbol{\theta}}_{\ell,t}\|_2 &\leq \|\widehat{\boldsymbol{\theta}}_{i,t} - \boldsymbol{\theta}_i\|_2 + \|\widehat{\boldsymbol{\theta}}_{\ell,t} - \boldsymbol{\theta}_\ell\|_2 \\ &< \widetilde{\gamma}/2.\end{aligned}$$

If $j(i) \neq j(\ell)$, Assumption 1 instead gives

$$\begin{aligned}\|\widehat{\boldsymbol{\theta}}_{i,t} - \widehat{\boldsymbol{\theta}}_{\ell,t}\|_2 &\geq \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_\ell\|_2 - \|\widehat{\boldsymbol{\theta}}_{i,t} - \boldsymbol{\theta}_i\|_2 - \|\widehat{\boldsymbol{\theta}}_{\ell,t} - \boldsymbol{\theta}_\ell\|_2 \\ &> \gamma - \widetilde{\gamma}/2 \geq \widetilde{\gamma}/2.\end{aligned}$$

Thus the edge test adds exactly the edges from i to its previously certified cluster mates. Starting from the empty graph, induction over admissions proves $G_t = G^\star[\mathcal{C}_t]$. Every represented true cluster is a clique, whence $\mathcal{N}_t[i] = \mathcal{C}_t \cap \mathcal{I}_{j(i)}$. These are deterministic consequences of $\mathcal{E}_{\mathrm{loc}}$ and introduce no additional failure event. No conclusion about graph correctness outside that event is needed. □

Lemma 4 (Full-history confidence conditional on the arrival path)**.** *Under either arrival model in Remark 2, condition on the arrival path (almost surely in the stochastic case). There is an event $\mathcal{E}_{\mathrm{sh}}$ of conditional probability at least $1-\delta$ such that, on $\mathcal{E}_{\mathrm{loc}} \cap \mathcal{E}_{\mathrm{sh}}$, every certified shared estimate at $t \le T$ satisfies*

$$\left|\boldsymbol{x}^\top(\widetilde{\boldsymbol{\theta}}_{i,t-1} - \boldsymbol{\theta}_i)\right| \le \beta_t \|\boldsymbol{x}\|_{\widetilde{\boldsymbol{V}}_{i,t-1}^{-1}} \quad (i \in \mathcal{C}_{t-1},\ \boldsymbol{x} \in \mathbb{R}^d). \tag{22}$$

Here β_t is the common coefficient in (3)*.*

Proof. Let σ_i be agent i's certification round, with $\sigma_i = \infty$ if it never reaches n_{cert} interactions. For each true cluster c, order the agents with finite certification rounds as $i_{c,1}, \dots, i_{c,h_c}$ and put $A_{c,k} = \{i_{c,1}, \dots, i_{c,k}\}$. The conditioned path fixes these sets and every σ_i. For $k \in [h_c]$, use the calendar-time mask

$$B_{c,k,s} = \mathbf{1}\{i_s \in A_{c,k}\}.$$

Define the corresponding chronological design and noise sums by

$$\boldsymbol{V}_{c,k}(t) = \lambda \boldsymbol{I} + \sum_{s \le t} B_{c,k,s} \boldsymbol{x}_s \boldsymbol{x}_s^\top, \qquad \boldsymbol{M}_{c,k}(t) = \sum_{s \le t} B_{c,k,s} \boldsymbol{x}_s \eta_s.$$

Conditional on the path, each masked feature is known before its reward. Lemma 7, with the path included in its fields, at level δ/u and a union bound over $\sum_c h_c \le u$ masks give an event $\mathcal{E}_{\mathrm{sh}}$ on which

$$\|\boldsymbol{M}_{c,k}(t)\|_{\boldsymbol{V}_{c,k}(t)^{-1}} \le R\sqrt{\log \frac{\det \boldsymbol{V}_{c,k}(t)}{\det(\lambda \boldsymbol{I})} + 2\log\frac{u}{\delta}} \quad \text{for every } c, k, t.$$

This event is established before intersecting with graph correctness. The masks are only proof objects; the algorithm does not observe future arrivals, and all noise is analyzed in chronological order.

Now take a shared action of cluster c at round t and let k be its number of certified agents before that action. The prefix of $B_{c,k,s}$ through $t-1$ contains exactly the complete personal histories of the certified owners in $A_{c,k}$, each once. On $\mathcal{E}_{\mathrm{loc}}$, the neighborhood identity therefore gives

$$\widetilde{\boldsymbol{V}}_{i_t,t-1} = \boldsymbol{V}_{c,k}(t-1), \qquad \widetilde{\boldsymbol{\theta}}_{i_t,t-1} - \boldsymbol{\theta}^c = \boldsymbol{V}_{c,k}(t-1)^{-1}\big(\boldsymbol{M}_{c,k}(t-1) - \lambda \boldsymbol{\theta}^c\big).$$

Cauchy–Schwarz and $\boldsymbol{V}_{c,k}(t-1) \succeq \lambda \boldsymbol{I}$ imply

$$\left|\boldsymbol{x}^\top(\widetilde{\boldsymbol{\theta}}_{i_t,t-1} - \boldsymbol{\theta}^c)\right| \le \|\boldsymbol{x}\|_{\boldsymbol{V}_{c,k}(t-1)^{-1}} \left(\|\boldsymbol{M}_{c,k}(t-1)\|_{\boldsymbol{V}_{c,k}(t-1)^{-1}} + \sqrt{\lambda} S\right).$$

The pool contains at most $t-1$ distinct bounded records. Hence

$$\log \frac{\det \boldsymbol{V}_{c,k}(t-1)}{\det(\lambda \boldsymbol{I})} \le d \log\left(1 + \frac{(t-1)L^2}{d\lambda}\right) \le d \log\left(1 + \frac{tL^2}{d\lambda}\right).$$

The coefficient in (3) upper-bounds the resulting radius, proving (22). The same argument applies to any certified agent's pool, whether or not it acts at t. □

For Full-history pooling, let $\mathcal{T}_c = \{t \le T : i_t \in \mathcal{I}_c,\ i_t \in \mathcal{C}_{t-1}\}$ be the certified action rounds of cluster c, and write $Q_c = |\mathcal{T}_c| = \sum_{i \in \mathcal{I}_c} (T_i - n_{\mathrm{cert}})_+$. On $\mathcal{E}_{\mathrm{loc}}$, all certified agents of cluster c have the same pool; denote its design after round t by $\boldsymbol{V}_c(t)$, using $\lambda \boldsymbol{I}$ before the cluster is represented.

Lemma 5 (Elliptical potential over certified rounds)**.** *On $\mathcal{E}_{\mathrm{loc}}$ and with $\lambda \ge L^2$,*

$$\sum_{t \in \mathcal{T}_c} \|\boldsymbol{x}_t\|^2_{\boldsymbol{V}_c(t-1)^{-1}} \le 2 \log \frac{\det \boldsymbol{V}_c(T)}{\det(\lambda \boldsymbol{I})} \le 2d \log\left(1 + \frac{TL^2}{d\lambda}\right). \tag{23}$$

Proof. Apply Lemma 8 to $\boldsymbol{W}_t = \boldsymbol{V}_c(t)$ with $\mathcal{R} = \mathcal{T}_c$. On a certified round the acting agent is already certified, so $\boldsymbol{V}_c(t) = \boldsymbol{V}_c(t-1) + \boldsymbol{x}_t \boldsymbol{x}_t^\top$. At an admission, the new owner's stored history joins the neighbor sum as a positive-semidefinite block; all other rounds leave $\boldsymbol{V}_c$ unchanged. Since $\lambda \ge L^2$, the minimum in (39) can be removed, and $\boldsymbol{V}_c(T)$ contains at most T distinct records. □

Theorem 4 (Finite-sample regret)**.** *Under Assumptions 1 and 2, the reward-noise model of Section 3, and stochastic or oblivious arrivals (Remark 2), the following bounds hold. For any $T \geq 1$ and $0 < \delta \leq 1/2$, Algorithm 1 with $\lambda \geq L^2$, β_t from* (3)*, and n_{cert} from* (4) *satisfies, conditionally on almost every arrival path and hence unconditionally, with probability at least $1 - 2\delta$,*

$$\mathrm{Reg}(T) \leq b_{\max} \sum_{i=1}^{u} \min\{T_i, n_{\mathrm{cert}}\} + 2\beta_T \sqrt{2dmT \log(1 + TL^2/(d\lambda))}, \tag{24}$$

where T_i is the number of arrivals of agent i through T. Moreover,

$$\mathbb{E}[\mathrm{Reg}(T)] \leq b_{\max} \sum_{i} \min\{\mathbb{E}[T_i], n_{\mathrm{cert}}\} + 2\beta_T \sqrt{2dmT \log(1 + TL^2/(d\lambda))} + 2b_{\max}\delta T. \tag{25}$$

For i.i.d. arrivals with probabilities p_i, $\mathbb{E}[T_i] = p_i T$. Neither bound requires a positive minimum arrival frequency or within-cluster frequency alignment.

Proof. Condition on the arrival path. By Remark 2, the conditional context and noise laws remain valid. The personal-accuracy argument selected by (4) (Lemma 2 or 6) and Lemma 4 give $\mathcal{E}_{\mathrm{loc}} \cap \mathcal{E}_{\mathrm{sh}}$ conditional probability at least $1 - 2\delta$. Work on this event. Fix $t \in \mathcal{T}_c$, and let $\boldsymbol{x}_t^*$ maximize the true reward over the offered arms. Write $\boldsymbol{V} = \boldsymbol{V}_c(t-1)$ and $\widetilde{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}_{i_t,t-1}$. Confidence and the UCB maximization rule give

$$\begin{aligned}
\Delta_t &= (\boldsymbol{x}_t^*)^\top \boldsymbol{\theta}^c - \boldsymbol{x}_t^\top \boldsymbol{\theta}^c \\
&\leq (\boldsymbol{x}_t^*)^\top \widetilde{\boldsymbol{\theta}} + \beta_t \|\boldsymbol{x}_t^*\|_{\boldsymbol{V}^{-1}} - \boldsymbol{x}_t^\top \boldsymbol{\theta}^c \\
&\leq \boldsymbol{x}_t^\top \widetilde{\boldsymbol{\theta}} + \beta_t \|\boldsymbol{x}_t\|_{\boldsymbol{V}^{-1}} - \boldsymbol{x}_t^\top \boldsymbol{\theta}^c \\
&\leq 2\beta_t \|\boldsymbol{x}_t\|_{\boldsymbol{V}^{-1}}.
\end{aligned}$$

Using Cauchy–Schwarz within and across clusters, and then Lemma 5, we obtain

$$\begin{aligned}
\sum_{c=1}^{m} \sum_{t \in \mathcal{T}_c} \Delta_t &\leq 2\beta_T \sum_{c=1}^{m} \sqrt{Q_c \sum_{t \in \mathcal{T}_c} \|\boldsymbol{x}_t\|^2_{\boldsymbol{V}_c(t-1)^{-1}}} \\
&\leq 2\beta_T \sqrt{2d \log\left(1 + \frac{TL^2}{d\lambda}\right)} \sum_{c=1}^{m} \sqrt{Q_c} \\
&\leq 2\beta_T \sqrt{2dm \left(\sum_c Q_c\right) \log\left(1 + \frac{TL^2}{d\lambda}\right)}.
\end{aligned}$$

Since $\sum_c Q_c \leq T$, this is the claimed learning term. Each agent has exactly $\min\{T_i, n_{\mathrm{cert}}\}$ pre-certification actions, including its triggering action, each incurring regret at most $b_{\max}$. Adding the regret bounds for these two disjoint sets of rounds proves (24) conditionally on the path. No independence between the two confidence events is needed. For stochastic arrivals, averaging over the arrival path proves the unconditional high-probability claim; for oblivious arrivals, the path is fixed.

For the expectation, use $\mathrm{Reg}(T) \leq b_{\max} T$ on failure, first conditionally and then averaged over paths:

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}(T)] &\leq b_{\max} \sum_i \mathbb{E}[\min\{T_i, n_{\mathrm{cert}}\}] + 2\beta_T \sqrt{2dmT \log(1 + TL^2/(d\lambda))} + 2b_{\max}\delta T \\
&\leq b_{\max} \sum_i \min\{\mathbb{E}[T_i], n_{\mathrm{cert}}\} + 2\beta_T \sqrt{2dmT \log(1 + TL^2/(d\lambda))} + 2b_{\max}\delta T,
\end{aligned}$$

where concavity of $x \mapsto \min\{x, n_{\mathrm{cert}}\}$ gives the second line. Choosing $\delta = 1/T$ in the certification threshold and confidence radius makes the failure contribution $2b_{\max}$. For i.i.d. arrivals, $\mathbb{E}[T_i] = p_i T$, yielding

$$\mathbb{E}[\mathrm{Reg}(T)] \leq b_{\max} \sum_i \min\{p_i T, n_{\mathrm{cert}}\} + 2\beta_T \sqrt{2dmT \log(1 + TL^2/(d\lambda))} + 2b_{\max}. \tag{26}$$

This gives the i.i.d. specialization in Remark 3. For oblivious arrivals, the proof applies to each fixed path; no event common to all counterfactual paths is asserted. □

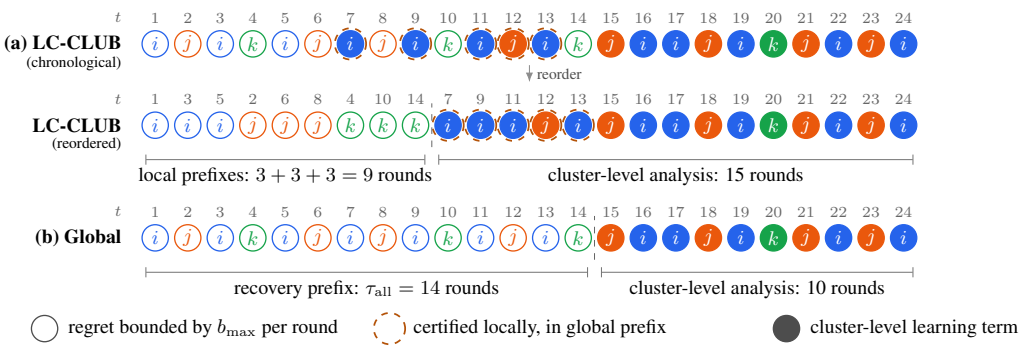

Figure 2: An example of our proof technique: local certification versus global recovery on one arrival path, with $u = 3$ agents in one cluster whose realized arrival frequencies over $T = 24$ rounds are $1/2$, $1/3$ and $1/6$, so $(T_i, T_j, T_k) = (12, 8, 4)$, and $n_{\text{cert}} = 3$. Certification occurs at $\sigma_i = 5$, $\sigma_j = 8$ and $\sigma_k = 14$. (a) The first LC-CLUB row preserves chronological order. The second groups nine pre-certification actions by agent, followed by 15 certified actions in their original relative order. Dashed circles identify the same five early certified actions in both LC-CLUB rows; these actions remain inside the global recovery prefix. (b) Global recovery splits the chronological path at $\tau_{\text{all}} = 14$, leaving ten rounds for cluster-level analysis. Numbers above the nodes are original rounds; reordering is only for the regret decomposition. Actions and rewards are not shown.

Local certification versus global recovery. A global-recovery proof first establishes correct clustering for all agents, then identifies each learned cluster with an entire true cluster and applies clusterwise confidence and potential bounds to the remaining rounds. UniCLUB follows this order after its calibrated exploration length T_0 (Li et al., 2025, Lemma 3 and proof of Theorem 1). Figure 2 gives an example of our proof technique, contrasting this common recovery-time decomposition with LC-CLUB's growing certified cluster slices. The comparison uses the same local threshold to isolate the timing distinction; UniCLUB's own T_0 and uniform-arrival assumption remain those of the cited theorem.

Recall that σ_i is agent i's certification round, with $\sigma_i = \infty$ if it does not reach n_{cert} observations by T. The corresponding all-agent time is $\tau_{\text{all}} := \max_{i \in [u]} \sigma_i$. A decomposition that uses complete-cluster recovery at this time bounds the whole prefix through $\min\{T, \tau_{\text{all}}\}$ by $b_{\max}$ per round. LC-CLUB instead applies this bound only to each agent's own pre-certification actions, including its trigger. Thus the two prefix terms satisfy

$$\underbrace{b_{\max} \sum_{i=1}^{u} \min\{T_i, n_{\text{cert}}\}}_{\text{local pre-certification term}} \le \underbrace{b_{\max} \min\{T, \tau_{\text{all}}\}}_{\text{global recovery prefix}}. \tag{27}$$

Every local pre-certification action occurs within the global prefix, proving the inequality. This compares the rounds bounded by $b_{\max}$, not the realized regrets of two policies. If one agent never reaches the threshold, the global prefix extends through T, while the local proof still controls certified agents at the cluster level.

For the local proof, fix the arrival path and work on $\mathcal{E}_{\text{loc}} \cap \mathcal{E}_{\text{sh}}$. Lemma 3 identifies every certified neighborhood with the currently certified slice of its true cluster, giving one shared design $\boldsymbol{V}_c$ per represented cluster even before τ_{all}. In the figure, i uses its certified pool at $t = 7$, while k certifies only at $t = 14$. The five rounds $7, 9, 11, 12, 13$ therefore enter the local learning term instead of the global recovery prefix. A certified action adds $\boldsymbol{x}_t \boldsymbol{x}_t^\top$ to the shared design; at σ_i, the owner's earlier records enter as

$$H_i := \sum_{s \le \sigma_i :\, i_s = i} \boldsymbol{x}_s \boldsymbol{x}_s^\top \succeq 0.$$

These earlier actions already belong to the local prefix term. Inserting their records adds no new regret term and can only increase the log-determinant potential. Lemma 5 controls all Q_c certified actions, and $\sum_c \sqrt{Q_c} \le \sqrt{mT}$ yields the cluster-level learning term without waiting for population-wide recovery.

The reordering is algebraic. Lemma 4 establishes shared confidence separately: conditioning on the arrival path fixes the owner sets, and every masked noise sum retains its original calendar order before intersecting with $\mathcal{E}_{\mathrm{loc}}$.

Corollary 1 (Coarser rate)**.** *Under the assumptions of Theorem 4, let $\delta = 1/T$. For bounded L, R, S, λ,*

$$\mathbb{E}[\mathrm{Reg}(T)] = O\left(\frac{ud}{\widetilde{\gamma}^2\lambda_x}\log T + d\sqrt{mT}\log T\right). \tag{28}$$

Proof. Take $\delta = 1/T$ in Theorem 4, whose failure contribution is then $2b_{\max}$. With bounded L, R, S, λ, we have $\beta_T^2 = O(d\log T)$ and $\log(2ud/\delta) = O(d\log T)$. Since $\widetilde{\gamma} \le 2S$ and $\widetilde{\gamma}^2\lambda_x \le 4S^2L^2$, the simple threshold (17) is $O(d\log T/(\widetilde{\gamma}^2\lambda_x))$. Since (4) does not exceed that threshold, $b_{\max}\sum_i \min\{\mathbb{E}[T_i], n_{\mathrm{cert}}\} \le b_{\max}un_{\mathrm{cert}} = O(ud\log T/(\widetilde{\gamma}^2\lambda_x))$. The learning term is $2\beta_T\sqrt{2dmT\log(1+TL^2/(d\lambda))} = O(d\sqrt{mT}\log T)$. Adding the three contributions proves (28). □

Unknown horizon. To establish Remark 4, fix $0 < \delta \le 1/2$ and restart Algorithm 1 in stages of length $H_s = 2^s$, with confidence input $\delta_s = \delta/2^{s+1}$, resetting the graph and all statistics. Recompute the threshold and confidence coefficients using H_s as the prescribed horizon. Up to time T, the entered stages have indices $0 \le s \le J := \lfloor \log_2 T \rfloor$; the final stage may be truncated. Apply Theorem 4 within each stage. Since $H_s \le T$ and $\sum_{s=0}^{J}\sqrt{H_s} = O(\sqrt{T})$, the learning terms sum to $O(d\sqrt{mT}\log T)$ for fixed δ. Summing (34) gives an identification term $O(u\log^2 T/(\widetilde{\gamma}^2\lambda_x) + uC_\gamma\log T)$. The expected regret from confidence failures is at most $\sum_{s=0}^{J} 2b_{\max}\delta_s H_s = b_{\max}\delta(J+1) = O(\log T)$. For the fixed problem parameters in Theorem 1, this preserves the order of the learning term and enlarges the identification term by at most a logarithmic factor as $T \to \infty$.

A.2 Refined Certification Threshold

In Lemma 2, the confidence numerator of a personal estimate is evaluated at the horizon, which places $d\log T$ in (17). Here we keep the personal-count logarithm $d\log(1+nL^2/(d\lambda))$ until the accuracy condition is solved. Because latest personal estimates are compared after certification, this requires information about post-certification actions, which pooled LinUCB chooses. Assumption 2 alone gives a weak information floor that holds for every arm-selection rule. Define

$$\begin{aligned} \nu &:= L^2(\lambda_x/L^2)^K, \qquad D_\gamma := \frac{128R^2}{\widetilde{\gamma}^2}\left[2\log\frac{2u}{\delta} + d\log\left(1 + \frac{512R^2L^2}{\widetilde{\gamma}^2\lambda\nu}\right)\right] + \frac{8\lambda S}{\widetilde{\gamma}}, \\ n_{\mathrm{ref}} &:= \left\lceil \max\left\{\frac{8L^2}{\lambda_x}\log\frac{4ud}{\delta},\ \frac{4D_\gamma}{\lambda_x}\right\}\right\rceil. \end{aligned} \tag{29}$$

Lemma 6 (Refined persistent accuracy)**.** *Under Assumptions 1 and 2, run Algorithm 1 with $n_{\mathrm{cert}} = q := n_{\mathrm{ref}}$, so that each agent explores uniformly for its first q own interactions. With probability at least $1-\delta/2$, simultaneously for every agent and every reached own count $q+s$,*

$$\lambda_{\min}(\boldsymbol{V}_{i,q+s}) \ge D_s := \lambda + q\lambda_x/4 + \nu s/2. \tag{30}$$

Define, for $D > 0$,

$$g(D) := R\sqrt{\frac{2\log(2u/\delta) + d\log(1+4L^2D/(d\lambda\nu))}{D}} + \frac{\lambda S}{D}. \tag{31}$$

Consequently, under the noise condition of Section 3, with probability at least $1-\delta$, simultaneously for every agent and every reached count $q \le n \le T$, $\|\widehat{\boldsymbol{\theta}}_{i,n} - \boldsymbol{\theta}_i\|_2 \le \rho_{i,n}(\delta/2) \le g(\lambda + q\lambda_x/4) < \widetilde{\gamma}/4$.

Proof. The prefix argument in the proof of Lemma 2, now at level $\delta/(4u)$ per agent and using the first term in (29), gives $\lambda_{\min}(\boldsymbol{V}_{i,q}) \ge \lambda + q\lambda_x/2$ for every agent that reaches q, with failure at most $\delta/4$.

For subsequent information, fix a unit vector $\boldsymbol{v}$ and set $Z_a = (\boldsymbol{v}^\top \boldsymbol{x}_{t,a})^2/L^2 \in [0,1]$. Every selected arm satisfies $Z_{a_t} \geq \min_a Z_a \geq \prod_a Z_a$. Conditional independence in Assumption 2 therefore gives

$$\mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^\top \mid \mathcal{F}_{t-1}, i_t] \succeq \nu \boldsymbol{I}, \qquad \nu = L^2(\lambda_x/L^2)^K \leq \lambda_x, \tag{32}$$

for every adaptive selector and without any graph-correctness event. The exponent cannot be improved at this generality: if $\boldsymbol{x}$ equals $\boldsymbol{e}_1$ or $\boldsymbol{e}_2$ with probability $1/2$ each, a selector that prefers $\boldsymbol{e}_2$ gives $\mathbb{E}[(\boldsymbol{e}_1^\top \boldsymbol{x}_t)^2] = 2^{-K} = (\lambda_x/L^2)^K$. Apply Lemma 9 with $B_t = \mathbf{1}\{i_t = i,\ n_i(t-1) \geq q\}$, which is known before the current contexts, $\boldsymbol{Y}_t = \boldsymbol{x}_t \boldsymbol{x}_t^\top$, $H = L^2$, $a = \nu$, and failure level $\delta/(4u)$. Since $1 - e^{-1} > 1/2$, a union bound gives, with failure at most $\delta/4$, for every reached post-certification count s,

$$\lambda_{\min}\left(\sum_{k=q+1}^{q+s} \boldsymbol{x}_{i,k}\boldsymbol{x}_{i,k}^\top\right) \geq \nu s/2 - L^2 \log(4ud/\delta). \tag{33}$$

The process pauses between included arrivals, and the argument does not condition on future arrivals. The first term in (29) implies $L^2 \log(4ud/\delta) \leq q\lambda_x/8$. Weyl's inequality applied to the prefix and post-certification designs proves (30), with total design failure at most $\delta/2$.

Intersect the design events with Lemma 1 at level $\delta/2$, and call the result $\mathcal{E}_{\mathrm{loc}}$ as before. Write $n = q + s$. Since $\nu \leq \lambda_x$, $D_s \geq \nu n/4$, so (14) and (30) give $\rho_{i,q+s}(\delta/2) \leq g(D_s)$. Apply Lemma 10 with $A = 2\log(2u/\delta)$ and $c = 4L^2/(d\lambda\nu)$. Its logarithmic factor has $128R^2dc/\widetilde{\gamma}^2 = 512R^2L^2/(\widetilde{\gamma}^2\lambda\nu)$, so its threshold is exactly D_γ. The second term in (29) gives $D_0 > D_\gamma$, hence $g(D_s) \leq g(D_0) < \widetilde{\gamma}/4$. Only this deterministic envelope is monotone; the realized estimation error need not be. The total failure probability is at most $\delta/2 + \delta/2 = \delta$. Later adaptive actions are not claimed to retain the uniform-exploration rate $\lambda_x/2$; their rate is only $\nu/2$. □

Using the smaller threshold. Both (17) and (29) are deterministic functions of the inputs, fixed before interaction. The main-text choice (4) takes their minimum. Whichever value attains it, Lemma 2 or Lemma 6 gives an event $\mathcal{E}_{\mathrm{loc}}$ of probability at least $1 - \delta$ on which every personal estimate is strictly within $\widetilde{\gamma}/4$ of its parameter from certification through T. The remaining proofs in Appendix A.1 use only this conclusion, so Lemmas 3 to 5 and Theorem 4 hold verbatim with (4).

Theorem 1 (Regret of LC-CLUB-U)**.** *Under Assumptions 1–2 and the noise and arrival models of Section 3, LC-CLUB-U with $\lambda \geq L^2$ and $\delta = 1/T$ satisfies,*

$$\mathbb{E}[\mathrm{Reg}(T)] = O\left(\frac{u}{\widetilde{\gamma}^2\lambda_x}\log T + d\sqrt{mT}\log T\right). \tag{5}$$

Proof. Take $\delta = 1/T$ in Theorem 4 with (4), so that $n_{\mathrm{cert}} \leq n_{\mathrm{ref}}$. Expanding D_γ and bounding the maximum by the sum gives

$$n_{\mathrm{cert}} \leq \left(\frac{8L^2}{\lambda_x} + \frac{1024R^2}{\widetilde{\gamma}^2\lambda_x}\right)\log\frac{2u}{\delta} + C_\gamma, \tag{34}$$

where

$$C_\gamma = 1 + \frac{8L^2}{\lambda_x}\log(2d) + \frac{32\lambda S}{\widetilde{\gamma}\lambda_x} + \frac{512R^2d}{\widetilde{\gamma}^2\lambda_x}\log\left(1 + \frac{512R^2L^2}{\widetilde{\gamma}^2\lambda\nu}\right). \tag{35}$$

This remainder does not depend on u, δ, or T. Since $\widetilde{\gamma} \leq 2S$ and $\log(2uT) \leq 3\log T$, setting $\delta = 1/T$ yields

$$n_{\mathrm{cert}} \leq C_\gamma + O\left(\frac{\log T}{\widetilde{\gamma}^2\lambda_x}\right). \tag{36}$$

For these u and $T \geq e^{\lambda S^2}$, the bounds $\lambda \geq L^2$ and $\sqrt{\lambda}S \leq \sqrt{\log T}$ give $\beta_T \leq (2\sqrt{2}R+1)\sqrt{d\log T}$. Substituting into (24) gives, with probability at least $1 - 2/T$, a bound of order $u\log T/(\widetilde{\gamma}^2\lambda_x) + uC_\gamma + d\sqrt{mT}\log T$. Using (25) adds only the $2b_{\max}$ failure term, so

$$\mathbb{E}[\mathrm{Reg}(T)] = O\left(\frac{u}{\widetilde{\gamma}^2\lambda_x}\log T + uC_\gamma + d\sqrt{mT}\log T\right). \tag{37}$$

The constants hidden in (36) and (37) then depend only on L, R, S; the dependence on d, K, λ, λ_x beyond the displayed factors is confined to C_γ and the condition $T \geq e^{\lambda S^2}$. Set $T_0 :=$

$\max\{2, e^{\lambda S^2}, \exp(\widetilde{\gamma}^2 \lambda_x C_\gamma)\}$, which depends only on $d, K, L, R, S, \lambda, \lambda_x, \widetilde{\gamma}$. For $T \geq T_0$ we have $uC_\gamma \leq u \log T/(\widetilde{\gamma}^2 \lambda_x)$, uniformly in u. Absorbing this term proves (5), and the corresponding high-probability bound. Because (4) never exceeds (17), bound (28) also holds. □

The improvement comes from solving the accuracy condition in the personal count, not from discarding dimension-dependent terms. In particular, $\log(1/\nu) = K \log(L^2/\lambda_x) - \log L^2$ enters C_γ, and $d\lambda_x \leq L^2$. Hence the full remainder uC_γ must be retained when comparing finite horizons or when d or K grows with T. At the parameters of Section 6, for instance, n_{ref} is larger than (17), and (4) coincides with (17).

A.3 Technical Inequalities

We collect the standard concentration and potential tools used in the proofs, together with the conditional matrix bound used for post-certification information growth and the Gaussian extension.

Lemma 7 (Self-normalized concentration)**.** *Let $\mathcal{F}_{t-1} \subseteq \mathcal{G}_t \subseteq \mathcal{F}_t$ be interleaved fields such that η_t is $\mathcal{F}_t$-measurable and $\mathbb{E}[e^{z\eta_t} \mid \mathcal{G}_t] \leq e^{z^2R^2/2}$ for all $z \in \mathbb{R}$. In Section 3, take $\mathcal{G}_t = \mathcal{H}_t$. Under stochastic or oblivious arrivals (Remark 2), these fields may also include the complete arrival path, since conditioning on it preserves the noise bound. If $\boldsymbol{z}_t \in \mathbb{R}^d$ is $\mathcal{G}_t$-measurable, put*

$$\boldsymbol{W}_t = \lambda \boldsymbol{I} + \sum_{s \leq t} \boldsymbol{z}_s \boldsymbol{z}_s^\top, \qquad \boldsymbol{M}_t = \sum_{s \leq t} \boldsymbol{z}_s \eta_s, \qquad \lambda > 0.$$

For every $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\|\boldsymbol{M}_t\|_{\boldsymbol{W}_t^{-1}} \leq R\sqrt{\log \frac{\det \boldsymbol{W}_t}{\det(\lambda \boldsymbol{I})} + 2\log \frac{1}{\delta}} \quad \textit{simultaneously for all } t. \tag{38}$$

Proof. This is Abbasi-Yadkori et al. (2011, Theorem 1), applied to the interleaved filtration: the feature is $\mathcal{G}_t$-measurable before the conditionally R-sub-Gaussian reward noise is revealed. □

Lemma 8 (Elliptical potential with additional matrix increments)**.** *Let $\boldsymbol{W}_0 = \lambda \boldsymbol{I}$, $\lambda > 0$, and suppose a sequence of design matrices never decreases. For a designated set of rounds $\mathcal{R}$, $\boldsymbol{W}_t \succeq \boldsymbol{W}_{t-1} + \boldsymbol{x}_t \boldsymbol{x}_t^\top$; on other rounds it may receive arbitrary positive-semidefinite block additions. Then*

$$\sum_{t \in \mathcal{R}} \min\{1, \|\boldsymbol{x}_t\|^2_{\boldsymbol{W}_{t-1}^{-1}}\} \leq 2 \log \frac{\det \boldsymbol{W}_T}{\det(\lambda \boldsymbol{I})}. \tag{39}$$

If $\lambda \geq L^2$ and $\|\boldsymbol{x}_t\|_2 \leq L$, the minimum can be removed. If $\boldsymbol{W}_T = \lambda \boldsymbol{I} + \sum_{r=1}^N \boldsymbol{y}_r \boldsymbol{y}_r^\top$ consists of N distinct records with $\|\boldsymbol{y}_r\|_2 \leq L$, the right side is at most $2d \log(1 + NL^2/(d\lambda))$.

Proof. This is a slight extension of the standard elliptical potential lemma (Abbasi-Yadkori et al., 2011, Lemma 11). Write $w_t = \|\boldsymbol{x}_t\|_{\boldsymbol{W}_{t-1}^{-1}}$. Determinant monotonicity and the matrix determinant lemma give, for $t \in \mathcal{R}$,

$$\log \det \boldsymbol{W}_t - \log \det \boldsymbol{W}_{t-1} \geq \log(1 + w_t^2).$$

All other log-determinant increments are nonnegative. Sum these inequalities and use $\min\{1, y\} \leq 2\log(1+y)$ for $y \geq 0$. If $\lambda \geq L^2$, $w_t^2 \leq L^2/\lambda \leq 1$. The last claim follows from the trace–determinant inequality $\det \boldsymbol{W}_T \leq (\operatorname{tr} \boldsymbol{W}_T/d)^d$. The additional matrix increments increase the potential without adding terms to the sum over $\mathcal{R}$. □

Lemma 9 (Conditional information accumulation)**.** *Let $\mathcal{F}_{t-1} \subseteq \mathcal{P}_t \subseteq \mathcal{F}_t$, where $\mathcal{P}_t$ contains the current arrival but precedes the current contexts. Let $B_t \in \{0,1\}$ be known in $\mathcal{P}_t$. On included rounds, suppose $\boldsymbol{Y}_t$ is $\mathcal{F}_t$-measurable, $0 \preceq \boldsymbol{Y}_t \preceq H\boldsymbol{I}$ and $\mathbb{E}[\boldsymbol{Y}_t \mid \mathcal{P}_t] \succeq a\boldsymbol{I}$, where $H, a > 0$ are fixed. Set $N_t = \sum_{s \leq t} B_s$ and $\boldsymbol{S}_t = \sum_{s \leq t} B_s \boldsymbol{Y}_s$. For every $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\lambda_{\min}(\boldsymbol{S}_t) \geq (1 - e^{-1}) a N_t - H \log(d/\delta) \quad \textit{simultaneously for all } t. \tag{40}$$

Proof. Write $c_e = 1 - e^{-1}$. The secant bound $e^{-\boldsymbol{Y}_t/H} \preceq \boldsymbol{I} - c_e \boldsymbol{Y}_t/H$ implies $\mathbb{E}[e^{-\boldsymbol{Y}_t/H} \mid \mathcal{P}_t] \preceq e^{-c_e a/H}\boldsymbol{I}$. The conditional matrix exponential argument (Tropp, 2012, Section 2), using Golden–Thompson and this scalar upper bound, makes

$$Z_t = \operatorname{tr}\exp(-\boldsymbol{S}_t/H + c_e a N_t \boldsymbol{I}/H)$$

a nonnegative supermartingale starting at d; it is unchanged when $B_t = 0$. A failure of (40) entails $Z_t > d/\delta$, so Ville's inequality proves the claim. This argument controls adaptive observations after the uniform prefix and is also used by the Gaussian extension. The initial uniform prefix itself uses independent matrix Chernoff. □

Lemma 10 (Scalar accuracy calibration)**.** *For $A, c, R, \lambda, S, \widetilde{\gamma}, d > 0$, the function*

$$g(D) := R\sqrt{\frac{A + d\log(1+cD)}{D}} + \frac{\lambda S}{D}$$

is decreasing on $D > 0$, and $g(D) < \widetilde{\gamma}/4$ whenever

$$D > \frac{128R^2}{\widetilde{\gamma}^2}\left[A + d\log\left(1 + \frac{128R^2 dc}{\widetilde{\gamma}^2}\right)\right] + \frac{8\lambda S}{\widetilde{\gamma}}.$$

Proof. The derivative of the squared noise envelope is negative, since

$$\frac{\mathrm{d}}{\mathrm{d}D}\frac{A + d\log(1+cD)}{D} = \frac{dcD/(1+cD) - A - d\log(1+cD)}{D^2} < 0,$$

using $x/(1+x) \le \log(1+x)$. The bias term also decreases. Set $b = \widetilde{\gamma}^2/(128R^2 d)$. The inequality $1 + cD \le (1 + c/b)(1 + bD) \le (1 + c/b)e^{bD}$ gives $\log(1 + cD) \le bD + \log(1 + c/b)$. Under the displayed threshold,

$$\frac{R^2[A + d\log(1+cD)]}{D} \le \frac{R^2[A + d\log(1+c/b)]}{D} + R^2 db < \frac{\widetilde{\gamma}^2}{128} + \frac{\widetilde{\gamma}^2}{128} = \frac{\widetilde{\gamma}^2}{64}, \qquad \frac{\lambda S}{D} < \frac{\widetilde{\gamma}}{8}.$$

Taking the square root and adding the bias proves the claim. □

A.4 Approximate-Cluster Certification

Remark 5 (Certification margin)**.** Under the relaxed model in Remark 1, let $\gamma = \min_{j \neq j'} \|\boldsymbol{\theta}^j - \boldsymbol{\theta}^{j'}\|_2$ denote the representative separation and $0 < \widetilde{\gamma} \le \gamma$ a known lower bound. If $4\epsilon < \widetilde{\gamma}$ and every post-admission personal estimate has error at most $r < \widetilde{\gamma}/4 - \epsilon$ relative to its agent's parameter $\boldsymbol{\theta}_i$, estimated within-cluster distances are at most $2\epsilon + 2r < \widetilde{\gamma}/2$, while cross-cluster distances are at least $\gamma - 2\epsilon - 2r > \widetilde{\gamma}/2$. Thus the same threshold proves (20) and (21). A known bound ϵ allows $\widetilde{\gamma}$ in the definition of n_{cert} in (4) to be replaced by $\widetilde{\gamma} - 4\epsilon$, while the edge threshold remains $\widetilde{\gamma}/2$. This extends graph certification only: pooled rewards have heterogeneous parameters, so the pooled-confidence and regret bounds require a separate bias analysis.

A.5 Smoothed Gaussian Contexts and LC-CLUB-L

We prove Theorem 2 directly for untruncated Gaussian perturbations. The first step is a covariance bound for the selected context, including its nonzero conditional mean.

Lemma 11 (Residual variance outside an interval)**.** *Let $G \sim \mathcal{N}(0,1)$ and let E be the complement of an interval, with $p = \mathbb{P}(G \in E) > 0$. Then*

$$\operatorname{Var}(G \mid G \in E) \ge \frac{1}{256\log(2e/p)}.$$

Intervals may be empty, unbounded, open, or closed.

Proof. A real random variable with density bounded by $f_{\max}$ has variance at least $1/(32 f_{\max}^2)$: an interval of radius $1/(4f_{\max})$ about its mean has probability at most $1/2$, so integrate the squared distance outside it. For a Gaussian conditioned on a half-line of probability q, its density is bounded by $2\sqrt{\log(e/q)}$. To verify this, by symmetry consider $G \ge a$. If $a < 0$, then $q \ge 1/2$ and the

density is at most $2/\sqrt{2\pi}$. If $a \geq 0$, the bound $\overline{\Phi}(a) \geq \phi(a)/(a+1)$ follows by differentiating $\overline{\Phi}(a) - \phi(a)/(a+1)$, whose derivative is $-a\phi(a)/(a+1)^2$ and whose limit at infinity is zero. Also $q \leq e^{-a^2/2}$, so the maximal conditional density satisfies

$$\frac{\phi(a)}{q} \leq a+1 \leq \sqrt{2(a^2+1)} \leq 2\sqrt{\log(e/q)}.$$

Consequently the half-line conditional variance is at least $1/[128\log(e/q)]$. An interval complement is a union of at most two disjoint half-lines, up to null endpoints; partition the full line at zero if necessary. One piece has mass $q \geq p/2$ and conditional mixture weight at least $1/2$. The variance decomposition for this mixture gives the claim. □

Explicit anytime constants. Put $B = M_0/\sigma_-$ and $\kappa = \sigma_+^2/\sigma_-^2$. Let $r \geq 1$ be the smallest number satisfying

$$K2^{d/2}e^{-r/4}\{2B^2 + 2\kappa(r+4)\} \leq \alpha_K/2.$$

The left side decreases in r. Since $r + 4 \leq 8e^{r/8}$, the condition holds at $r_* := \max\{1, 8\log(4K2^{d/2}(B^2+8\kappa)/\alpha_K)\}$, so $r \leq r_*$. Define

$$R_{\mathrm{SA}} = 2B^2 + 2\kappa r, \qquad n_{\mathrm{SA}}(\delta) = \left\lceil \frac{16R_{\mathrm{SA}}}{\alpha_K}\log\frac{d}{\delta} \right\rceil. \tag{41}$$

Theorem 2 (Information under smoothed contexts)**.** *Under Assumption 3, let a_t maximize $s_t(\boldsymbol{x}_{t,a})$, with smallest-index ties, for a finite convex score s_t chosen from $\mathcal{P}_t$ (jointly measurably). With $\alpha_K = [256\log(2eK)]^{-1}$,*

$$\mathrm{Cov}(\boldsymbol{x}_t \mid \mathcal{P}_t) \succeq \alpha_K \boldsymbol{\Sigma}_t. \tag{9}$$

For any stream selected before its contexts are revealed, with probability at least $1-\delta$, its n-round Gram matrix $\boldsymbol{S}_n$ satisfies

$$\lambda_{\min}(\boldsymbol{S}_n) \geq \tfrac{1}{4}\alpha_K n\sigma_-^2 \quad \textit{for all attained } n \geq n_{\mathrm{SA}}(\delta), \tag{10}$$

with the burn-in $n_{\mathrm{SA}}(\delta)$ in (41)*.*

Proof. We first prove (9). Fix $\mathcal{P}_t$ and a unit information direction $\boldsymbol{v}$ in whitened coordinates. Write $\boldsymbol{\Sigma}_t^{-1/2}\boldsymbol{x}_{t,k} = \boldsymbol{m}_k + G_k\boldsymbol{v} + \boldsymbol{U}_k$ for candidate $k \in [K]$, where G_k is standard normal, $\boldsymbol{U}_k \perp \boldsymbol{v}$, and all Gaussian components are independent. Condition on $\boldsymbol{U}_k$ and all other candidates, denoting this information by W_k. Along this remaining scalar coordinate the score is convex. Let b_k be the largest competing score. If a smaller-index competitor attains b_k, candidate k wins exactly when its score exceeds b_k; otherwise it wins when its score is at least b_k. A strict or non-strict superlevel set of a convex scalar function is the complement of an interval. For $K = 1$ the winning set is the whole line. Let $\pi_k(W_k)$ be its conditional winning probability. For every fixed $b \in \mathbb{R}$, Lemma 11 and the variance decomposition imply

$$\mathbb{E}[(\boldsymbol{v}^\top\boldsymbol{\Sigma}_t^{-1/2}\boldsymbol{x}_t - b)^2\mathbf{1}_{\{a_t=k\}} \mid W_k, \mathcal{P}_t] \geq \frac{\pi_k(W_k)}{256\log(2e/\pi_k(W_k))},$$

with the right side zero at $\pi_k = 0$. Translations affect the conditional mean but not this residual variance bound. The function $h(p) = p/\log(2e/p)$, extended by $h(0) = 0$, is convex: if $\ell = \log(2e/p)$, then $h''(p) = (\ell+2)/(p\ell^3) > 0$. By iterated expectation, $\mathbb{E}[\pi_k(W_k) \mid \mathcal{P}_t] = \mathbb{P}(a_t = k \mid \mathcal{P}_t)$. Thus $\sum_k \mathbb{E}[\pi_k(W_k) \mid \mathcal{P}_t] = 1$, and Jensen's inequality gives

$$\sum_k \mathbb{E}[h(\pi_k(W_k)) \mid \mathcal{P}_t] \geq Kh(1/K) = \frac{1}{\log(2eK)}.$$

Summing over winners and minimizing over b proves the directional conditional variance bound. A countable dense set of directions and congruence by $\boldsymbol{\Sigma}_t^{1/2}$ give the covariance inequality. Adding the outer product of the conditional mean gives the raw-moment bound. This one-round result only requires finite centers and a positive definite common covariance; uniform spectral bounds enter the realized Gram and certification arguments below.

We next prove (10). Use $\boldsymbol{y}_t = \boldsymbol{x}_t/\sigma_-$ on the declared stream. For any candidate, $\|\boldsymbol{y}_{t,a}\| \le B + \sqrt{\kappa}\|\boldsymbol{Z}_{t,a}\|$. For $Q = \|\boldsymbol{Z}\|^2$, the exponential moment $\mathbb{E}e^{Q/4} = 2^{d/2}$ and tail integration give

$$\mathbb{P}(Q > r) \le 2^{d/2}e^{-r/4}, \qquad \mathbb{E}[Q\mathbf{1}_{\{Q>r\}}] \le 2^{d/2}(r+4)e^{-r/4}.$$

The event $\|\boldsymbol{y}_{t,a}\|^2 > R_{\mathrm{SA}}$ implies $Q > r$. Bounding the selected tail by the sum of the candidate tails thus gives

$$\mathbb{E}[\|\boldsymbol{y}_t\|^2\mathbf{1}_{\{\|\boldsymbol{y}_t\|^2>R_{\mathrm{SA}}\}} \mid \mathcal{P}_t] \le \alpha_K/2.$$

Hence $\boldsymbol{Y}_t = \boldsymbol{y}_t\boldsymbol{y}_t^\top\mathbf{1}_{\{\|\boldsymbol{y}_t\|^2\le R_{\mathrm{SA}}\}}$ obeys $0 \preceq \boldsymbol{Y}_t \preceq R_{\mathrm{SA}}\boldsymbol{I}$ and $\mathbb{E}[\boldsymbol{Y}_t \mid \mathcal{P}_t] \succeq \alpha_K\boldsymbol{I}/2$. Apply Lemma 9 to this predictable stream with increments $\boldsymbol{Y}_t$, $H = R_{\mathrm{SA}}$, $a = \alpha_K/2$, and failure level δ. Since $c_e = 1 - e^{-1} > 5/8$, it gives, simultaneously at every reached count n,

$$\lambda_{\min}\left(\sum_{s\le n}\boldsymbol{Y}_s\right) \ge c_e\alpha_K n/2 - R_{\mathrm{SA}}\log(d/\delta) \ge \alpha_K n/4 \quad \text{for } n \ge n_{\mathrm{SA}}(\delta).$$

The last step uses $(c_e/2 - 1/4)\alpha_K n \ge \alpha_K n/16$ and (41). The lemma includes streams that stop. Since $\boldsymbol{S}_n = \sigma_-^2\sum_{s\le n}\boldsymbol{y}_s\boldsymbol{y}_s^\top \succeq \sigma_-^2\sum_{s\le n}\boldsymbol{Y}_s$, the claimed lower bound on $\lambda_{\min}(\boldsymbol{S}_n)$ follows. The scaling and clipping are only in the proof; they change neither the policy nor the observations. □

Regret of LC-CLUB-L. Consider Algorithm 1 with its uniform draw replaced by personal LinUCB: before certification, agent i selects the smallest-index maximizer of $\boldsymbol{x}^\top\widehat{\boldsymbol{\theta}}_{i,n} + \bar{\beta}_n\|\boldsymbol{x}\|_{\boldsymbol{V}_{i,n}^{-1}}$, with the count-based coefficient $\bar{\beta}_n$ below, where $\boldsymbol{V}_{i,n} = \lambda\boldsymbol{I} + \boldsymbol{S}_{i,n}$ uses all of its own observations. It certifies at the fixed own count n_L using the latest personal estimates and the usual $\widetilde{\gamma}/2$ comparison, then uses direct-neighbor Full-history pooled LinUCB. The fixed count includes the triggering action; when an agent has fewer than n_L arrivals it never certifies. Let $\alpha_K = [256\log(2eK)]^{-1}$, $s_0 = \sigma_-^2$, and set, for $0 < \delta \le 1/2$,

$$\begin{aligned}
L_T &= M_0 + \sigma_+\sqrt{2d\log 2 + 4\log(3KT/\delta)},\\
\beta_t &= R\sqrt{d\log\left(1 + \frac{tL_T^2}{d\lambda}\right) + 2\log\frac{3u}{\delta}} + \sqrt{\lambda}S,\\
\bar{\beta}_n &= R\sqrt{d\log\left(1 + \frac{nL_T^2}{d\lambda}\right) + 2\log\frac{3u}{\delta}} + \sqrt{\lambda}S.
\end{aligned}$$

Assume $\lambda \ge L_T^2$. Define $A = 2\log(3u/\delta)$ and

$$\begin{aligned}
D_L &= \frac{128R^2}{\widetilde{\gamma}^2}\left[A + d\log\left(1 + \frac{512R^2L_T^2}{\widetilde{\gamma}^2\lambda\alpha_K s_0}\right)\right] + \frac{8\lambda S}{\widetilde{\gamma}},\\
n_L &= \left\lceil \max\left\{n_{\mathrm{SA}}\left(\frac{\delta}{3u}\right), \frac{4D_L}{\alpha_K s_0}\right\}\right\rceil.
\end{aligned} \tag{42}$$

Lemma 12 (Persistent personal accuracy for LC-CLUB-L)**.** *On an event $\mathcal{E}_{\mathrm{loc}}$ of probability at least $1-\delta$, every agent and every attained own count $n_L \le n \le T$ satisfy $\|\widehat{\boldsymbol{\theta}}_{i,n} - \boldsymbol{\theta}_i\|_2 < \widetilde{\gamma}/4$. Consequently its admission comparisons recover exactly the true-cluster edges to previously certified agents, and its closed certified neighborhood is its certified true-cluster slice.*

Proof. Let $\mathcal{E}_{\mathrm{norm}}$ be the event that all KT candidate norms are at most L_T. Since $\mathbb{E}e^{\|Z\|^2/4} = 2^{d/2}$, a union bound gives failure at most $\delta/3$. For each agent, its arrival indicator is known before contexts are drawn, and its selected context is the maximizer of a finite convex score measurable before that draw. Before certification this is the personal LinUCB index with coefficient $\bar{\beta}_n$. After certification, the direct-neighbor pooled LinUCB index is also finite and convex; its pool, estimate, design, and confidence coefficient are fixed before the Gaussian menu. Smallest-index tie breaking is exactly the convention in Theorem 2; thus its anytime Gram bound applies to the entire personal stream, on both sides of certification, without requiring graph correctness. A union bound at level $\delta/(3u)$ gives $\mathcal{E}_{\mathrm{gram}}$ with failure at most $\delta/3$ and $\lambda_{\min}(\boldsymbol{S}_{i,n}) \ge \alpha_K s_0 n/4$ for every attained $n \ge n_{\mathrm{SA}}(\delta/(3u))$.

Apply the determinant-form self-normalized argument of Lemma 1 with confidence parameter $\delta/3$; it does not require bounded contexts. Denote its simultaneous personal confidence event by $\mathcal{E}_{\mathrm{per}}$. On $\mathcal{E}_{\mathrm{per}} \cap \mathcal{E}_{\mathrm{norm}}$, the trace–determinant inequality bounds the error at count n by

$$R\sqrt{\frac{A + d\log(1 + nL_T^2/(d\lambda))}{\lambda_{\min}(\boldsymbol{V}_{i,n})}} + \frac{\lambda S}{\lambda_{\min}(\boldsymbol{V}_{i,n})}.$$

Put $D_n = \lambda + \alpha_K s_0 n/4$. On $\mathcal{E}_{\mathrm{gram}}$, $\lambda_{\min}(\boldsymbol{V}_{i,n}) \geq D_n$ and $n \leq 4D_n/(\alpha_K s_0)$, so the displayed error is at most the function $g(D_n)$ in Lemma 10 with $A = 2\log(3u/\delta)$ and $c = 4L_T^2/(d\lambda\alpha_K s_0)$. Here $128R^2dc/\widetilde{\gamma}^2 = 512R^2L_T^2/(\widetilde{\gamma}^2\lambda\alpha_K s_0)$, so the lemma's threshold is exactly D_L. Since $n \geq n_L$ implies $D_n > D_L$, the accuracy claim follows on $\mathcal{E}_{\mathrm{loc}} := \mathcal{E}_{\mathrm{norm}} \cap \mathcal{E}_{\mathrm{gram}} \cap \mathcal{E}_{\mathrm{per}}$. The exact-neighborhood conclusion follows from the usual triangle inequalities and latest estimates. The three events have total failure at most $\delta/3 + \delta/3 + \delta/3 = \delta$. □

Theorem 5 (Regret of LC-CLUB-L)**.** *Under Assumptions 1 and 3, the chronological noise condition of Section 3, and stochastic or oblivious arrivals as in Remark 2, suppose $\|\boldsymbol{\theta}_i\|_2 \leq S$, $\lambda \geq L_T^2$, and use the choices above. For any $T \geq 1$ and $0 < \delta \leq 1/2$, with probability at least $1 - 2\delta$,*

$$\begin{aligned}\mathrm{Reg}(T) \leq 2\bar{\beta}_{n_L}\sqrt{2d\log\left(1 + \frac{n_L L_T^2}{d\lambda}\right)}\sum_{i=1}^{u}\sqrt{\min\{T_i, n_L\}} \\ + 2\beta_T\sqrt{2dmT\log\left(1 + \frac{TL_T^2}{d\lambda}\right)}.\end{aligned} \tag{43}$$

Moreover, with the T-independent constant

$$C_x := 2M_0^2 + 2\sigma_+^2(2d\log 2 + 4\log K + 4),$$

the expected pseudo-regret satisfies

$$\begin{aligned}\mathbb{E}[\mathrm{Reg}(T)] \leq 2\bar{\beta}_{n_L}\sqrt{2d\log\left(1 + \frac{C_x n_L}{d\lambda}\right)}\sum_{i=1}^{u}\sqrt{\mathbb{E}[\min\{T_i, n_L\}]} \\ + 2\beta_T\sqrt{2dmT\log\left(1 + \frac{TL_T^2}{d\lambda}\right)} + 2ST\sqrt{2\delta C_x}.\end{aligned} \tag{44}$$

Proof. Condition on an arrival path permitted by Remark 2, enlarging the pre-context sigma-field $\mathcal{P}_t$ by that path. The context and chronological noise laws are preserved. The fixed own count n_L makes certification times and the true-cluster slices of certified agents deterministic given the path. Reapply Lemma 12 conditionally to obtain $\mathcal{E}_{\mathrm{loc}}$ with failure at most δ. The determinant-form calendar-time argument of Lemma 4 gives $\mathcal{E}_{\mathrm{sh}}$ with conditional failure at most δ; its radius is bounded by β_t on $\mathcal{E}_{\mathrm{norm}}$. Thus $\mathcal{E} := \mathcal{E}_{\mathrm{loc}} \cap \mathcal{E}_{\mathrm{sh}}$ has conditional probability at least $1 - 2\delta$, contains the required personal and shared confidence conclusions, and is a subset of $\mathcal{E}_{\mathrm{norm}}$. On $\mathcal{E}$, actual certified neighborhoods equal the fixed true-cluster slices; no independence of events is needed.

For agent i, the path fixes its pre-certification rounds $\mathcal{T}_i^{\mathrm{pre}}$ and their number $q_i := \min\{T_i, n_L\}$, including the triggering action. Set

$$\boldsymbol{V}_i^{\mathrm{pre}} := \lambda\boldsymbol{I} + \sum_{t \in \mathcal{T}_i^{\mathrm{pre}}} \boldsymbol{x}_t\boldsymbol{x}_t^\top, \qquad Z_i := \sum_{t \in \mathcal{T}_i^{\mathrm{pre}}} \|\boldsymbol{x}_t\|^2.$$

On $\mathcal{E}$, personal optimism gives $\Delta_t \leq 2\bar{\beta}_{n_L}\|\boldsymbol{x}_t\|_{\boldsymbol{V}_{i,n_i(t-1)}^{-1}}$ before certification, since $\bar{\beta}_n$ is increasing and $n_i(t-1) < n_L$. As $\|\boldsymbol{x}_t\|^2 \leq L_T^2 \leq \lambda$, Cauchy–Schwarz and the elliptical potential, followed by the trace–determinant inequality, give

$$\begin{aligned}\mathrm{Reg}_i^{\mathrm{pre}} &\leq 2\bar{\beta}_{n_L}\sqrt{2q_i\log(\det\boldsymbol{V}_i^{\mathrm{pre}}/\lambda^d)} \\ &\leq \psi_i(Z_i), \qquad \psi_i(z) := 2\bar{\beta}_{n_L}\sqrt{2dq_i\log\Big(1 + \frac{z}{d\lambda}\Big)}.\end{aligned}$$

Using $Z_i \le q_i L_T^2 \le n_L L_T^2$ here gives the first term of (43); we retain $\psi_i(Z_i)$ for expectation.

For certified rounds, invoke the Section 4 argument: Lemmas 3, 4, and 5, and the proof of Theorem 4, with context bound L_T on $\mathcal{E}$. The exact direct-neighbor pools and Full-history confidence yield the unchanged second term of (43). Adding the pre-certification term proves the high-probability bound conditionally on the path and hence unconditionally.

For expectation, Gaussian tail integration gives $\mathbb{E}[\max_{a\le K} \|\boldsymbol{Z}_{t,a}\|^2 \mid \mathcal{P}_t] \le 2d\log 2 + 4\log K + 4$: integrate the tail bound $\min\{1, K2^{d/2}e^{-z/4}\}$. Uniformly over the admissible centers and covariances, this implies

$$\mathbb{E}[\|\boldsymbol{x}_t\|^2 \mid \mathcal{P}_t] \le \mathbb{E}[\max_a \|\boldsymbol{x}_{t,a}\|^2 \mid \mathcal{P}_t] \le C_x, \qquad \mathbb{E}[Z_i \mid \text{path}] \le C_x q_i,$$

where the second inequality uses the tower property and the fixed round set $\mathcal{T}_i^{\text{pre}}$. Since $\psi_i \ge 0$,

$$\begin{aligned}\mathbb{E}[\text{Reg}_i^{\text{pre}}\mathbf{1}_{\mathcal{E}} \mid \text{path}] &\le \mathbb{E}[\psi_i(Z_i)\mathbf{1}_{\mathcal{E}} \mid \text{path}] \\ &\le \mathbb{E}[\psi_i(Z_i) \mid \text{path}] \le \psi_i(\mathbb{E}[Z_i \mid \text{path}]) \le \psi_i(C_x q_i).\end{aligned}$$

No Gaussian law is ever conditioned on $\mathcal{E}$; the indicator is simply dropped. The last two steps use concavity and monotonicity of $z \mapsto \sqrt{\log(1 + z/(d\lambda))}$ on $z \ge 0$. Use $q_i \le n_L$ inside the logarithm, then average over paths and apply $\mathbb{E}[\sqrt{q_i}] \le \sqrt{\mathbb{E}[q_i]}$ to obtain the first term of (44). For stochastic arrivals, $\mathbb{E}[\min\{T_i, n_L\}] \le \min\{\mathbb{E}[T_i], n_L\}$ by concavity of the minimum; for an oblivious path q_i is deterministic. The certified contribution on $\mathcal{E}$ retains its deterministic bound. Finally, $\text{Reg}(T) \le 2S\sum_{t\le T}\max_a \|\boldsymbol{x}_{t,a}\|$ and Cauchy–Schwarz give $\mathbb{E}[\text{Reg}(T)^2] \le 4S^2T^2C_x$. As $\mathbb{P}(\mathcal{E}^c) \le 2\delta$, another application gives

$$\mathbb{E}[\text{Reg}(T)\mathbf{1}_{\mathcal{E}^c}] \le \sqrt{\mathbb{E}[\text{Reg}(T)^2]\mathbb{P}(\mathcal{E}^c)} \le 2ST\sqrt{2\delta C_x},$$

which completes (44). □

Corollary 2 (Expected regret rate of LC-CLUB-L)**.** *Under the assumptions of Theorem 5, let $\delta = T^{-2}$ and $\lambda = L_T^2$, and fix d, R, S, M_0, σ_+. Then, as $T \to \infty$,*

$$\mathbb{E}[\text{Reg}(T)] = \widetilde{O}\left(\frac{u\sqrt{d}\log T}{\sqrt{\alpha_K}\,\sigma_-\widetilde{\gamma}} + d\sqrt{mT}\log T\right), \tag{45}$$

where $\widetilde{O}$ hides factors logarithmic in K, $1/\sigma_-$ and $1/\widetilde{\gamma}$, and the constants do not depend on u or m. This is the rate in Theorem 3.

Proof. Here $L_T^2 = \widetilde{O}(\log T)$ and $A = 2\log(3u/\delta) = \Theta(\log T)$. In (41), $B^2 = M_0^2/\sigma_-^2$, $\kappa = \sigma_+^2/\sigma_-^2$, and $r \le r_*$ is logarithmic in these quantities and in K, so $n_{\text{SA}}(\delta/(3u)) = \widetilde{O}(\log T/(\alpha_K\sigma_-^2))$. In D_L, the choice $\lambda = L_T^2$ makes the logarithmic factor $\log(1 + 512R^2/(\widetilde{\gamma}^2\alpha_K s_0)) = \widetilde{O}(1)$, and $\widetilde{\gamma} \le 2S$ bounds $\lambda S/\widetilde{\gamma}$ by $2S^2L_T^2/\widetilde{\gamma}^2$. The definition (42) therefore gives

$$n_L = \widetilde{O}\left(\frac{\log T}{\alpha_K\sigma_-^2\widetilde{\gamma}^2}\right).$$

The terms A and $\sqrt{\lambda}S = L_T S$ give $\bar{\beta}_{n_L} = \widetilde{O}(\sqrt{\log T})$ and $\beta_T = \widetilde{O}(\sqrt{d\log T})$. Moreover $C_x = \widetilde{O}(1)$ and $n_L/\lambda = \widetilde{O}(1/(\alpha_K\sigma_-^2\widetilde{\gamma}^2))$, so $\log(1 + C_x n_L/(d\lambda)) = \widetilde{O}(1)$. Substituting these bounds and $\sum_i \sqrt{\mathbb{E}[q_i]} \le u\sqrt{n_L}$ into (44) gives (45); with $\delta = T^{-2}$, the failure term is $2S\sqrt{2C_x} = \widetilde{O}(1)$. □

The pre-certification term scales with the square root of the certification count rather than the count itself, improving its leading polynomial dependence on the supplied separation lower bound from $1/\widetilde{\gamma}^2$ to $1/\widetilde{\gamma}$, up to logarithmic factors. Since $1/\sqrt{\alpha_K} = 16\sqrt{\log(2eK)}$, its information-growth factor is $\sqrt{\log K}/\sigma_-$. The expectation in this rate is substantive: it uses the bounded expected-potential logarithm $\log(1 + C_x n_L/(d\lambda))$. For fixed parameters, directly simplifying the high-probability bound (43) instead leaves a $\sqrt{\log\log T}$ factor in the pre-certification term, since its potential logarithm is $\log(1 + n_L/d)$ when $\lambda = L_T^2$.

A.6 UNIFORM-SPHERE CONTEXTS

Scope. This subsection records what we can prove when candidates are fresh uniform draws on the unit sphere $\mathbb{S}^{d-1}$, which satisfy Assumption 2 with $L = 1$ and $\lambda_x = 1/d$. No theorem in the main text uses it. Proposition 1 is a finite-time bound for linear scores without an exploration bonus. Corollary 3 covers LinUCB but is only asymptotic, so it yields no certification threshold. Its convergence statement holds for all fixed $d \geq 2$ and $K \geq 1$, but the proposition and the corollary's explicit $1/(2d)$ lower bound require $d \geq 4[\log(2K) + 1]$.

Proposition 1 (Greedy information on a high-dimensional sphere)**.** *Let the candidates be fresh independent uniform draws on $\mathbb{S}^{d-1}$, $d \geq 2$, and let a_t maximize a linear score chosen from $\mathcal{P}_t$, with smallest-index ties. If $d \geq 4[\log(2K) + 1]$, then*

$$\mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^\top \mid \mathcal{P}_t] \succeq \frac{1}{2d} \boldsymbol{I}.$$

Consequently, for any stream selected before its contexts are revealed, with probability at least $1-\eta$, its Gram matrix satisfies $\boldsymbol{S}_n \succeq n\boldsymbol{I}/(4d)$ for every attained $n \geq \lceil 16d \log(d/\eta) \rceil$.

Proof. For a nonzero score, rotate its unit direction to $\boldsymbol{e}_1$ and write $Z_a = \boldsymbol{e}_1^\top \boldsymbol{x}_{t,a}$ and $M_K = \max_a Z_a$. Rotational symmetry about $\boldsymbol{e}_1$ makes the selected context's second-moment matrix diagonal, with axial eigenvalue $\mathbb{E}M_K^2$ and, since $\|\boldsymbol{x}_t\| = 1$, transverse eigenvalues $(1-\mathbb{E}M_K^2)/(d-1)$. If F is the symmetric continuous CDF of Z_a, then for $r \geq 0$, $\mathbb{P}(|M_K| \leq r) = F(r)^K - (1-F(r))^K \leq 2F(r) - 1 = \mathbb{P}(|Z_1| \leq r)$, so $\mathbb{E}M_K^2 \geq \mathbb{E}Z_1^2 = 1/d$. The spherical coordinate moments satisfy

$$\mathbb{E}Z_1^{2j} = \frac{1 \cdot 3 \cdots (2j-1)}{d(d+2)\cdots(d+2j-2)} \leq \frac{(2j-1)!!}{d^j},$$

so $\mathbb{E}e^{sZ_1} \leq e^{s^2/(2d)}$ and $\mathbb{P}(\max_a Z_a^2 > z) \leq \min\{1, 2Ke^{-dz/2}\}$. Since $M_K^2 \leq \max_a Z_a^2$, tail integration gives $\mathbb{E}M_K^2 \leq 2[\log(2K) + 1]/d \leq 1/2$. Each transverse eigenvalue is therefore at least $1/[2(d-1)] \geq 1/(2d)$. A zero score selects the first candidate, whose moment is $\boldsymbol{I}/d$.

For the realized bound, apply Lemma 9 with $\boldsymbol{Y}_t = \boldsymbol{x}_t \boldsymbol{x}_t^\top$, $H = 1$ and $a = 1/(2d)$. Since $(1 - e^{-1})/(2d) - 1/(4d) > 1/(16d)$, its conclusion gives $\lambda_{\min}(\boldsymbol{S}_n) \geq n/(4d)$ whenever $n \geq 16d \log(d/\eta)$. □

Corollary 3 (Asymptotic spherical information for LinUCB)**.** *Fix $d \geq 2$ and $K \geq 1$. On one agent's own clock, let each context set consist of K fresh uniform draws on $\mathbb{S}^{d-1}$, and observe $y_t = \boldsymbol{x}_t^\top \boldsymbol{\theta}_* + \epsilon_t$ for a fixed $\boldsymbol{\theta}_* \neq \boldsymbol{0}$, where ϵ_t is R-sub-Gaussian given the history, the current context set and the action. For $\lambda > 0$, using only these observations, form $\boldsymbol{S}_n = \sum_{t \leq n} \boldsymbol{x}_t \boldsymbol{x}_t^\top$, $\boldsymbol{V}_n = \lambda \boldsymbol{I} + \boldsymbol{S}_n$ and $\widehat{\boldsymbol{\theta}}_n = \boldsymbol{V}_n^{-1} \sum_{t \leq n} \boldsymbol{x}_t y_t$, and select a maximizer of $\boldsymbol{x}^\top \widehat{\boldsymbol{\theta}}_{t-1} + \beta_t \|\boldsymbol{x}\|_{\boldsymbol{V}_{t-1}^{-1}}$ with smallest-index ties, where $\beta_t \geq 0$ is chosen before the context set and $\beta_t = o(\sqrt{t})$ almost surely. Let $\boldsymbol{G}(\boldsymbol{\theta}_*) = \mathbb{E}[\boldsymbol{X}_{A_*} \boldsymbol{X}_{A_*}^\top]$, where $A_* = \arg\max_{a \leq K} \boldsymbol{\theta}_*^\top \boldsymbol{X}_a$ on a fresh context set. Then $\boldsymbol{S}_n/n \to \boldsymbol{G}(\boldsymbol{\theta}_*)$ almost surely. If, in addition, $d \geq 4[\log(2K) + 1]$, then*

$$\liminf_{n \to \infty} \frac{\lambda_{\min}(\boldsymbol{S}_n)}{n} \geq \frac{1}{2d} \quad \textit{almost surely}.$$

Proof. Write $\boldsymbol{C}_t = \mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^\top \mid \mathcal{P}_t]$. The all-selector floor (32) with $L = 1$ and $\lambda_x = 1/d$ gives $\boldsymbol{C}_t \succeq d^{-K} \boldsymbol{I}$. The entries of $\boldsymbol{D}_t = \boldsymbol{x}_t \boldsymbol{x}_t^\top - \boldsymbol{C}_t$ are bounded martingale differences, so $\sum_t \boldsymbol{D}_t/t$ converges almost surely and Kronecker's lemma yields $n^{-1} \sum_{t \leq n} \boldsymbol{D}_t \to \boldsymbol{0}$. Hence $\boldsymbol{S}_n \succeq d^{-K} n \boldsymbol{I}/2$ eventually, almost surely.

Since $\operatorname{tr} \boldsymbol{S}_n = n$, the log-determinant ratio $\log[\det \boldsymbol{V}_n / \det(\lambda \boldsymbol{I})]$ is $O(\log n)$. Lemma 7 at level $(n+1)^{-2}$, Borel–Cantelli and the eigenvalue bound give $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_*\|_2 = O(\sqrt{\log n / n})$ almost surely. Moreover, $\sup_{\|\boldsymbol{x}\|=1} \beta_t \|\boldsymbol{x}\|_{\boldsymbol{V}_{t-1}^{-1}} \leq \beta_t / \sqrt{\lambda_{\min}(\boldsymbol{V}_{t-1})} \to 0$ almost surely. Fix such a probability-one path of scoring parameters and integrate over one fresh test context set. Because $\boldsymbol{\theta}_* \neq \boldsymbol{0}$, the true linear score has a unique maximizer on almost every fresh context set, so uniform convergence of the scores gives convergence of the selected outer product; it has norm one, and dominated convergence yields $\boldsymbol{C}_t \to \boldsymbol{G}(\boldsymbol{\theta}_*)$ almost surely. Cesàro averaging and the martingale limit give $\boldsymbol{S}_n/n \to \boldsymbol{G}(\boldsymbol{\theta}_*)$. Proposition 1, applied to the linear score $\boldsymbol{\theta}_*^\top \boldsymbol{x}$, gives $\boldsymbol{G}(\boldsymbol{\theta}_*) \succeq \boldsymbol{I}/(2d)$. □

What the sphere does not give. The coefficient d^{-K} above serves only to establish consistency; the corollary gives no finite-time rate and does not cover zero signal or estimates formed from heterogeneous pools. The sphere's symmetry also does not extend Proposition 1 to convex scores. Fix a unit vector $\boldsymbol{v}$ and $0 < \varepsilon < 1$. The score $\sqrt{\boldsymbol{x}^\top(\boldsymbol{I} - \varepsilon \boldsymbol{v}\boldsymbol{v}^\top)\boldsymbol{x}}$ selects the candidate with the smallest $|\boldsymbol{v}^\top \boldsymbol{x}|$. For fixed $d \geq 2$, the signed projection $\boldsymbol{v}^\top \boldsymbol{x}$ has density $c_d = \Gamma(d/2)/(\sqrt{\pi}\,\Gamma((d-1)/2))$ at zero, so $\mathbb{E}\min_{a \leq K}(\boldsymbol{v}^\top \boldsymbol{x}_{t,a})^2 \sim 1/(2c_d^2 K^2)$ as $K \to \infty$. Even for large d, a UCB state with $\widehat{\boldsymbol{\theta}} = \mathbf{0}$, $\beta > 0$, $K = 2$ and $\boldsymbol{V}^{-1} \propto \boldsymbol{I} - \varepsilon \boldsymbol{v}\boldsymbol{v}^\top$ gives $d\,\mathbb{E}\min_{a \leq 2}(\boldsymbol{v}^\top \boldsymbol{x}_{t,a})^2 \to 1 - 2/\pi < 1/2$, below the greedy floor. The conditional bound of Proposition 1 therefore does not hold for every LinUCB state; Corollary 3 recovers its rate only asymptotically, as the exploration bonus vanishes.

B Limitations and Future Directions

A general information condition. Certification requires the smallest eigenvalue of each agent's design to grow linearly in its observation count. LC-CLUB-U enforces this with n_{cert} uniform rounds. Theorem 2 shows that under Gaussian perturbations every predictable convex score, LinUCB included, achieves it at rate $\alpha_K \sigma_-^2$ with α_K of order $1/\log K$. For uniform-sphere contexts with $d \geq 4[\log(2K) + 1]$, Appendix A.6 proves a conditional second-moment floor $\boldsymbol{I}/(2d)$ for predictable linear scores and the same asymptotic Gram coefficient for personal LinUCB with nonzero signal and $\beta_t = o(\sqrt{t})$. Under Assumption 2 alone, however, the rate guaranteed for arbitrary selectors is only $L^2(\lambda_x/L^2)^K$, which a two-point context law attains. A general sufficient condition on the context law, stated for the selector actually used rather than for all selectors, would remove the uniform phase beyond these cases. Finding such a general sufficient condition on the contexts remains open.

Separation in the span of the contexts. Our analysis assumes full-rank contexts ($\lambda_x > 0$) and ℓ_2-separated parameters. Neither is necessary for clustering. Rewards depend on $\boldsymbol{\theta}_i$ only through $\boldsymbol{P}\boldsymbol{\theta}_i$, where $\boldsymbol{P}$ projects onto the span $\mathcal{U}$ of the candidate contexts. Agents with $\boldsymbol{P}\boldsymbol{\theta}_i = \boldsymbol{P}\boldsymbol{\theta}_\ell$ can be pooled even if $\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_\ell$, and parameters that differ only outside $\mathcal{U}$ need not be separated. Since ridge estimates lie in $\mathcal{U}$, a certificate for $\|\boldsymbol{P}(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\|_2$ needs only the smallest eigenvalue of Σ_x restricted to $\mathcal{U}$ and the projected gap $\min \|\boldsymbol{P}(\boldsymbol{\theta}^j - \boldsymbol{\theta}^{j'})\|_2$ over clusters that differ in $\mathcal{U}$. More generally, clusters may be distinguishable only through the rewards of the offered arms, $\boldsymbol{x}^\top(\boldsymbol{\theta}^j - \boldsymbol{\theta}^{j'})$. The difficulty is that $\mathcal{U}$ and the restricted eigenvalue floor must be learned from data.

Sharing first or certifying first. The bias term $\lambda S/\lambda_{\min}(\boldsymbol{V}_{i,n})$ in (13) is the error of initializing every agent at $\boldsymbol{\theta} = \mathbf{0}$. Initializing agent i at a shared estimate $\boldsymbol{\theta}_0$, such as the pooled estimate of a certified cluster, replaces it with $\lambda\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_0\|_2/\lambda_{\min}(\boldsymbol{V}_{i,n})$. A large λ then amounts to sharing from the start: the error is small if i belongs to that cluster and of order γ otherwise. A small λ recovers our independent start. The bias $\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_0\|_2$ therefore decides which start is cheaper, and an algorithm could set λ per agent from an estimate of it. The difficulty is that $\boldsymbol{\theta}_0$ is estimated from other agents' data, so the personal confidence bound must hold uniformly over a data-dependent center.

C Experimental Details

C.1 Protocol and Environments

The reported cohort comprises $u = 1000$ agents, $m = 10$ equal clusters, $d = 10$, $K = 40$, and horizon $T = 100{,}000{,}000$. Three environments, five arrival settings, seeds 10–14, and seven policies give 525 completed trajectories. Every table and plot in this section uses this cohort. Tables 2 and 3 in the main text select the raw sphere and constructed space under $\alpha = 0, 1$ and Sine. The main tables display five-seed means to four significant digits (exact zeros as 0); the appendix retains mean $\pm$ SE in tables and mean $\pm$ one SE bands in curves. These are standard errors, with SE equal to the sample standard deviation divided by $\sqrt{5}$. Here we retain all 15 environment–arrival combinations, including the augmented sphere and the two other exponents. The main tables use 10^3 regret units for cumulative regret and original units for the final-1M increment. The large UniCLUB mean increments carry an explicit $\times 10^6$ factor. All reported experimental means and SEs, including intermediate F1, native graph F1, certified-agent counts and paired differences, use four significant

digits; exact zeros are displayed as 0. The appendix retains its original units: cumulative regret in 10^3 and unscaled final-1M increments, as stated in each caption.

Fixed preferences. Write $\mathbb{R}^{10} = \mathcal{H}_2 \oplus \mathcal{A}_4 \oplus \mathcal{B}_4$. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_5$ be unit regular-simplex vertices in $\mathbb{R}^4$ with pairwise inner product $-1/4$, and set $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{10}) = \sqrt{5/3}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_5, -\boldsymbol{v}_1, \ldots, -\boldsymbol{v}_5)$. The construction uses the positive-eigenvalue eigenvectors of $\boldsymbol{I}_5 - \mathbf{1}\mathbf{1}^\top/5$, scaled by $\sqrt{5/4}$ and QR-rotated with seed 123. A Gaussian vector drawn with seed 999 is normalized to give a common $\boldsymbol{g} \in \mathbb{R}^4$ with norm 0.15. The ten preference centers are

$$\boldsymbol{\theta}^j = \left(\frac{1}{\sqrt{2}}, \sqrt{1.8}\, 0.003, \sqrt{1.8}\, \boldsymbol{g}, \sqrt{1.8}\, 0.35 \boldsymbol{w}_j \right).$$

Their minimum gap is $\gamma = \widetilde{\gamma} \approx 0.7425$ and their norms are approximately 0.9529. These base centers are fixed. For each experiment seed, a Haar-random orthogonal $\boldsymbol{Q} \in \mathbb{R}^{4\times 4}$ (the QR factor of a Gaussian matrix, with positive diagonal signs) is drawn once and applied to the $\mathcal{B}_4$ block of every center, $\boldsymbol{\theta}^j_{\mathcal{B}_4} \mapsto \boldsymbol{Q}^\top \boldsymbol{\theta}^j_{\mathcal{B}_4}$, and kept fixed throughout the run. It preserves all norms and pairwise distances, so the gap is unchanged, and all three arm distributions below are invariant under it.

Raw sphere and Gaussian rewards. Each candidate arm is independent and uniform on $\mathbb{S}^9$, so its second moment is $\boldsymbol{I}_{10}/10$. Rewards are $y = \boldsymbol{x}^\top \boldsymbol{\theta}_i + \eta$ with $\eta \sim \mathcal{N}(0, 1)$, without clipping. Here $\lambda_x = 1/10$, $R = 1$, and $\hat{\lambda} = 0.05$.

Bounded spaces and Bernoulli rewards. For each constructed arm, independently draw a symmetric sign ε, a uniform direction $\boldsymbol{v} \in \mathbb{S}^3$, and $J \in \{\mathcal{A}, \mathcal{B}\}$ with probability $1/2$ each, and set

$$\boldsymbol{x} = \left(\frac{1}{\sqrt{2}}, \frac{\varepsilon}{\sqrt{18}}, \frac{2}{3}\boldsymbol{v}\, \mathbf{1}_{\{J=\mathcal{A}\}}, \frac{2}{3}\boldsymbol{v}\, \mathbf{1}_{\{J=\mathcal{B}\}} \right).$$

The augmented-sphere arm is $(1, \boldsymbol{z})/\sqrt{2}$ with $\boldsymbol{z} \sim \mathrm{Unif}(\mathbb{S}^8)$. Both spaces have unit-norm arms and second moment $\mathrm{diag}(1/2, 1/18, \ldots, 1/18)$, giving $\lambda_x = 1/18$ and $\hat{\lambda} = 1/36$. Rewards are $y \sim \mathrm{Bernoulli}(\boldsymbol{x}^\top \boldsymbol{\theta}_i)$, with means in $[0, 1]$ and no extra additive noise or clipping. The empirical radius uses $R = 1/2$.

Context assumptions and interpretation. All three spaces satisfy the bounded i.i.d. candidate and positive second-moment requirements of Assumption 2; the fixed preferences satisfy Assumption 1.

Two-level Zipf arrivals. For agent i in cluster j, let r_j and $s_{i|j}$ be seed-specific random cluster and within-cluster ranks. With $H_{n,\alpha} = \sum_{r=1}^{n} r^{-\alpha}$,

$$p_i = \frac{r_j^{-\alpha}}{H_{10,\alpha}} \frac{s_{i|j}^{-\alpha}}{H_{100,\alpha}}, \qquad \alpha \in \{0, 0.5, 1, 1.5\}.$$

The path is sampled i.i.d. before interaction. The case $\alpha = 0$ is uniform; the other exponents vary the frequency imbalance without changing preferences, contexts or rewards.

Slow sinusoidal arrivals. Independently of contexts, rewards and policies, a seed-specific permutation r_i of $0, \ldots, 999$ defines $\phi_i = 2\pi r_i/1000$ and

$$p_i(t) = \frac{1 + \sin(2\pi t/10^7 + \phi_i)}{1000}.$$

The equally spaced phases make these probabilities sum to one. We sample the full path before interaction and condition on it as an oblivious arrival path. The horizon spans ten periods. Over a complete period every agent has the same expected count, but the realized counts remain random.

C.2 Policy Calibration and Reproducibility

Personal and shared learning. All methods use $\lambda = 1$, width $\sqrt{\boldsymbol{x}^\top \boldsymbol{V}^{-1} \boldsymbol{x}}$ and

$$\beta(N, t) = \sqrt{d \log(1 + N/d) + 4 \log t + \log 2} + 1,$$

where N counts the observations behind the selected estimate. LC-CLUB-L uses its own count during personal LinUCB. LinUCB maintains one model per agent, while the Oracle pools by the true labels.

Inspired by SCLUB's count-dependent confidence function (Li et al., 2019, Algorithm 3) and guided by the personal estimation bound (14) underlying our theoretical certification threshold n_{cert} in (4), we use the empirical certification radius (12) with $\hat{\lambda} = \lambda_x/2$ and $\lambda = S = 1$. Both LC policies use this radius, latest personal estimates, direct certified neighbors and Full-history pooling. LC-CLUB-U explores uniformly until $\rho(n) < \widetilde{\gamma}/4$. LC-CLUB-L uses personal LinUCB, substitutes $n/\log K$ only in the certification radius, and keeps unweighted sufficient statistics. Certification is checked after every personal observation. The exact own-count thresholds are

	LC-CLUB-U	LC-CLUB-L
Raw sphere	15,685	57,859
Both Bernoulli spaces	6,796	25,067

These thresholds are fixed across seeds and arrival laws.

Native clustering baselines. CLUB starts with a complete graph, prunes incident edges using coefficient 1, and pools connected components. SCLUB retains its phase, split and merge rules with coefficients 1 and $\sqrt{2}$, and pools its current sets.

UniCLUB retains its original theoretical exploration length at $\delta = 0.05$ and $L = 1$:

$$T_0 = \left\lceil 16u \log \frac{u}{\delta} + 4u \max\left\{ \frac{8L^2}{\lambda_x} \log \frac{ud}{\delta}, \frac{512d}{\widetilde{\gamma}^2 \lambda_x} \log \frac{u}{\delta} \right\} \right\rceil .$$

This is 3,679,494,966 rounds on the raw sphere and 6,622,964,173 on either Bernoulli space. Its incident-edge deletion radius remains

$$f(n) = \frac{\sqrt{2\log(u/\delta) + d\log(1 + nL^2/d)} + 1}{\sqrt{1 + n\lambda_x/2}}.$$

Edges are deleted when the distance between the two personal estimates exceeds the sum of their radii, including during exploration. Since $T_0 > T$, every UniCLUB action in this cohort is uniform.

Pairing and recording. The experiment seed fixes arrivals, context blocks, reward draws and uniform-arm draws shared across policies. All random streams use NumPy's PCG64 generator seeded by `SeedSequence` with entropy s and spawn key $(2, \text{stream}, b)$, for experiment seed s, zero-based 1000-round block start b ($b = 0$ for non-context streams), and separate streams for arrival ranks, arrival draws, contexts, rewards, uniform-arm choices and the preference rotation. Gaussian noise or Bernoulli uniform draws are shared before applying the policy's selected action. Identical preferences and input-stream hashes were checked across hosts. Each trajectory retains 3701 exact cumulative-regret samples and 52 compact state snapshots. The report uses explicit saved rounds, without interpolation or smoothing. Every setting has the same five seeds. For values z_s, the reported SE is their sample standard deviation divided by $\sqrt{5}$; paired differences and window increments are formed within each seed first.

C.3 Complete Final-Regret and Paired Comparisons

The following tables retain all seven methods and all five arrival laws for each space. Within each setting, all methods face the same inputs. LC-CLUB-L has lower final regret than LinUCB and CLUB in all 75 paired cases, and than SCLUB in $70/75$. Its sole setting-level exception is the augmented sphere at Zipf $\alpha = 1.5$, where it is worse than SCLUB in all five seeds. LC-CLUB-U's regret is dominated by its uniform exploration before certification. Reported percentage reductions are differences of seed-mean final regrets divided by the baseline mean.

Table 4: Raw sphere / Gaussian: final pseudo-regret ($\times 10^3$, mean $\pm$ SE); $u = 1000$, $T = 10^8$, five seeds.

Method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
UniCLUB	60900 ± 1.414	60900 ± 1.336	60900 ± 2.073	60900 ± 0.8872	60900 ± 1.328
LC-CLUB-U	9553 ± 0.5878	9554 ± 0.5603	8884 ± 0.4706	5069 ± 0.3648	9552 ± 0.5897
LinUCB	375.8 ± 0.5857	367.8 ± 1.113	332.1 ± 1.463	265.9 ± 0.5329	377.7 ± 0.4175
CLUB	370.5 ± 1.061	362.3 ± 0.6518	326.2 ± 1.162	326.2 ± 5.362	385.4 ± 1.577
SCLUB	374.1 ± 0.6683	367.0 ± 0.3656	332.4 ± 1.051	265.3 ± 0.5413	376.5 ± 0.7457
LC-CLUB-L	350.4 ± 0.4531	349.1 ± 1.181	319.2 ± 1.532	258.7 ± 0.4555	352.3 ± 0.3591
Oracle	5.702 ± 0.1477	5.712 ± 0.1207	5.554 ± 0.1143	5.481 ± 0.2524	5.958 ± 0.1081

Table 5: Constructed / Bernoulli: final pseudo-regret ($\times 10^3$, mean $\pm$ SE); $u = 1000$, $T = 10^8$, five seeds.

Method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
UniCLUB	33370 ± 0.7838	33370 ± 0.2094	33370 ± 1.287	33360 ± 0.8796	33370 ± 0.8408
LC-CLUB-U	2270 ± 0.1291	2271 ± 0.1737	2272 ± 0.1628	1616 ± 0.2311	2270 ± 0.1243
LinUCB	1816 ± 0.9723	1724 ± 0.2217	1418 ± 1.161	921.5 ± 0.6724	1815 ± 0.7434
CLUB	1816 ± 0.6689	1725 ± 0.9342	1433 ± 1.261	1147 ± 12.90	1845 ± 0.7837
SCLUB	1793 ± 0.7060	1706 ± 1.086	1392 ± 1.805	904.8 ± 1.471	1799 ± 1.299
LC-CLUB-L	1116 ± 0.6424	1123 ± 0.1983	1054 ± 1.037	721.9 ± 0.6207	1116 ± 0.4896
Oracle	76.57 ± 0.3317	75.50 ± 0.08283	71.77 ± 0.2228	65.70 ± 0.1218	76.24 ± 0.3508

Table 6: Augmented sphere / Bernoulli: final pseudo-regret ($\times 10^3$, mean $\pm$ SE); $u = 1000$, $T = 10^8$, five seeds.

Method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
UniCLUB	30180 ± 0.6843	30180 ± 0.8281	30180 ± 0.4589	30180 ± 0.4666	30180 ± 0.6467
LC-CLUB-U	2051 ± 0.2513	2051 ± 0.1980	2051 ± 0.1775	1452 ± 0.1477	2051 ± 0.2056
LinUCB	325.6 ± 0.5203	323.1 ± 0.3255	310.9 ± 0.4448	259.2 ± 0.3325	325.8 ± 0.3409
CLUB	325.7 ± 0.2098	323.4 ± 0.2105	323.9 ± 1.213	427.1 ± 9.278	449.2 ± 2.141
SCLUB	312.7 ± 0.2356	310.9 ± 0.4828	301.3 ± 1.210	251.9 ± 0.5871	315.0 ± 0.6343
LC-CLUB-L	298.7 ± 0.5735	300.4 ± 0.3963	298.8 ± 0.3457	253.5 ± 0.2829	298.9 ± 0.3299
Oracle	3.548 ± 0.07141	3.615 ± 0.03625	3.536 ± 0.06723	3.505 ± 0.04888	3.538 ± 0.05502

Table 7: Seed-paired final regret differences: LC-CLUB-L minus the named method ($\times 10^3$, mean $\pm$ SE). Parentheses count seeds with lower LC-CLUB-L regret; negative is favorable to LC-CLUB-L.

Space	Zipf α / Sine	LinUCB	CLUB	SCLUB
Raw sphere	0	-25.48 ± 0.1715 (5/5)	-20.18 ± 1.239 (5/5)	-23.71 ± 0.9559 (5/5)
Raw sphere	0.5	-18.69 ± 0.1129 (5/5)	-13.22 ± 1.342 (5/5)	-17.89 ± 1.199 (5/5)
Raw sphere	1	-12.86 ± 0.09032 (5/5)	-7.023 ± 1.084 (5/5)	-13.21 ± 1.360 (5/5)
Raw sphere	1.5	-7.157 ± 0.1464 (5/5)	-67.49 ± 4.941 (5/5)	-6.564 ± 0.4824 (5/5)
Raw sphere	Sine	-25.37 ± 0.1032 (5/5)	-33.11 ± 1.399 (5/5)	-24.22 ± 0.5590 (5/5)
Constructed	0	-699.5 ± 0.5072 (5/5)	-700.0 ± 0.4701 (5/5)	-676.9 ± 1.149 (5/5)
Constructed	0.5	-601.7 ± 0.3721 (5/5)	-602.3 ± 1.118 (5/5)	-583.2 ± 1.277 (5/5)
Constructed	1	-364.3 ± 0.2412 (5/5)	-378.9 ± 1.102 (5/5)	-338.3 ± 2.461 (5/5)
Constructed	1.5	-199.6 ± 0.3066 (5/5)	-425.3 ± 12.62 (5/5)	-182.9 ± 1.394 (5/5)
Constructed	Sine	-699.4 ± 0.3667 (5/5)	-729.4 ± 0.7646 (5/5)	-682.7 ± 0.9645 (5/5)
Augmented sphere	0	-26.88 ± 0.1097 (5/5)	-27.00 ± 0.4947 (5/5)	-13.95 ± 0.4646 (5/5)
Augmented sphere	0.5	-22.70 ± 0.08444 (5/5)	-22.97 ± 0.4467 (5/5)	-10.48 ± 0.1735 (5/5)
Augmented sphere	1	-12.05 ± 0.1160 (5/5)	-25.06 ± 1.017 (5/5)	-2.419 ± 0.9103 (5/5)
Augmented sphere	1.5	-5.773 ± 0.08416 (5/5)	-173.6 ± 9.198 (5/5)	1.572 ± 0.4408 (0/5)
Augmented sphere	Sine	-26.81 ± 0.1051 (5/5)	-150.2 ± 1.950 (5/5)	-16.09 ± 0.8458 (5/5)

Table 8: Raw sphere / Gaussian: final effective-pair F1 (%, mean ± SE). LC uses direct certified edges; CLUB/UniCLUB use component pairs; SCLUB uses set pairs.

Method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
UniCLUB	100.0 ± 0	100.0 ± 0	18.03 ± 0	18.03 ± 0	100.0 ± 0
LC-CLUB-U	100.0 ± 0	100.0 ± 0	76.36 ± 0.03923	30.02 ± 0.04000	100.0 ± 0
LinUCB	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
CLUB	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
SCLUB	0.008080 ± 0.002213	0.002424 ± 0.001616	0.004040 ± 0.001278	0 ± 0	0.003232 ± 0.0008081
LC-CLUB-L	100.0 ± 0	68.21 ± 0.06242	28.69 ± 0.03189	7.168 ± 0	100.0 ± 0
Oracle	100.0 ± 0	100.0 ± 0	100.0 ± 0	100.0 ± 0	100.0 ± 0

Table 9: Constructed / Bernoulli: final effective-pair F1 (%, mean ± SE). LC uses direct certified edges; CLUB/UniCLUB use component pairs; SCLUB uses set pairs.

Method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
UniCLUB	100.0 ± 0	98.17 ± 1.835	18.03 ± 0	18.03 ± 0	100.0 ± 0
LC-CLUB-U	100.0 ± 0	100.0 ± 0	99.66 ± 0.02392	49.53 ± 0.03723	100.0 ± 0
LinUCB	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
CLUB	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
SCLUB	0.1784 ± 0.006428	0.1712 ± 0.01017	0.2631 ± 0.01369	0.3026 ± 0.01671	0.1462 ± 0.009666
LC-CLUB-L	100.0 ± 0	100.0 ± 0	57.41 ± 0.03254	20.87 ± 0.05270	100.0 ± 0
Oracle	100.0 ± 0	100.0 ± 0	100.0 ± 0	100.0 ± 0	100.0 ± 0

Table 10: Augmented sphere / Bernoulli: final effective-pair F1 (%, mean ± SE). LC uses direct certified edges; CLUB/UniCLUB use component pairs; SCLUB uses set pairs.

Method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
UniCLUB	100.0 ± 0	98.17 ± 1.835	18.03 ± 0	18.03 ± 0	100.0 ± 0
LC-CLUB-U	100.0 ± 0	100.0 ± 0	99.66 ± 0.02392	49.53 ± 0.03723	100.0 ± 0
LinUCB	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
CLUB	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
SCLUB	0.6050 ± 0.03374	0.4418 ± 0.01915	0.3920 ± 0.01889	0.3316 ± 0.01091	0.5729 ± 0.02252
LC-CLUB-L	100.0 ± 0	100.0 ± 0	57.41 ± 0.03254	20.87 ± 0.05270	100.0 ± 0
Oracle	100.0 ± 0	100.0 ± 0	100.0 ± 0	100.0 ± 0	100.0 ± 0

Table 11: Native graph-edge F1 (%, mean ± SE), distinct from component-pair F1.

Space and method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
Raw sphere: CLUB	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Raw sphere: UniCLUB	100.0 ± 0	100.0 ± 0	74.88 ± 0.1452	23.82 ± 0.05623	100.0 ± 0
Constructed: CLUB	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Constructed: UniCLUB	100.0 ± 0	100.0 ± 0.0002020	40.34 ± 0.1471	20.57 ± 0.01575	100.0 ± 0
Augmented sphere: CLUB	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Augmented sphere: UniCLUB	100.0 ± 0	100.0 ± 0.0008080	40.31 ± 0.1745	20.57 ± 0.01066	100.0 ± 0

C.4 Graph Recovery and Certified Coverage

For each seed, the true graph contains $49{,}500$ unordered within-cluster pairs. We compute $F_1 = 2\mathrm{TP}/(2\mathrm{TP} + \mathrm{FP} + \mathrm{FN})$ before taking the seed mean and SE. Predicted pairs are direct certified edges for LC, connected-component pairs for CLUB and UniCLUB, and same-set pairs for SCLUB. LinUCB predicts no pairs, and the Oracle predicts exactly the true pairs. Native CLUB/UniCLUB edge F1 is tabulated separately from their component-pair F1.

All final LC graphs have zero false pairs. Under Sine, both LC variants certify all agents and recover the true graph in all five seeds of every space. Under skewed Zipf arrivals, rare agents may remain uncertified: at $\alpha = 1.5$ on the raw sphere, LC-CLUB-L certifies 144.0 agents on average. Table 13 tracks the pair F1 of both LC variants at ten unevenly spaced rounds; it rises with T in every setting.

Table 12: Certified agents at $T = 10^8$ (out of 1000, mean $\pm$ SE). The final LC graphs contain no false pair in any of the 75 setting/seed cases.

Space	Zipf α / Sine	LC-CLUB-U	LC-CLUB-L
Raw sphere	0	1000 ± 0	1000 ± 0
Raw sphere	0.5	1000 ± 0	673.2 ± 0.6633
Raw sphere	1	752.2 ± 0.3742	315.2 ± 0.3742
Raw sphere	1.5	329.4 ± 0.6782	144.0 ± 0
Raw sphere	Sine	1000 ± 0	1000 ± 0
Constructed	0	1000 ± 0	1000 ± 0
Constructed	0.5	1000 ± 0	1000 ± 0
Constructed	1	996.6 ± 0.2449	572.4 ± 0.2449
Constructed	1.5	498.2 ± 0.3742	253.0 ± 0.3162
Constructed	Sine	1000 ± 0	1000 ± 0
Augmented sphere	0	1000 ± 0	1000 ± 0
Augmented sphere	0.5	1000 ± 0	1000 ± 0
Augmented sphere	1	996.6 ± 0.2449	572.4 ± 0.2449
Augmented sphere	1.5	498.2 ± 0.3742	253.0 ± 0.3162
Augmented sphere	Sine	1000 ± 0	1000 ± 0

Table 13: Pair F1 of the certified graph over time (%, five-seed means) at ten unevenly spaced rounds. No cross-cluster pair is certified at any round. The two Bernoulli spaces share arrival paths and admission counts, so their values coincide.

Arrivals	Method	10^5	5×10^5	10^6	5×10^6	10^7	2×10^7	4×10^7	6×10^7	8×10^7	10^8
Raw sphere											
$\alpha = 0$	LC-CLUB-U	0	0	0	0	0	100.0	100.0	100.0	100.0	100.0
	LC-CLUB-L	0	0	0	0	0	0	0	100.0	100.0	100.0
$\alpha = 0.5$	LC-CLUB-U	0	0	0	0.2873	6.197	44.58	95.24	100.0	100.0	100.0
	LC-CLUB-L	0	0	0	0	0.01212	0.4193	8.494	30.45	50.64	68.21
$\alpha = 1$	LC-CLUB-U	0	0.004040	0.02828	1.139	4.881	19.06	40.01	55.60	67.30	76.36
	LC-CLUB-L	0	0	0	0.04443	0.2824	1.332	5.802	12.86	22.08	28.69
$\alpha = 1.5$	LC-CLUB-U	0	0.01212	0.04443	0.6845	1.822	4.734	11.90	19.91	26.94	30.02
	LC-CLUB-L	0	0	0.004040	0.07674	0.2421	0.7487	2.024	3.524	5.252	7.168
Sine	LC-CLUB-U	0	0	0	0	0	100.0	100.0	100.0	100.0	100.0
	LC-CLUB-L	0	0	0	0	0	0	0	100.0	100.0	100.0
Constructed											
$\alpha = 0$	LC-CLUB-U	0	0	0	0	100.0	100.0	100.0	100.0	100.0	100.0
	LC-CLUB-L	0	0	0	0	0	0	100.0	100.0	100.0	100.0
$\alpha = 0.5$	LC-CLUB-U	0	0	0.004040	10.73	55.25	99.82	100.0	100.0	100.0	100.0
	LC-CLUB-L	0	0	0	0.02828	0.8657	14.72	62.11	91.80	100.0	100.0
$\alpha = 1$	LC-CLUB-U	0	0.02828	0.1857	6.630	24.65	46.29	73.22	87.98	96.06	99.66
	LC-CLUB-L	0	0	0.004040	0.4097	1.841	7.765	27.20	37.61	49.40	57.41
$\alpha = 1.5$	LC-CLUB-U	0.004040	0.07674	0.1857	2.218	5.786	14.25	28.76	35.84	43.21	49.53
	LC-CLUB-L	0	0.004040	0.02828	0.3074	0.9386	2.532	6.393	10.94	15.96	20.87
Sine	LC-CLUB-U	0	0	0	17.30	100.0	100.0	100.0	100.0	100.0	100.0
	LC-CLUB-L	0	0	0	0	0	0	100.0	100.0	100.0	100.0
Augmented sphere											
$\alpha = 0$	LC-CLUB-U	0	0	0	0	100.0	100.0	100.0	100.0	100.0	100.0
	LC-CLUB-L	0	0	0	0	0	0	100.0	100.0	100.0	100.0
$\alpha = 0.5$	LC-CLUB-U	0	0	0.004040	10.73	55.25	99.82	100.0	100.0	100.0	100.0
	LC-CLUB-L	0	0	0	0.02828	0.8657	14.72	62.11	91.80	100.0	100.0
$\alpha = 1$	LC-CLUB-U	0	0.02828	0.1857	6.630	24.65	46.29	73.22	87.98	96.06	99.66
	LC-CLUB-L	0	0	0.004040	0.4097	1.841	7.765	27.20	37.61	49.40	57.41
$\alpha = 1.5$	LC-CLUB-U	0.004040	0.07674	0.1857	2.218	5.786	14.25	28.76	35.84	43.21	49.53
	LC-CLUB-L	0	0.004040	0.02828	0.3074	0.9386	2.532	6.393	10.94	15.96	20.87
Sine	LC-CLUB-U	0	0	0	17.30	100.0	100.0	100.0	100.0	100.0	100.0
	LC-CLUB-L	0	0	0	0	0	0	100.0	100.0	100.0	100.0

Table 14: Raw sphere / Gaussian: final-1M pseudo-regret increments, $R(100\mathrm{M}) - R(99\mathrm{M})$, for all seven methods (unscaled, five-seed mean $\pm$ SE). Endpoints are subtracted within each seed before aggregation.

Method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
UniCLUB	608900 ± 131.5	608900 ± 134.0	608600 ± 25.48	608600 ± 44.03	608700 ± 119.5
LC-CLUB-U	3.917 ± 0.2387	4.375 ± 0.3222	17140 ± 63.06	19400 ± 64.44	4.328 ± 0.3170
LinUCB	471.7 ± 4.438	464.2 ± 1.578	475.1 ± 4.032	509.3 ± 5.477	474.6 ± 2.648
CLUB	473.6 ± 4.664	466.4 ± 1.670	476.8 ± 3.609	506.6 ± 4.117	474.5 ± 4.099
SCLUB	471.7 ± 3.356	463.5 ± 2.201	474.7 ± 2.717	512.7 ± 3.915	475.0 ± 3.375
LC-CLUB-L	4.513 ± 0.2473	157.5 ± 1.956	330.9 ± 2.042	448.0 ± 5.231	4.633 ± 0.4541
Oracle	4.485 ± 0.1790	4.602 ± 0.3463	4.837 ± 0.3107	4.748 ± 0.3802	4.834 ± 0.4619

Table 15: Constructed / Bernoulli: final-1M pseudo-regret increments, $R(100\mathrm{M}) - R(99\mathrm{M})$, for all seven methods (unscaled, five-seed mean $\pm$ SE). Endpoints are subtracted within each seed before aggregation.

Method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
UniCLUB	333500 ± 123.6	333600 ± 96.07	333500 ± 113.7	333500 ± 74.20	333700 ± 37.75
LC-CLUB-U	39.89 ± 0.7887	39.65 ± 0.5151	115.7 ± 6.647	4848 ± 19.14	40.23 ± 0.8196
LinUCB	6157 ± 9.614	5899 ± 7.530	4963 ± 10.13	3881 ± 13.19	6145 ± 16.39
CLUB	6139 ± 9.658	5897 ± 18.64	4982 ± 7.031	3874 ± 13.03	6135 ± 11.03
SCLUB	6440 ± 95.92	6064 ± 80.76	5021 ± 103.2	3905 ± 45.09	6220 ± 51.73
LC-CLUB-L	16.03 ± 0.3424	16.28 ± 0.2616	1555 ± 4.101	2248 ± 14.63	17.05 ± 0.5459
Oracle	255.1 ± 2.604	255.0 ± 1.928	243.8 ± 1.302	220.6 ± 1.242	254.3 ± 2.467

Table 16: Augmented sphere / Bernoulli: final-1M pseudo-regret increments, $R(100\mathrm{M}) - R(99\mathrm{M})$, for all seven methods (unscaled, five-seed mean $\pm$ SE). Endpoints are subtracted within each seed before aggregation.

Method	Zipf $\alpha = 0$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	Sine
UniCLUB	301800 ± 125.3	301800 ± 27.72	301700 ± 75.32	301700 ± 55.86	301700 ± 58.90
LC-CLUB-U	1.294 ± 0.05684	1.592 ± 0.05996	75.37 ± 5.987	4347 ± 10.56	1.539 ± 0.1020
LinUCB	169.9 ± 0.6691	180.7 ± 0.9164	240.1 ± 1.955	499.4 ± 3.695	173.7 ± 2.209
CLUB	169.5 ± 0.5946	180.9 ± 0.6038	240.3 ± 2.281	488.9 ± 3.944	171.0 ± 1.984
SCLUB	160.8 ± 1.346	173.8 ± 1.259	228.8 ± 4.573	475.0 ± 6.454	165.4 ± 2.442
LC-CLUB-L	1.409 ± 0.06652	1.601 ± 0.04203	135.1 ± 2.075	454.2 ± 4.032	1.491 ± 0.09421
Oracle	1.247 ± 0.07286	1.635 ± 0.04126	1.510 ± 0.1082	1.452 ± 0.07737	1.551 ± 0.07842

C.5 Complete Curves and Fixed-Window Increments

The cumulative-regret figures compare LinUCB, CLUB, SCLUB and LC-CLUB-L over the full horizon in a single panel per space. Both axes are linear and start at zero; bands show one SE. Tables 14–16 compare all seven methods, including the Oracle, UniCLUB and LC-CLUB-U, using $\Delta R = R(100\mathrm{M}) - R(99\mathrm{M})$. We subtract the exact saved endpoints within each seed before computing the mean and SE. Values are in unscaled regret units. All methods share this 10^6-round window; dividing each increment by 10^6 gives its average regret per round. The Oracle serves as a reference rather than a lower bound.

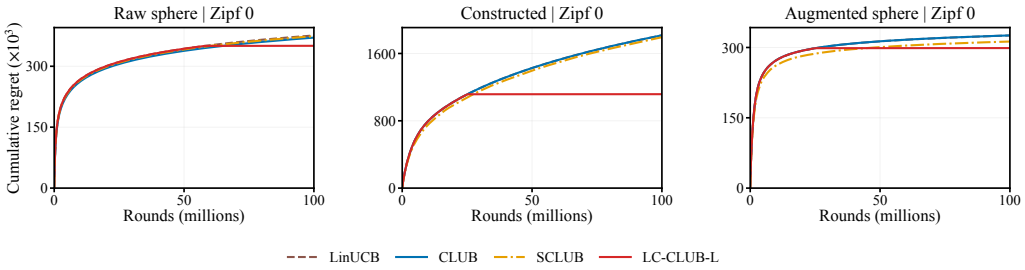

Figure 3: Uniform arrivals (Zipf $\alpha = 0$), all three spaces, five-seed mean $\pm$ SE. Each panel compares LinUCB, CLUB, SCLUB and LC-CLUB-L over the full horizon.

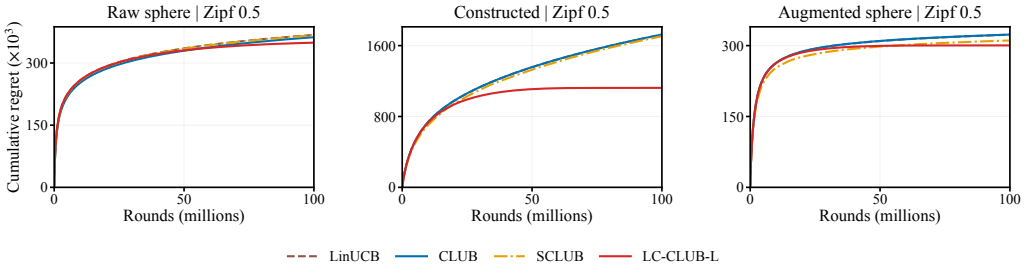

Figure 4: Zipf $\alpha = 0.5$, with the same presentation and complete horizon.

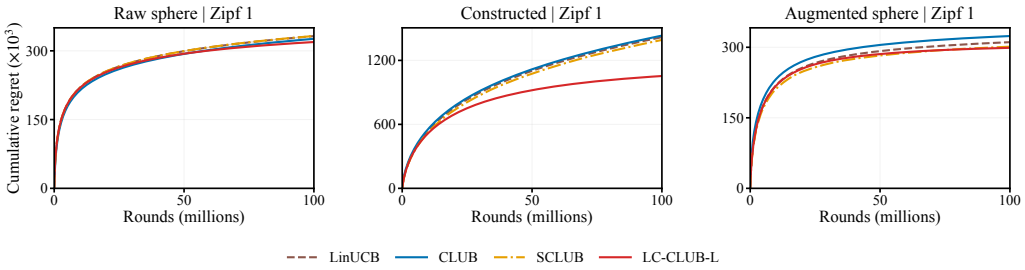

Figure 5: Zipf $\alpha = 1$, with the same presentation and complete horizon.

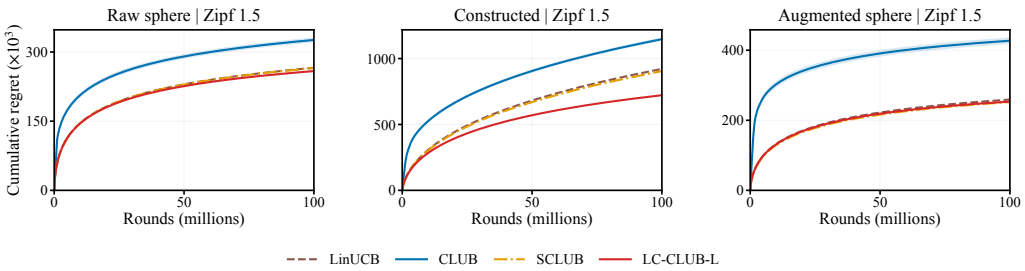

Figure 6: Zipf $\alpha = 1.5$, with the same presentation and complete horizon.

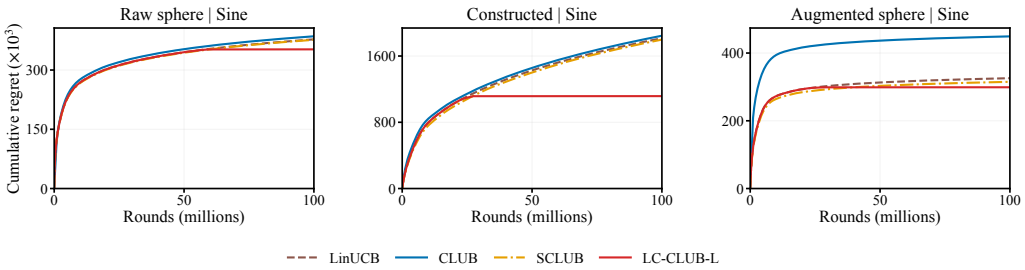

Figure 7: Slow Sine with period 10^7, ten complete periods and five-seed mean $\pm$ SE. Each panel compares LinUCB, CLUB, SCLUB and LC-CLUB-L.