arXiv:submit/6169390 [cs.LG] 31 Jan 2025

Multi-agent Multi-armed Bandit with Fully Heavy-tailed Dynamics

Xingyu Wang

x.wang4@uva.nl
Quantitative Economics, University of Amsterdam
Amsterdam, 1018 WB, NL

Mengfan Xu*

mengfanxu@umass.edu
Mechanical and Industrial Engineering, University of Massachusetts Amherst
Amherst, MA, 01003, USA

January 31, 2025

Abstract

We study decentralized multi-agent multi-armed bandits in fully heavy-tailed settings, where clients communicate over sparse random graphs with heavy-tailed degree distributions and observe heavy-tailed (homogeneous or heterogeneous) reward distributions with potentially infinite variance. The objective is to maximize system performance by pulling the globally optimal arm with the highest global reward mean across all clients. We are the first to address such fully heavy-tailed scenarios, which capture the dynamics and challenges in communication and inference among multiple clients in real-world systems. In homogeneous settings, our algorithmic framework exploits hub-like structures unique to heavy-tailed graphs, allowing clients to aggregate rewards and reduce noises via hub estimators when constructing UCB indices; under M clients and degree distributions with power-law index $\alpha > 1$, our algorithm attains a regret bound (almost) of order $O(M^{1-\frac{1}{\alpha}} \log T)$. Under heterogeneous rewards, clients synchronize by communicating with neighbors, aggregating exchanged estimators in UCB indices; With our newly established information delay bounds on sparse random graphs, we prove a regret bound of $O(M \log T)$. Our results improve upon existing work, which only address time-invariant connected graphs, or light-tailed dynamics in dense graphs and rewards.

1 Introduction

Multi-armed Bandit (MAB) is an online sequential decision-making framework where a decision maker, or a client, pulls an arm from a finite set of arms at each time step, receives the reward of the pulled arm, and aims to maximize the cumulative received reward, or equivalently, minimize the regret compared to always pulling the optimal arm [1, 2]. Recent advancements have focused on its multi-agent variant, namely Multi-agent Multi-armed Bandit (MA-MAB), capturing the complexity of networks of decision makers in real-world scenarios. A notable focus in on cooperative MA-MAB,

*Corresponding Author

where M clients aim to optimize the regret of the entire network via communication, with respect to a globally optimal arm defined by the global reward averaged across all clients. In this work, we consider a general and widely applicable setting with decentralization, where clients communicate on a graph without the stringent assumption of a central server.

Unique to decentralized MA-MAB is the coupling between graphs and time in the sequential regime, where two clients can communicate only when there is an edge between them on the graph. Many studies have focused on time-invariant graphs (i.e., with edges remaining constant over time). However, the emergence of examples such as ad-hoc wireless networks [18] has motivated the recent interest in time-varying graphs. In particular, the random-graph setting allows graphs to be redrawn from a distribution at each time step. For instance, [26] has recently explored the classical light-tailed Erdos-Renyi graphs in MA-MAB, where the edge between each pair of clients is sampled under a Bernoulli distribution with constant probability. Notably, this model is symmetric and dense, with each client having exactly the same expected degree (i.e., the count of connected clients) of order $O(M)$. However, real-world tasks often involve uneven and sparse communications: the degree—representing the communication resource assigned to each client—can be rather heterogeneous among clients, and the resource for each individual may not scale with M. Enabling collaboration over such asymmetric and sparse graphs facilitates system-wide performance in cooperative MA-MAB, which depends on the collective performance of all clients rather than any single individual, and has profound impact on promoting fairness and enhancing social good. This setting, however, remains unexplored and leaves a significant research gap.

This work focuses on sparse random graphs with power-law heavy-tailed degree distributions that capture the highly asymmetric network dynamics in real-world tasks, where a few vertices play hub-like roles with many connections to others, while most vertices have limited degrees. Such power-law heavy tails prevail in real-world networks across numerous contexts, including finance and economics, transportation networks, online retailing, supply chains, social communications, and epidemiology, to name a few; see, e.g., [8, 6, 12, 15, 7, 23, 16]. Therefore, it is essential to address this research gap and build solid theoretical and algorithmic foundations for MA-MAB over sparse and heavy-tailed random graphs, thus enabling efficient and robust decision-making via time-evolving, uneven, and limited communication and coordination over a wide range of real-world networks.

Two aspects of the reward distributions are central to multi-agent MAB problems. First, if the expected reward of an arm is the same for all clients, it is categorized into homogeneous settings; otherwise, it is heterogeneous. While homogeneous MA-MAB is generally well understood, heterogeneous MA-MAB has recently gained attention and presents additional challenges. This is particularly true in the decentralized setting, as inferring the global optimal arm requires the reward information from all clients. While Erdos-Renyi graphs have been addressed in [26], their approach does not apply to our sparse and asymmetric setting. Second, the intensity of randomness (more specifically, tail behaviors) of reward distributions significantly impacts the complexity of the problem. Aside from the long-standing focus on sub-Gaussian rewards in the bandit community, there has been a recent interest in distributions with heavier tails, including the sub-Exponential class and Exponential families [14, 13], and those with even heavier tails and infinite p^{th} order moments [4, 20, 9, 19]. Inferring the mean value using reward observations under such extreme randomness becomes quite challenging, often necessitating the use of robust estimators in the algorithms. Notably, such efforts are mostly limited to single-agent MAB. A recent work [10] addresses heavy-tailed rewards with $(1+\epsilon)^{\text{th}}$-order moments in MA-MAB, but only considers the homogeneous-rewards setting and with time-invariant connected graphs. Considering heavy-tailed rewards in a more challenging and general setting with heterogeneous rewards and time-varying random graphs remains unexplored, a gap we address herein as well

In this paper, we focus on the following question: *Can we formulate and solve the multi-agent multi-armed bandit problem with heavy-tailed random graphs and heavy-tailed rewards in both homogeneous and heterogeneous settings?*

1.1 Main Contributions

We hereby provide an affirmative answer to the research question through our contributions, elaborated as follows. We formulate the multi-agent multi-armed bandit problem over sparse, asymmetric, and heavy-tailed random graphs, and under rewards with potentially infinite variance. Specifically, we consider rank-1 inhomogeneous random graphs [3, 5] with heavy-tailed degree distributions, a standard setting in literature (e.g., [21, 22]). In this framework, the probability for having an edge between a pair of clients at each time step is dictated by (normalization of) attraction weights of clients; under heavy-tailed weight distributions, some clients consistently play hub-like roles and often connect to many other clients. Moreover, the graphs we consider are much more sparse (with $O(1)$ expected degree for each client) compared to Erdos-Renyi graphs (with $O(M)$ expected degree), which translates to significantly reduced communication costs and is of broad interest in large-scale multi-agent learning problems.

Methodologically, we propose new algorithms for homogeneous and heterogeneous settings. In the homogeneous-reward setting, we characterize and exploit the notion of hubs exclusive to heavy-tailed graphs: clients over the hub communicate and aggregate rewards, achieving variance reduction proportional to the hub size, while other clients use delayed aggregation through a hub representative. This principle guides the design of our novel UCB index, which also incorporates the median-of-means estimator for robust estimation under heavy-tailed rewards. In the heterogeneous setting, another challenge is asynchronization (differences in arm pulls) among clients due to variations in their reward distributions. To address this, clients use random sampling when asynchronization occurs, and deploy UCB-based strategies otherwise based on our newly constructed reward estimators. Specifically, we propose an aggregation method that integrates the most recent heavy-tailed reward information from all clients, introducing novel information update mechanisms.

We establish theoretical guarantees for the proposed algorithms through comprehensive regret analyses. In homogeneous settings, we obtain a regret upper bound that is (almost) of order $O(M^{1-\frac{1}{\alpha}}\log T)$, which is sublinear in M. This improves upon the potential linearity in [10] and demonstrates sample complexity reduction even under sparse graph structures. Under heterogeneous rewards, we derive an upper bound of order $O(M\log T)$, extending the bound in [28] to sparse and asymmetric graphs with heavy-tailed rewards. The results highlight the consistency, effectiveness, and robustness of our approach.

The paper is organized as follows. Section 2 sets notations and formulates the problem. Section 3 explores properties that are exclusive to sparse, heavy-tailed graphs and useful for our MA-MAB setting. Then, we propose algorithms and conduct regret analyses for the homogeneous and heterogeneous settings in Sections 4 and 5, respectively. Lastly, we conclude the paper and discuss future work in Section 6.

2 Problem Formulation

We start with notations used throughout this paper. Given a positive integer k, let $[k]=\{1,2,\ldots,k\}$. We adopt the convention that $[0]=\emptyset$. Given two sequences of non-negative real numbers $(x_n)_{n\geq 1}$ and $(y_n)_{n\geq 1}$, we say that $x_n = O(y_n)$ (as $n\to\infty$) if there exists some $C\in[0,\infty)$ such that $x_n \leq Cy_n\ \forall n\geq 1$. Besides, we say that $x_n = o(y_n)$ if $\lim_{n\to\infty} x_n/y_n = 0$.

Let M denote the number of clients, which are labeled by $[M]=\{1,2,\ldots,M\}$. At each time $t=1,2,\ldots$, the clients are distributed over an undirected graph $G_t=(V,E_t)$, where $V=[M]$, and E_t is the set of edges (i.e. two clients communicate at time t only if $(i,j)\in E_t$) generated by the following distribution: independently for each pair, $(i,j)\in E_t$ with probability $P(h_i,h_j)$, and $(i,j)\notin E_t$ with probability $1-P(h_i,h_j)$; where

$$P(u,v)=\min\{1,\ uv/(\theta M)\}\qquad \forall u,v\geq 0, \tag{2.1}$$

$(h_i)_{i\geq 1}$ are independent copies of a positive random variable h, and $\theta=\mathbf{E}h$. Note that this model is the standard rank-1 inhomogeneous random graphs; e.g., [3, 5]. Intuitively speaking, h_i is the weight

assigned to the i^{th} node, representing its *attraction* to the other nodes. The weight h_i does not change with time t, and is close to the expected degree of the i^{th} node (especially under large M) over the graphs G_t's. We use $\mathcal{N}_i(t)$ to denote the neighborhood set of the i^{th} node at time t, which includes all the other nodes that are connected to i over the graph G_t. Equivalently, the graph G_t can be represented by the adjacency matrix $(X^t_{i,j})_{1\le i,j\le M}$ where $X^t_{i,j}=1$ if there is an edge between nodes i and j, and $X^t_{i,j}=0$ otherwise. We set $X^t_{i,i}\equiv 1$ for any $1\le i\le M$. We also define the empirical adjacency matrix by $P_t(i,j)\triangleq\frac{\sum_{s=1}^t X^s_{i,j}}{t}$ and $P_t=\big(P_t(i,j)\big)_{1\le i,j\le M}$. We note that each node m only knows its own neighbors and can only observe the m-th row of P_t, i.e., node m has access to $\{P_t(m,j)\}_j$ but not $\{P_t(k,j)\}_j$ for $k\neq m$.

Power-law heavy tails are typically captured through the notion of regular variation. Given a measurable function $\phi:(0,\infty)\to(0,\infty)$, we say that ϕ is regularly varying as $x\to\infty$ with index β (denoted as $\phi(x)\in\mathcal{RV}_\beta(x)$ as $x\to\infty$) if $\phi(x)=x^\beta\cdot l(x)$ for some function $l:(0,\infty)\to(0,\infty)$ with $\lim_{x\to\infty} l(tx)/l(x)=1$ for all $t>0$. That is, $\phi(x)$ roughly follows a power-law tail with index β. For a standard treatment on the properties of regularly varying functions, see, e.g., [17]. The next assumption specifies the law of weights h_i and the choice of θ in (2.1).

Assumption 1 (Heavy-Tailed Graph)**.** *The sequence $(h_i)_{i\ge1}$ are iid copies of h, with law $\mathbf{P}(h>x)\in\mathcal{RV}_{-\alpha}(x)$ for some $\alpha>1$, and $\theta=\mathbf{E}h$ in (2.1).*

We impose the next assumption to exclude the pathological case that some clients are (almost) never connected to others at any time t.

Assumption 2 (Lower Bound for h)**.** *There exists $c_h>0$ such that $\mathbf{P}(h\ge c_h)=1$.*

We add two comments for Assumptions 1 and 2: (1) Assumption 2 does not prevent isolated clients in G_t; in fact, given time $t\ge1$ and client i, the probability that i is not connected to any others over G_t equals $\prod_{j\in[M]:\ j\neq i}\big[1-\mathbf{E}P(h_i,h_j)\big]$, which is strictly positive; (2) the graph G_t is both *heavy-tailed* and *sparse*; in particular, the heavy-tailedness in h implies that that the client with the highest weight h_i can have degree of order M^α; however, since the law of h does not vary with M, the expectation of the degree for each client is only $O(1)$ (i.e., most nodes are only connected to a small number of nodes).

Let K be the number of arms. For each client $i\in[M]$, we denote the reward of arm $1\le k\le K$ at time t by $r^i_k(t)$, which is an i.i.d. sequence from a time-invariant distribution with mean value μ^i_k. When $\mu^i_k=\mu^j_k\ \forall i,j\in[M]$ holds for any arm k, it is referred to as a homogeneous-reward setting; otherwise it is heterogeneous. Our Assumption 3 is sufficiently general to account for both heavy-tailed reward distributions (potentially with infinite variance) and light-tailed distributions (with finite moments of any order, e.g. sub-Gaussian and sub-exponential classes).

Assumption 3 (Rewards with Uniformly Bounded $(1+\epsilon)^{\text{th}}$ Central Moments)**.** *Given $i\in[M]$ and $k\in[K]$, rewards $(r^k_i(t))_{t\ge1}$ are iid copies from the distribution F^k_i. Furthermore, there exist $\epsilon\in(0,1]$ and $\rho\in(0,\infty)$ such that $\sup_{i\in[M],\ k\in[K]}\mathbf{E}|r^k_i(1)-\mu^k_i|^{1+\epsilon}\le\rho$, where we use $\mu^k_i\triangleq\mathbf{E}r^k_i(1)=\int xF^k_i(dx)$ to denote expected rewards.*

We use a^t_m to denote the arm pulled by client m at time t. We define the global reward of arm i at each time step t as $r_i(t)=\frac{1}{M}\sum_{m=1}^M r^m_i(t)$, and the expected value of the global reward of arm i by $\mu_i=\frac{1}{M}\sum_{m=1}^M\mu^m_i$. We denote the *global optimal arm* by $i^*=\arg\max_i\mu_i$, and consider the cooperative setting where all clients would, ideally, pull the globally optimal arm i^*. The optimality gap for arm i is $\Delta_i=\mu_{i^*}-\mu_i$. This motivates the definition of the global regret by $R_T=T\cdot\mu_{i^*}-\frac{1}{M}\sum_{t=1}^T\sum_{m=1}^M\mu_{a^t_m}$, which measures the difference in the cumulative expected reward between the global optimal arm and the action sequence. The main objective of this paper is to develop a multi-agent MAB algorithm and minimize R_T for clients given the sparse communications available on $(G_t)_{t\ge1}$.

3 Analyses on Random Graphs

In this section, we establish useful properties for the hub structures and information delay over random graphs G_t introduced in Section 2. The results lay the foundation for our subsequent analysis of multi-agent multi-armed bandits. We collect the proofs in the Appendix.

3.1 Hub on Heavy-Tailed Graphs

A feature exclusive to heavy-tailed graphs is the arise of hub-like nodes with disproportionally large degrees (i.e., being connected to a large number of nodes), which enables efficient communication among all clients through hubs. We first consider a *deterministic* characterization of the hub. Let $\hat{i}$ be the client with the highest degree at time 1 (arbitrarily pick one if there are ties). Let $S_0^t \triangleq \{i \in [M] : (i, \hat{i}) \in E_t\}$ be the clients communicating with $\hat{i}$ at time t. Note that, for any $i \in [M]$ with $h_i > \theta M / h_{\hat{i}}$, by (2.1) we know that such i must be *deterministically* (i.e., with probability 1) connected to $\hat{i}$ for all t. Lemma 3.1 confirms that, with high probability, such nodes i are plenty. Specifically, by standard techniques in extreme value theory for heavy-tailed variables, one can show that $h_{\hat{i}}$ is roughly of order $M^{1/\alpha}$, and (under $\alpha \in (1,2)$) the count of $i \in [M]$ with $h_i > \theta M / h_{\hat{i}} \approx \theta M^{1-\frac{1}{\alpha}}$ is roughly of order $M^{2-\alpha}$. By taking the ζ-slackness in Lemma 3.1, we are able to ensure the exponentially decaying bound for pathological cases.

Lemma 3.1. *Let Assumptions 1 and 2 hold with $\alpha \in (1,2)$. Given $\zeta \in (0, 2-\alpha)$, there exists $\gamma > 0$ such that*

$$\mathbf{P}(|S_0| \le M^{2-\alpha-\zeta}) = o(\exp(-M^{\gamma})),$$

where $S_0 = \cap_{t \ge 1} S_0^t$.

In fact, we can further improve upon Lemma 3.1 by considering the following *stochastic* characterization of hubs. Given $\zeta > 0$, let $\tau(t) \triangleq \max\{u \le t : \ |S_0^u| > M^{\frac{1}{\alpha}-\zeta}\}$ be the last time the hub is large (w.r.t. threshold $M^{\frac{1}{\alpha}-\zeta}$) up until time t, under the convention that $\tau(t) = 0$ when taking maximum over empty sets. Lemma 3.2 bounds the time gap between the emergence of large hubs. The proof builds upon extreme value theory and a straightforward bound of $\sup_{t \le T} t - \tau(t)$ using geometric random variables.

Lemma 3.2. *Let Assumptions 1 and 2 hold. Define event $A_{\alpha,\zeta} \triangleq \{h_{\hat{i}} \ge M^{\frac{1}{\alpha}-\frac{\zeta}{2}}\}$. Let $\zeta \in (0, 1-\frac{1}{\alpha})$. There exists $\gamma > 0$ such that*

$$\mathbf{P}\big((A_{\alpha,\zeta})^c\big) = o\big(\exp(-M^{\gamma})\big).$$

Furthermore, there exists $M_0 > 0$ such that for any $M \ge M_0$ and $T \ge 1$,

$$\mathbf{P}\left(\sup_{t \le T} t - \tau(t) > \log T \ \middle| \ A_{\alpha,\zeta}\right) \le \frac{1}{MT}.$$

Comparison to dense and light-tailed graphs [26] considered dense E-R graphs, where any pair of clients connects on a regular basis. In this work, we show that under the presence of heavy tails in degree distributions, clients can afford to collaborate over much sparse communication by sending messages to and receiving messages from the hub center $\hat{i}$ that integrates all information.

3.2 Information Delay over Sparse Graphs

The sparsity of graphs G_t makes existing analysis in [26] largely incompatible, and requires new approach to obtain detailed bounds regarding the information delay under sparse communication. Specifically, given some non-empty subset of clients $S \subseteq [M]$, let $\bar{S}^0 = S$, and $\bar{S}^t \triangleq \{i \in [M] : \ i \in \bar{S}^{t-1}$; or $\exists j \in \bar{S}^{t-1}$ s.t. $(i,j) \in E_t\}$ for each $t \ge 1$. That is, if all clients in S send a piece of message

at time 1, which will passed to neighbors over graph G_t at each time t, then $\bar{S}^t$ is the collection of clients that have received the message at time t. Lemma 3.3 shows that, with high probability, the information delay uniformly for any client $i \in [M]$ is at most $O\big((\log M)^2\big)$. Our proof strategy is to establish a coupling between the sequence of graphs $(G_t)_{t\geq 1}$ and a branching process, whose size grows geometrically fast in expectation.

Lemma 3.3. *Under Assumption 2, there exists $\kappa \in (0,\infty)$ such that*

$$\mathbf{P}\big(j \notin \bar{S}^{\gamma\cdot\kappa\cdot(\log M)^2} \text{ for some } j \in [M]\big) \leq M^{-\gamma}$$

holds for any $\gamma > 0$, $M \geq 1$, and any non-emtpy $S \subseteq [M]$.

4 Homogeneous Rewards

This section considers the homogeneous-reward setting. The algorithmic framework will be extended to the heterogeneous-reward setting in the next section. Specifically, we propose the algorithm in Section 4.1, addressing the challenges from both heavy-tailed rewards and sparse graphs. Then, we establish the theoretical effectiveness of the proposed algorithm in Section 4.2.

4.1 Algorithm

Under homogeneous rewards, we propose a new algorithm called HT-HMUCB (**H**eavy-**T**ailed **H**o**M**ogeneous **U**pper **C**onfidence **B**ounds); pseudocode is provided below. The algorithm consists of several stages in the following order.

Hub identification. A novel step in our algorithm concerns the identification of hub center $\hat{i}$, i.e., the client with the highest degree at time $t = 1$. Specifically, the degree information of all clients at time 1 will be passed over G_t at each step t so that, with high probability, each client $i \in [M]$ is able to tell whether itself is the hub center (i.e., $i = \hat{i}$) or not after $O\big((\log M)^2\big)$ steps; see Algorithm 4 (Appendix). Afterwards, all non-center clients will follow the reward information processed by the hub center $\hat{i}$.

Arm selection. During this stage, the clients decide which arm to pull by executing a UCB-based strategy, where each arm i is assigned a UCB index, formally expressed as $\hat{\mu}_i^m(t) + \rho^{\frac{1}{1+\epsilon}}\big(\frac{c\log(t)}{N_{m,i}(t)}\big)^{\frac{\epsilon}{1+\epsilon}}$. Here, $\hat{\mu}_i^m(t)$ and $N_{m,i}(t)$ represent the global reward estimators and sample counts of arm i by client m, respectively, defined in Rule 1 below. Constants ρ and ϵ are characterized in Assumption 3, and c is specified below in Theorem 4.1.

Transmission. We define an information filtration $\mathcal{F}_m(t)$ as the information available to m up to time t, which reads as $\mathcal{F}_m(0) = \{(m,1), r_i^m(1), N_{m,i}(1), \hat{\mu}_i^m(1)\}, \mathcal{F}_m(t) = \cup_{j\in\mathcal{N}_m(t-1)}\mathcal{F}_j(t-1)$. Each client m communicates with its neighbors $\mathcal{N}_m(t)$ by sending an message, composed of $(m,t), r_i^m(t)$, $N_{m,i}(t)$, $\hat{\mu}_i^m(t)$, and $\mathcal{F}_m(t)$, while collecting messages from its neighbors.

Information update. After pulling arms and receiving feedback from the environment, as well as information from others, the clients proceed to update their information based on **Rule 1**, which is detailed below:
1) Local estimation

$$t_{m,j} = \max\{s \leq t : \ (j,s) \in \mathcal{F}_m(t)\}$$
$$\text{local sample counts: } n_{m,i}(t+1) = n_{m,i}(t) + \mathbb{1}_{a_m^t = i},$$

2) Global estimation

$$\text{Sample counts: } N_{m,i}(t+1) = \sum_{m \in \mathcal{N}_m(t)} n_{m,i}(t+1)$$

$$\text{Estimator: if } m \text{ is center (i.e., } m = \hat{i}\text{)},$$

$$\hat{\mu}_i^{\hat{i}}(t+1) = MoM_B(\{r_i^j(s) : r_i^j(s) \in \mathcal{F}_m(t)\})$$

$$\text{Estimator: if } m \text{ is not center, } \hat{\mu}_i^m(t+1) = \hat{\mu}_i^{\hat{i}}(\max_{j \in S_0} t_{m,j})$$

Here, $MoM_B\big((X_i)_{i\in[n]}\big)$ denote the **median of mean** estimator with B batches, i.e., the meadian of the estimators $\hat{\mu}_1, \ldots, \hat{\mu}_B$ defined by $\hat{\mu}_j = \frac{1}{N}\sum_{t=(j-1)N+1}^{jN} X_t$ with $N = \lfloor n/B \rfloor$. MoM estimators have been applied in [4] for UBC algorithms in single-agent settings. Our work further demonstrates its use for robust estimation under heavy tails in multi-agent MAB problems.

Comparison with prior works Our algorithm differs from the existing algorithms in [10] for homogeneous settings with heavy-tailed rewards in the following ways: 1) the clients identify the hub center to maximize information efficiency, which is computationally more tractable than the clique search in [10]; and 2) the clients rely solely on information from hub center, rather than maintaining global estimators individually, to reduce noise in estimation.

Algorithm 1 HT-HMUCB (Heavy-tailed Homogeneous UCB)

Initialization: For each client m and arm $i \in \{1, 2, \ldots, K\}$, we set $N_{m,i}(L+1) = n_{m,i}(L)$; all other values at $L+1$ are initialized as 0

for $t = 1, 2, \ldots, L$ **do**
  Indentify hub center $\hat{i}$ by running Algo. 4; `// Hub`
end
for $t = L+1, L+2, \ldots, T$ **do**
  for *each client* m **do** `// UCB`
    $a_m^t = \arg\max_i \hat{\mu}_i^m(t) + \rho^{\frac{1}{1+\epsilon}}\big(\frac{c\log(t)}{N_{m,i}(t)}\big)^{\frac{\epsilon}{1+\epsilon}}$
    Pull arm a_m^t and receive reward $r_{a_m^t}^m(t)$
  end
  The environment generates the graph G_t; `// Env`
  Each client m sends (m,t), $r_i^m(t)$, $N_{j,i}(t)$, $\tilde{\mu}_i^m(t)$, $\mathcal{F}_m(t)$ to each client in $\mathcal{N}_m(t)$ `// Transmission`
  for *each client* m **do**
    for $i = 1, \ldots, K$ **do**
      Update $\bar{\mu}_i^m(t), n_{m,i}(t), N_{m,i}(t)$ and $\tilde{\mu}_i^m(t)$ based on Rule 1 `// Update`
    end
  end
end

4.2 Regret Analyses

In this section, we demonstrate the effectiveness of the proposed algorithm through regret analyses. Notably, the tail index α for degree distributions in Assumption 1 plays a key role in our regret bound. First, we establish Theorem 4.1 for $\alpha \in (1, 2)$.

Theorem 4.1 ($\alpha \in (1,2)$)**.** *Let Assumptions 1–3 hold with $1 < \alpha < 2$. Let Algorithm 1 run under Rule 1. Then, given $\zeta \in (0, 2-\alpha)$, there exists $\eta > 0$ such that, for any T and M, the event $A_{\zeta,\delta}$*

holds with probability at least $(1-\frac{2\eta}{M}-\frac{\eta}{TM})$*, we have*

$$\mathbf{E}[R_T|A_{\zeta,\delta}] \le L + M\sum_i \left(\frac{2c\Delta_i \log T}{M^{2-\alpha-\zeta}\cdot(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} + \frac{\pi^2}{3}\Delta_i\right)$$
$$= O\Big((1+2M^{\alpha-1+\zeta})\cdot\rho^{\frac{1}{\epsilon}}\sum_{i\in[K]}\Delta_i^{-\epsilon}\cdot\log T\Big),$$

where $L = 2\kappa(\log M)^2\log T$*,* κ *is the constant characterized in Lemma 3.3,* $c = (16\log 2e^{1/8})^{\frac{\epsilon}{1+\epsilon}}$*,* $C = (12)^{\frac{1}{1+\epsilon}}$*, we set* $B = 8\log(e^{1/8}T)$ *for the count of batches in MoM estimators,* $|S_0|$ *denotes the size of* S_0 *(see Lemma 3.1), and the event is defined by* $A_{\zeta,\delta} = A^1_{\zeta,\delta}\cap A^2_{\zeta,\delta}\cap A^3_{\zeta,\delta}$ *with* $A^1_{\zeta,\delta} = \{|S_0| \ge M^{2-\alpha-\zeta}\}$*,* $A^2_{\zeta,\delta} = \{n_{m,i}(t_{m,j}) \ge n_{m,i}(t) - \kappa\log M\log T\ \ \forall t \le T,\ \ \forall m,i,j\}$*, and* $A^3_{\zeta,\delta} = \{\hat{i}(m) \ne \hat{i}$ *for some* $m\in[M]\}$ *(see Algo. 4). In particular,* $\mathbf{E}[R_T|A_{\zeta,\delta}] \le O(M^{\alpha-1+\zeta}\cdot\log T) = o(M)\cdot O(\log T)$.

Proof Sketch. The proof hinges on the key observation that, with high probability, $|S_0|$ is at least $M^{2-\alpha-\zeta}$; see Lemma 3.1. The communication delay between the clients is then bounded by Lemma 3.3 (see also Lemma B.8), hence the clients in the hub enjoy the reduction in sample complexity. This is achieved by utilizing a concentration inequality with respect to $\sum_{m\in S_0} n_{m,i}(t)$ instead of $n_{m,i}(t)$, resulting in an individual regret of order $\frac{1}{|S_0|}\cdot\log T$. Consequently, the total regret is of order $\frac{M-|S_0|}{|S_0|}\cdot\log T$. The full proof is provided in the Appendix. □

Remark 1 (Expected regret)**.** *We emphasize that the expected regret* $\mathbf{E}[R_T]$ *is upper bounded by* $\mathbf{E}[R_T] = \mathbf{E}[R_T|A_{\zeta,\delta}]\mathbf{P}(A_{\zeta,\delta}) + \mathbf{E}[R_T|(A_{\zeta,\delta})^c]\mathbf{P}((A_{\zeta,\delta})^c) \le O(M^{\alpha-1+\zeta}\cdot\log T) + O(M\log T)\cdot(\frac{2\eta}{M}+\frac{\eta}{TM}) = O(M^{\alpha-1+\zeta}\cdot\log T)$*. Here, the inequality follows from Theorem 4.1, as well as the observation that, in the pathological case where the "good" event* $A_{\zeta,\delta}$ *does not happen, the regret would still be no worse than the total regret of multiple single-agent bandits, which results in a regret of order* $O(M\log T)$.

We stress that the regret bound in Theorem 4.1 depends on α (see Assumption 1), which characterizes the relationship between the regret and the heavy-tailed graph dynamics, and reflects the reduction of complexity that is unique to our setting. Additionally, the regret bound depends on ρ and ϵ (see Assumption 3), capturing the influence of the heavy-tailed rewards. The dependency on optimality gaps Δ_i's is standard in MAB, but in our case there is an additional ϵ-polynomial factor due to heavy-tailed rewards.

In fact, we can obtain even stronger regret bound—and even without the requirement of $\alpha\in(1,2)$ in Theorem 4.1—by considering the the stochastic characterization of hubs S_0^t. This is demonstrated in Theorem 4.2.

Theorem 4.2 ($\alpha > 1$)**.** *Let Assumptions 1–3 hold. Let Algorithm 1 run under Rule 1. Then, given* $\zeta\in(0,2-\alpha)$ *there exists* $\eta > 0$ *such that, for any* T *and* M*, the event* $A_{\alpha,\delta,\zeta}$ *holds with probability at least* $(1-\frac{\eta}{M}-\frac{\eta}{TM})$*, and*

$$\mathbf{E}[R_T|A_{\alpha,\delta,\zeta}] \le L + M\sum_i(\frac{2c\log T}{M^{\frac{1}{\alpha}-\zeta}(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} + \frac{\pi^2}{3})\Delta_i$$
$$= O\Big(M^{1-\frac{1}{\alpha}+\zeta}\cdot\log T\Big),$$

where $A_{\alpha,\delta,\zeta} = A_{\zeta,\delta}\cap A_{\alpha,\zeta}$*, with event* $A_{\alpha,\zeta}$ *defined in Lemma 3.2 and event* $A_{\zeta,\delta}$ *defined in Theorem 4.1, and parameters* L,κ,c,C,B *are specified as in Theorem 4.1.*

Proof Sketch. The information delay characterized in Lemma 3.3 (and hence Lemma B.8) for sparse graphs still holds here. Meanwhile, Lemma 3.2 proves that hub size $|S_0^t|$, despite being time-varying, is often times large (of order $M^{\frac{1}{\alpha}-\frac{\zeta}{2}}$). This tighter lower bound for hub size leads to further variance reduction: as clients can efficiently collect information sent by (often times) even larger S_0^t, and thereby minimize regret. The detailed proof is in Appendix. □

Remark 2 (Expected regret). *Analogous to Remark 1, we can derive an upper bound on* $\mathbf{E}[R_T]$*, which has the same order as the high-probability bound above, i.e.,* $\mathbf{E}[R_T] \leq O(M^{1-\frac{1}{\alpha}+\zeta}\log T)$.

Remark 3 (Comparison to Theorem 4.1). *Given* $\alpha > 1$*, the index* $1-\frac{1}{\alpha}$ *for regret bound in Theorem 4.2 is always smaller than the index* $2-\alpha$ *in Theorem 4.1, due to the preliminary inequality* $-\frac{1}{x} \leq 1-x$ *for any* $x > 1$*. This implies that by considering a dynamic, time-dependent hub, the coefficient in our regret bound is upper bounded by a smaller constant compared to Theorem 4.1. This highlights the advantage of leveraging the tighter characterization for notion of time-varying hub as in Lemma 3.2 and the significant contributions of this work, as we obtain a regret upper bound of even smaller order and under more relaxed assumptions (i.e., without the requirement of* $\alpha < 2$*).*

Comparison with existing literature on homogeneous MA-MAB We begin the comparison by focusing on the perspective of regret bound and emphasizing the order of M, as naive UCB already leads to a regret of order $\log T$. First, the regret bound in [10] is $R_T \leq O(\alpha(G)\rho^{\frac{1}{\epsilon}}(\sum_i \Delta_i^{-\epsilon})\log T)$, and their algorithm relies on solving an NP-hard problem to find the clique (i.e., the largest independent set). Also, the quantity $\alpha(G)$—the independence number of graph G—may not have an explicit form and has been an active research topic. For a connected graph, some known bounds on $\alpha(G)$ are [25]: $\alpha(G) \leq M - \frac{M-1}{\Delta}$, where Δ is the maximum degree, and $\alpha(G) \geq \frac{M}{1+\Delta}$, implying that $R_T \leq O(M \cdot \log T)$. In contrast, we obtain an improved regret bound that is sub-linear in M. Specifically, since the slackness parameter ζ in Theorem 4.2 can be set arbitrarily close to 0, our regret bound is almost of order $O(M^{1-\frac{1}{\alpha}+\zeta})$, which improves upon the regret bound $O(M\log T)$ of executing individual bandits without communication in homogeneous MAB (i.e., single-player bandit). In particular, our bound is established for sparse graphs with total degree (i.e., count of edges over the graph) of order $O(M)$. While [29] establishes a regret bound of order $O(\log T)$ independent of M, the authors assume that the graph is connected or complete, which can have $O(M^2)$ total degree, and they only consider sub-Gaussian rewards. Their assumptions allow the use of arm elimination instead of UCB, which is largely different from our problem setting.

Regarding assumptions, we 1) do not require the graph to be connected (or l-periodically connected, as in [30]); 2) allow the graph to change over time; and 3) address sparse graphs. Points 1) and 3) address limitations present in almost all existing work on multi-agent multi-armed bandit problems, to the best of our knowledge, while point 2) resolves an open problem identified in [10], which suggested time-varying network analysis and tested this case numerically. Our heavy-tailed reward assumption is in the same spirit as that in [10]. To the best of our knowledge, our framework, the most general to date in homogeneous MA-MAB, relaxes these widely used assumptions and thus opens new research directions.

5 Heterogeneous Rewards

This section addresses the more general heterogeneous setting. Specifically, we present the algorithm in Section 5.1 and the regret analysis in Section 5.2. The results are well beyond the scope of the existing work on MA-MAB with random graphs and heterogeneous rewards [26], which only considered limited to Erdős–Rényi graphs with light-tailed, dense dynamics.

5.1 Algorithm

Under heterogeneous rewards, we propose a new algorithm, namely HT-HTUCB (**H**eavy-**T**ailed **H**e**T**erogeneous UCB), as the presence of heterogeneity in rewards necessitates different approaches to informative communication and information updates across clients. The pseudo-code is provided in Algorithm 3 (Appendix) and Algorithm 2 for the burn-in and learning stages, respectively.

Algorithm 3 (Appendix) for the burn-in period is designed to accumulate local reward and graph information, and is identical to that in [26] so we defer the details to Appendix. This prepares

Algorithm 2 HT-HTUCB (Heavy-Tailed Heterogeneous UCB): Learning period

Initialization: For each client m and arm $i \in \{1, 2, \ldots, K\}$, we have $\tilde{\mu}_i^m(L+1)$, $N_{m,i}(L+1) = n_{m,i}(L)$; all other values at $L+1$ are initialized as 0

for $t = L+1, L+2, \ldots, T$ **do**

  for *each client* m **do** // UCB

    if *there is no arm* i *such that* $n_{m,i}(t) \leq N_{m,i}(t) - 2\kappa(\log M)^2 \log T$ **then**

      $a_m^t = \arg\max_i \hat{\mu}_i^m(t) + \rho^{\frac{1}{1+\epsilon}} \left(\frac{c \log(t)}{N_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}$

    else

      Randomly sample an arm $a_m^t = t \mod K$

    end

    Pull arm a_m^t and receive reward $r_{a_m^t}^m(t)$

  end

  The environment generates the graph $G_t = (V, E_t)$; // Env

  Each client m sends (m,t), $r_i^m(t)$, $N_{j,i}(t)$, $\bar{\mu}_i^m(t)$, $\tilde{\mu}_i^m(t)$, $\mathcal{F}_m(t)$ to $\mathcal{N}_m(t)$ // Transmission

  for *each client* m **do**

    for $i = 1, \ldots, K$ **do**

      Update $n_{m,i}(t), N_{m,i}(t)$ and $\tilde{\mu}_i^m(t)$ based on Rule 2

    end

  end

end

the clients with initial reward and graph estimators, enabling them to communicate and integrate information in the subsequent learning stage. Specifically, during the burn-in period, clients pull each arm $1 \leq i \leq K$ sequentially and update the average reward of each arm as local reward estimators. Simultaneously, clients observe the graph, updating the edge frequency and average degree of clients. At the end of the burn-in period, the clients output an initial global estimator, calculated as the weighted average of the local reward estimators, using the edge frequencies as weights.

Moving onto the learning period, clients employ UCB-based strategies to pull arms, communicate with each other, and integrate the information collected from neighbors. The steps are executed in the following order.

Arm Selection. We still employ a UCB-based strategy, but the global estimator is constructed differently, with an additional condition during the execution of the UCB-based strategy: $n_{m,i}(t) \leq N_{m,i}(t) - 2\kappa(\log M)^2 \log T$. This novel condition ensures that clients remain synchronized and accounts for the longer information delay caused by heavy-tailed graph dynamics, which differs from the setting in [26, 30].

Transmission. This step is almost identical to that of Section 4.1, except for the message components. Here, an information filtration $\mathcal{F}_m(t)$ reads as $\mathcal{F}_m(0) = \{(m,1), r_i^m(1), N_{m,i}(1), \bar{\mu}_i^m(1), \tilde{\mu}_i^m(1)\}$ and $\mathcal{F}_m(t) = \cup_{j \in \mathcal{N}_m(t-1)} \mathcal{F}_j(t-1)$. Each client m communicates with its neighbors $\mathcal{N}_m(t)$ by sending an message, including $(m,t), r_i^m(t)$, $N_{m,i}(t)$, $\bar{\mu}_i^m(t)$, $\tilde{\mu}_i^m(t)$ and $\mathcal{F}_m(t)$, while collecting messages from its neighbors.

Information update. Since the heterogeneity in rewards necessitates obtaining reward information from all clients, we propose a new information update step to aggregate information, represented by **Rule 2**:

1) Local estimation

$$\text{local sample counts: } n_{m,i}(t+1) = n_{m,i}(t) + \mathbb{1}_{a_m^t = i},$$

$$\text{Local estimator: } \bar{\mu}_i^m(t+1) = MoM_B(\{r_i^m(s)\}_{1\le s\le t})$$

2) Global estimation

$$\begin{aligned}
&t_{m,j} = \max\{s \le t : \ (j,s) \in \mathcal{F}_m(t)\}\\
&\hat{N}_{m,i}(t+1) = \max\{n_{m,i}(t+1), \{N_{j,i}(t)\}_{j\in\mathcal{N}_m(t)}\}\\
&\tilde{\mu}_i^m(t+1) = \textstyle\sum_{j=1}^M P_t' \tilde{\mu}_{i,j}^m(t_{m,j}) + d_{m,t}\sum_j \hat{\mu}_{i,j}^m(t_{m,j})\\
&P_t' = \frac{(N - M2^{\frac{1}{(\epsilon+1)}})}{MN2^{\frac{1}{\epsilon+1}}}, N = (12^{\frac{1}{1+\epsilon}})^{\frac{1+\epsilon}{\epsilon}} + 1\\
&d_{m,t} = \frac{(1 - \sum_{j=1}^M P_t')}{M}
\end{aligned}$$

Comparison with existing work and Section 4.1 Compared to Section 4.1, the arm selection step imposes an extra requirement that $n_{m,i}(t) \le N_{m,i}(t) - 2\kappa(\log M)^2 \log T$ in UCB, which balances exploitation and exploration given the noise in the global reward estimator $\hat{\mu}$ in the heterogeneous case. Secondly, we remove the hub estimation step in the heterogeneous setting, because each client must collect information from all clients by message passing with through neighbors sets $\mathcal{N}_m(t)$. Lastly, we run Rule 2 instead of Rule 1 for information update.

Compared to [26], our UCB index $\hat{\mu}_i^m(t) + \rho^{\frac{1}{1+\epsilon}}(\frac{c\log(t)}{N_{m,i}(t)})^{\frac{\epsilon}{1+\epsilon}}$ address (potentially) heavy-tailed rewards, while [26] considers $\hat{\mu}_i^m(t) + F(m,i,t)$ where $F(m,i,t) = \sqrt{\frac{C_1 \ln t}{n_{m,i}(t)}}$ (for sub-Gaussian rewards) and $F(m,i,t) = \sqrt{\frac{C_1 \ln T}{n_{m,i}(t)}} + \frac{C_2 \ln T}{n_{m,i}(t)}$ (for sub-exponential rewards). Also, we propose new the information update rule due to the differences in reward and graph dynamics.

5.2 Regret Analysis

Importantly, we next demonstrate the effectiveness of Algorithm 3 and Algorithm 2 by examining the regret upper bound. In particular, we stress that Theorem 5.1 does not rely on Assumption 1, meaning that the results address both light-tailed and heavy-tailed degree distributions in the random graph model (2.1).

Theorem 5.1. *Let Assumptions 2 and 3 hold. Let Algorithm 2 run with Rule 2. Then, we have that with* $\mathbf{P}(A_{\zeta,\delta}) \ge 1 - 7/T$ *and*

$$\begin{aligned}
\mathbf{E}[R_T|A_{\zeta,\delta}] &\le 2\kappa K(\log M)^2 \log T + \sum_{i\in[K]} M\Delta_i \cdot \left(\max\left\{\frac{2cN\log T}{(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}}, 2\kappa \log M \log T\right\}\right)\\
&\quad + \sum_{i\in[K]} M\Delta_i \cdot \left(\frac{2\pi^2}{3} + 2\kappa(\log M)^2 \log T\right)\\
&= O(M \log T)
\end{aligned}$$

where K is the count of arms, the event is defined by $A_{\zeta,\delta} = \{n_{m,i}(t_{m,j}) \ge n_{m,i}(t) - \kappa(\log M)^2 \log T \ \forall t \le T, \ \forall m,i,j\}$, $N = (12^{\frac{1}{1+\epsilon}})^{\frac{1+\epsilon}{\epsilon}} + 1$, and $c, B, \kappa, \rho, \epsilon$ are defined in Theorem 4.1.

Proof Sketch. Note that we do not exploit the hub structure in this setting, as clients need to collect information from all other clients rather than relying solely on a hub that contains only a subset of information. Nevertheless, the information delay bounds in Lemma 3.3 (and hence Lemma B.8) for sparse graphs still hold. Using the estimators constructed in Rule 2, which leverage neighbor

information to collect and integrate global information, we prove a concentration inequality for the global estimator $\hat{\mu}$ with respect to the global mean values:

$$|\hat{\mu}_m^i(n_{m,i}(t);k) - \mu_m^i| \leq 2\rho^{\frac{1}{1+\epsilon}}\left(\frac{Mc\log(T)}{\min_m n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}.$$

This ensures that clients can identify the globally optimal arm using UCB after

$$2cM\log T \Big/ \left(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}}\right)^{\frac{1+\epsilon}{\epsilon}}$$

steps with high probability. Another possible scenario is that, when clients are not synchronized, they randomly select arms instead of using UCB, which is proved to be upper bounded by the second constant term. As a result, the total number of pulls of sub-optimal arms $n_{m,i}(t)$ can be bounded by $O(\log T)$. Lastly, regret decomposition shows that the regret is dominated by $n_{m,i}(t)$, and thus has the upper bound. See Appendix for the detailed proof. □

Remark 4 (Extension)**.** *We emphasize that this theorem holds for both light-tailed and heavy-tailed random graphs as it does not rely on Assumption 1. This also highlights a key difference between the homogeneous and heterogeneous settings: in the homogeneous case, heavy-tailed h_i's lead to sample complexity reduction compared to the light-tailed case (with regret of order $O(M\log T)$), whereas the heterogeneous result holds true for both light-tailed and heavy-tailed distributions for the attraction weight h_i's.*

Comparison with the existing work We analyze the regret order with respect to T, as the absence of communication can lead to regret of order $O(T)$ [27]. The most relevant work [26] establishes a regret upper bound of $O(\log T)$ specifically for Erdős–Rényi graph with light-tailed and dense dynamics. In particular, the dense connectivity in [26] requires any pair of clients connects with a rather high probability (of order $O(1)$) at any time step, which limits practical applicability. Besides, their work covers some reward distributions that are heavier than the sub-Gaussian class, but still have finite moment-generating functions (MGFs). In contrast, we impose no such assumptions on graph connectivity, consider sparse graphs with only $O(M)$ total degree (instead of the $O(M^2)$ total degree in [26]), and allow for a significantly more general class of reward distributions, potentially with infinite variance and lacking finite MGFs. On the other hand, [30] achieves an regret bound of the same order, but assumes a connected or periodically connected graph under sub-Gaussian rewards, which may not hold for sparse graphs. Last but not least, our results close an open problem in [10] regarding homogeneous rewards under heavy-tailed settings (Section 4.2), thus bridging the gap in existing literature and addressing challenges posed by heavy-tailed graphs.

6 Conclusion and Future Work

We characterize the multi-agent multi-armed bandit problem with heavy tails in both rewards and graphs to capture complex real-world scenarios. We consider a setting where M clients have asymmetric and low degrees on graphs, and reward observations can deviate significantly from the mean. These complexities introduce challenges in both client communication and statistical inference of reward mean values, potentially leading to larger regret. Surprisingly, we prove that with novel algorithm designs, regret is sublinear in M and T, with an (almost) order of $O(M^{1-\frac{1}{\alpha}}\log T)$ and $O(M\log T)$, in homogeneous and heterogeneous settings, respectively. These results improve the regret bounds in existing work that considers less complex settings.

Moving forward, it would be valuable to explore other types of random graphs and analyze how regret scales with different graph dynamics. Also, while our framework allows clients to share information about their neighbors, removing this assumption in future research would be exciting.

References

[1] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.

[2] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

[3] M. Boguná and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68(3):036112, 2003.

[4] S. Bubeck, N. Cesa-Bianchi, and G. Lugosi. Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711–7717, 2013.

[5] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002.

[6] D. J. Clancy. Epidemics on critical random graphs with heavy-tailed degree distribution, 2021.

[7] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[8] J. P. da Cruz and P. G. Lind. The bounds of heavy-tailed return distributions in evolving complex networks. *Physics Letters A*, 377(3):189–194, 2013.

[9] A. Dubey and A. Pentland. Thompson sampling on symmetric α-stable bandits. *arXiv preprint arXiv:1907.03821*, 2019.

[10] A. Dubey and A. Pentland. Cooperative multi-agent bandits with heavy tails. In *International Conference on Machine Learning*, 2730–2739, 2020.

[11] T. E. Harris et al. *The theory of branching processes*, volume 6. Springer Berlin, 1963.

[12] E. J. Hearnshaw and M. M. Wilson. A complex network approach to supply chain network theory. *International Journal of Operations & Production Management*, 33(4):442–469, 2013.

[13] H. Jia, C. Shi, and S. Shen. Multi-armed bandit with sub-exponential rewards. *Operations Research Letters*, 49(5):728–733, 2021.

[14] N. Korda, E. Kaufmann, and R. Munos. Thompson sampling for 1-dimensional exponential family bandits. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

[15] J. Kunegis. Konect: the koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, page 1343–1350, New York, NY, USA, 2013. Association for Computing Machinery.

[16] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Phys. Rev. E*, 65:035108, Mar 2002.

[17] S. I. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media, 2007.

[18] R. Roman, J. Zhou, and J. Lopez. On the features and challenges of security and privacy in distributed internet of things. *Computer networks*, 57(10):2266–2279, 2013.

[19] Y. Tao, Y. Wu, P. Zhao, and D. Wang. Optimal rates of (locally) differentially private heavy-tailed multi-armed bandits. In G. Camps-Valls, F. J. R. Ruiz, and I. Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 1546–1574. PMLR, 28–30 Mar 2022.

[20] S. Vakili, K. Liu, and Q. Zhao. Deterministic sequencing of exploration and exploitation for multi-armed bandit problems. *IEEE Journal of Selected Topics in Signal Processing*, 7(5):759–767, 2013.

[21] R. van der Hofstad, A. J. E. M. Janssen, J. S. H. van Leeuwaarden, and C. Stegehuis. Local clustering in scale-free networks with hidden variables. *Phys. Rev. E*, 95:022307, Feb 2017.

[22] R. van der Hofstad, P. van der Hoorn, N. Litvak, and C. Stegehuis. Limit theorems for assortativity and clustering in null models for scale-free networks. *Advances in Applied Probability*, 52(4):1035–1084, 2020.

[23] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the internet. *Phys. Rev. E*, 65:066130, Jun 2002.

[24] X. Wang, S. Oh, and C.-H. Rhee. Eliminating sharp minima from SGD with truncated heavy-tailed noise. In *International Conference on Learning Representations*, 2022.

[25] W. Willis. Bounds for the independence number of a graph. 2011.

[26] M. Xu and D. Klabjan. Decentralized randomly distributed multi-agent multi-armed bandit with heterogeneous rewards. 2023.

[27] M. Xu and D. Klabjan. Regret lower bounds in multi-agent multi-armed bandit. *arXiv preprint arXiv:2308.08046*, 2023.

[28] M. Xu and D. Klabjan. Decentralized randomly distributed multi-agent multi-armed bandit with heterogeneous rewards. *Advances in Neural Information Processing Systems*, 36, 2024.

[29] L. Yang, X. Wang, M. Hajiesmaili, L. Zhang, J. C. Lui, and D. Towsley. Cooperative multi-agent bandits: Distributed algorithms with optimal individual regret and communication costs. In *Coordination and Cooperation for Multi-Agent Reinforcement Learning Methods Workshop*, 2023.

[30] J. Zhu and J. Liu. Distributed multi-armed bandits. *IEEE Transactions on Automatic Control*, 2023.

A Pseudo Code

Algorithm 3 HT-HTUCB (Heavy-Tailed Heterogeneous UCB): Burn-in period

Initialization: The length of the burn-in period is L and we are also given $\tau_1 < L$; In the time step $t = 0$, the estimates are initialized as $\bar{\mu}_i^m(0) = 0$, $n_{m,i}(0) = 0$, $\hat{\mu}_{i,j}^m(0) = 0$, and $P_0(m, j) =$ for any arm i and clients m, j

for $\tau_1 < t \leq L$ **do**

  The environment generates a sample graph $G_t = (V, E_t)$

  for *each client* m **do**

    Sample arm $a_t^m = (t \mod K)$

    Receive rewards $r_{a_t^m}^m(t)$ and update $n_{m,i}(t) = n_{m,i}(t-1) + \mathbb{1}_{a_m^t = i}$

    Update the local estimates for any arm i: $\bar{\mu}_i^m(t) = \frac{n_{m,i}(t-1)\bar{\mu}_i^m(t-1) + r_{a_t^m}^m(t) \cdot 1_{a_t^m = i}}{n_{m,i}(t-1) + 1_{a_t^m = i}}$

    Update the maintained matrix $P_t(m, j) = \frac{(t-1)P_{t-1}(m,j) + X_{m,j}^t}{t}$ for each $j \in V$

    Send $\{\bar{\mu}_i^m(t)\}_{i=1}^{i=K}$ to all clients in $\mathcal{N}_m(t)$

    Receive $\{\bar{\mu}_i^j(t)\}_{i=1}^{i=K}$ from all clients $j \in \mathcal{N}_m(t)$ and store them as $\hat{\mu}_{i,j}^m(t)$.

  end

end

for *each client* m *and arm* i **do**

  For client $1 \leq j \leq M$, set $h^L(m, j) = \max_{s \geq 1}\{(m, j) \in E_s\}$ or 0 if such s does not exist

  $\tilde{\mu}_i^m(L+1) = \sum_{j=1}^M P'_{m,j}(L)\hat{\mu}_{i,j}^m(h_{m,j}^L)$ where $P'_{m,j}(L) = \begin{cases} \frac{1}{M} & \text{if } P_L(m, j) > 0 \\ 0 & \text{otherwise} \end{cases}$

end

Algorithm 4 HT-HMUCB (Heavy-Tailed Homogeneous UCB): Identification of Hub Center

Initialization: The length of search period L; set $d_i(j) = -1$ for any $i, j \in [M]$

for $t = 1$ **do**

  The environment generates graph $G_1 = (V, E_1)$ at time $t = 1$

  for *each client* $m \in [M]$ **do**

    Set $d_m(m) \leftarrow d_m$, where d_m is the degree of m on G_1

  end

end

for *t = 2,3,. . . , L* **do**

  for *each client* $m \in [M]$ **do**

    If $d_j(m) \geq 0$ for some $j \in [M]$, send $d_j(m)$ to all clients on the neighbor set $\mathcal{N}_m(t)$

    For any $d_j(i)$ received from a neighbor $i \in \mathcal{N}_m(t)$, if $d_j(m) = -1$, then set $d_j(m) \leftarrow d_j(i)$

  end

end

for *each client* $m \in [M]$ **do**

  Set $\hat{i}(m) \leftarrow \arg\max_{i \in [M]} d_i(m)$; when the argument minimum is not unique, pick the smallest index

  If $\hat{i}(m) = m$, act as the *hub center* from now on; otherwise, act as *non-center*

end

B Proof for Section 3

We first set a few notations that will be used frequently in this section. Let $\mathbb{Z}$ be the set of integers. For any $x \in \mathbb{R}$, we use $\lfloor x \rfloor \triangleq \max\{n \in \mathbb{Z} : \ n \leq x\}$, $\lceil x \rceil \triangleq \min\{n \in \mathbb{Z} : \ n \geq x\}$ to denote the floor

and ceiling function. For any $x, y \in \mathbb{R}$, let $x \wedge y \triangleq \min\{x, y\}$ and $x \vee y \triangleq \max\{x, y\}$.

B.1 Proof of Lemmas 3.1 and 3.2

To prove Lemma 3.1, we first prepare a few technical tools. Let $H_M \triangleq \max_{i\in[M]} h_i$, where h_i's are iid copies of some random variable h taking values in $[0, \infty)$. By imposing Assumption 1 on the law of h, Lemma B.1 establishes a high-probability bound for H_M.

Lemma B.1. *Let Assumption 1 hold. For any $\Delta \in (0, 1/\alpha)$,*

$$\mathbf{P}\big(H_M \le M^{\frac{1}{\alpha}-\Delta}\big) = o\Big(\exp\big(-M^{\alpha\Delta/2}\big)\Big) \qquad \text{as } M \to \infty.$$

Proof. We write $n(M) = M^{\frac{1}{\alpha}-\Delta}$ in this proof, and observe that

$$\begin{aligned}
\mathbf{P}(H_M \le n(M)) &= \mathbf{P}(h_i \le n(M)\ \forall i \in [M]) \\
&= \Big(1 - \mathbf{P}(h > n(M))\Big)^M \qquad \text{by the iid nature of } h_i\text{'s;} \\
\Longrightarrow \log \mathbf{P}(H_M \le n(M)) &= M \cdot \log\Big(1 - \mathbf{P}(h > n(M))\Big) \\
&\le -M \cdot \mathbf{P}(h > n(M)). \qquad \text{(B.1)}
\end{aligned}$$

The inequality follows from $\log(1-x) \le -x$ for all $x \in (0,1)$. Due to our choice of $\Delta > 0$, we can fix $\epsilon > 0$ small enough such that $(\frac{1}{\alpha} - \Delta)(\alpha + \epsilon) < 1 - \frac{\alpha\Delta}{2}$. By Potter's bound for regularly varying functions (see, e.g., Proposition 2.6 of [17]), it holds for all x large enough that $\mathbf{P}(h > x) \ge x^{-(\alpha+\epsilon)}$. Therefore, for any M large enough, we have

$$\mathbf{P}(h > n(M)) \ge M^{-(\frac{1}{\alpha}-\Delta)(\alpha+\epsilon)} > M^{-1+\frac{\alpha\Delta}{2}}.$$

As a result, we have in (B.1) that

$$\mathbf{P}(H_M \le n(M)) = o\Big(\exp\big(-M^{\alpha\Delta/2}\big)\Big)$$

for all M large enough. This concludes the proof. □

Let $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ be the set of non-negative integers. Let d_i^t be the degree of the i^{th} node over graph G_t. To lighten notations we write $d_i = d_i^1$ at time $t = 1$. Under Assumption 2, Lemma B.2 provides useful bounds for the conditional law of the degree d_i.

Lemma B.2. *Let Assumption 2 hold. Let $\theta = \mathbf{E}h$ in (2.1). For any $M \in \mathbb{Z}_+$, $\beta \in (0,1)$ and $\Delta \in (0, \beta)$,*

$$\max_{i\in[M]} \mathbf{P}\big(d_i \ge M^{\beta+\Delta} \;\big|\; h_i \le M^{\beta}\big) = o\big(\exp(-M^{\beta-\Delta})\big), \qquad \text{(B.2)}$$

$$\max_{i\in[M]} \mathbf{P}\big(d_i \le M^{\beta-\Delta} \;\big|\; h_i \ge M^{\beta}\big) = o\big(\exp(-M^{\beta-\Delta})\big). \qquad \text{(B.3)}$$

Proof. **Proof of Claim** (B.2). Since the marginal distribution of $\big(h_i, d_i\big)$ is identical for each $i \in [M]$, it suffices to show that

$$\mathbf{P}\big(d_1 \ge M^{\beta+\Delta} \;\big|\; h_1 \le M^{\beta}\big) = o\big(\exp(-M^{\beta-\Delta})\big).$$

On event $\{h_1 \le M^\beta\}$, by the definition of kernel $P(\cdot,\cdot)$ in (2.1), we have

$$d_1 | \{h_1 \le M^\beta\} \underset{\text{s.t.}}{\le} \sum_{i=2}^{M} \underbrace{\text{Bernoulli}\Big(\frac{h_i M^\beta}{\theta M} \wedge 1\Big)}_{=V_i}.$$

Here, $X \underset{\text{s.t.}}{\leq} Y$ refers the the stochastic dominance where $\mathbf{P}(X > z) \leq \mathbf{P}(Y > z)\ \forall z \in \mathbb{R}$, and $X|A$ denote the conditional law $\mathbf{P}(Z > x) = \mathbf{P}(X > x|A)$. Besides, note that V_i's are iid, with $0 \leq V_i \leq 1$ and (recall that $\theta = \mathbf{E}h_i$)

$$\begin{aligned}
\mathbf{E}V_i = \mathbf{E}\bigg(\frac{h_i M^\beta}{\theta M} \wedge 1\bigg) \leq \mathbf{E}\bigg(\frac{h_i M^\beta}{\theta M}\bigg) = \frac{M^\beta}{\theta M} \cdot \mathbf{E}h_i = \frac{M^\beta}{M}, \\
\Longrightarrow \mathbf{E}\bigg[\sum_{i=2}^{M} V_i\bigg] = (M-1)\mathbf{E}V_2 \leq M^\beta < \frac{1}{2}M^{\beta+\Delta} \qquad \text{for any } M \text{ large enough.} \qquad \text{(B.4)}
\end{aligned}$$

On the other hand, by Assumption 2,

$$\begin{aligned}
\mathbf{E}V_i \geq \frac{c_h M^\beta}{\theta M} \qquad \text{for all } M \text{ large enough due to } \beta < 1 \\
\Longrightarrow \mathbf{E}\bigg[\sum_{i=2}^{M} V_i\bigg] \geq (M-1) \cdot \frac{c_h M^\beta}{M} \geq \frac{c_h}{2} M^\beta \qquad \text{for any } M \text{ large enough.} \qquad \text{(B.5)}
\end{aligned}$$

Therefore, for any M sufficiently large,

$$\begin{aligned}
\mathbf{P}\bigg(\sum_{i=2}^{M} V_i \geq M^{\beta+\Delta}\bigg) &\leq \mathbf{P}\bigg(\sum_{i=2}^{M} V_i \geq 2\bigg[\mathbf{E}\sum_{i=2}^{M} V_i\bigg]\bigg) \qquad \text{by (B.4)} \\
&\leq \exp\bigg(-\frac{1}{3}\mathbf{E}\bigg[\sum_{i=2}^{M} V_i\bigg]\bigg) \qquad \text{by Chernoff bound} \\
&\leq \exp\bigg(-\frac{1}{6}c_h M^\beta\bigg) \qquad \text{by (B.5)} \\
&= o\Big(\exp(-M^{\beta-\Delta})\Big).
\end{aligned}$$

This concludes the proof of (B.2).

Proof of Claim (B.3). Analogously, it suffices to prove the claim for $i = 1$. Due to $\beta < 1$, it holds for any M large enough that $\frac{c_h M^\beta}{\theta M} < 1$. Let V_i's be iid copies of Bernoulli($\frac{c_h M^\beta}{\theta M}$). On event $\{h_1 \geq M^\beta\}$, it follows from Assumption 2 and the definition of kernel $P(\cdot,\cdot)$ in (2.1) that

$$\sum_{i=2}^{M} V_i \underset{\text{s.t.}}{\leq} d_1|\{h_1 \geq M^\beta\}.$$

Furthermore,

$$\begin{aligned}
\mathbf{E}\bigg[\sum_{i=2}^{M} V_i\bigg] \geq (M-1) \cdot \frac{c_h M^\beta}{\theta M} \geq \frac{c_h}{2\theta} M^\beta \qquad \text{for all } M \geq 2 \\
\geq 2M^{\beta-\Delta} \qquad \text{for any } M \text{ large enough.}
\end{aligned}$$

As a result, for any M sufficiently large,

$$\begin{aligned}
\mathbf{P}\bigg(\sum_{i=2}^{M} V_i \leq M^{\beta-\Delta}\bigg) &\leq \mathbf{P}\bigg(\sum_{i=2}^{M} V_i \leq \frac{1}{2}\mathbf{E}\bigg[\sum_{i=2}^{M} V_i\bigg]\bigg) \\
&\leq \exp\bigg(-\frac{1}{4}\mathbf{E}\bigg[\sum_{i=2}^{M} V_i\bigg]\bigg) \qquad \text{by Chernoff bound}
\end{aligned}$$

$$\leq \exp\Big(-\frac{c_h}{8\theta}M^{\beta}\Big)$$
$$= o\big(\exp(-M^{\beta-\Delta})\big).$$

This concludes the proof of (B.3). □

Recall the definition of

$$\hat{i} \triangleq \arg\max_{i\in[M]} d_i, \tag{B.6}$$

which is the node with the highest degree over graph G_1 (i.e., at time $t = 1$). When there are ties, we arbitrarily pick one of argument maximum as $\hat{i}$. As an immediate consequence from Lemma B.2, the next Lemma shows that the node $\hat{i}$ will almost always have a large weight $h_{\hat{i}}$; in other words, the node with empirically largest degree (at time $t = 1$) will almost always have a large attraction weight.

Lemma B.3. *Let Assumptions 1 and 2 hold. Define event*

$$B(M,\Delta) = \big\{h_{\hat{i}} > M^{\frac{1}{\alpha}-\Delta}\big\}.$$

For any $\Delta \in (0, \frac{1}{2\alpha})$, and $\gamma > 0$ small enough such that

$$\gamma < \frac{\alpha\Delta}{4}, \qquad \gamma < \frac{1}{\alpha} - 2\Delta, \tag{B.7}$$

we have (as $M \to \infty$)

$$\mathbf{P}\Big(B(M,\Delta)^c\Big) = o\Big(\exp(-M^{\gamma})\Big).$$

Proof. In this proof, let $i^* = \arg\max_{i\in[M]} h_i$ denote the index of the node with largest weight h_i. Again, when there are ties, we arbitrarily pick one of such i^*'s. Note that on event $\big\{d_i < M^{\frac{1}{\alpha}-\frac{\Delta}{2}}\ \forall i \in [M] \text{ with } h_i < M^{\frac{1}{\alpha}-\Delta}\big\}$, the claim $h_{\hat{i}} \geq M^{\frac{1}{\alpha}-\Delta}$ must hold if, for some $i \in [M]$ with $d_i \geq M^{\frac{1}{\alpha}-\frac{\Delta}{2}}$, we have $h_i \geq M^{\frac{1}{\alpha}-\Delta}$. In particular, note that

$$B(M,\Delta) \supseteq \Big\{h_{i^*} \geq M^{\frac{1}{\alpha}-\frac{\Delta}{2}},\ d_{i^*} \geq M^{\frac{1}{\alpha}-\Delta}\Big\} \cap \Big\{d_i < M^{\frac{1}{\alpha}-\frac{\Delta}{2}}\ \forall i \in [M] \text{ with } h_i < M^{\frac{1}{\alpha}-\Delta}\Big\},$$

which leads to the upper bound

$$\mathbf{P}\Big(B(M,\Delta)^c\Big) \leq \underbrace{\mathbf{P}\big(h_{i^*} \leq M^{\frac{1}{\alpha}-\frac{\Delta}{2}}\big)}_{=p_1(M,\Delta)} + \underbrace{\mathbf{P}\Big(d_{i^*} < M^{\frac{1}{\alpha}-\Delta} \,\Big|\, h_{i^*} \geq M^{\frac{1}{\alpha}-\frac{\Delta}{2}}\Big)}_{p_2(M,\Delta)}$$
$$+ M \cdot \underbrace{\mathbf{P}\Big(d_1 > M^{\frac{1}{\alpha}-\frac{\Delta}{2}} \,\Big|\, h_1 < M^{\frac{1}{\alpha}-\Delta}\Big)}_{p_3(M,\Delta)}.$$

By Lemma B.1, $p_1(M,\Delta) = o\big(\exp(-M^{\alpha\Delta/4})\big)$. By Lemma B.2, we get $p_2(M,\Delta) = o\big(\exp(-M^{\frac{1}{\alpha}-\Delta})\big)$ and $p_3(M,\Delta) = o\big(\exp(-M^{\frac{1}{\alpha}-2\Delta})\big)$. By our choice of $\gamma > 0$ in (B.7), we conclude the proof. □

Recall that $\theta = \mathbf{E}h$. Let

$$S_{*,\Delta} = \big\{i \in [M] :\ h_i > \theta M^{1+\Delta-\frac{1}{\alpha}}\big\} \tag{B.8}$$

be the collection of nodes with large weights h_i w.r.t. threshold $\theta M^{1+\Delta-\frac{1}{\alpha}}$. The next lemma develops high-probability bounds for the size of $S_{*,\Delta}$.

Lemma B.4. *Let Assumption 1 hold with $\alpha \in (1,2)$. For any $\zeta \in (0, 2-\alpha)$, any $\Delta > 0$ small enough such*

$$(\alpha+\Delta)(1+\Delta-\frac{1}{\alpha}) \le \alpha - 1 + \frac{\zeta}{3}, \tag{B.9}$$

we have

$$\mathbf{P}\big(|S_{*,\Delta}| \le M^{2-\alpha-\zeta}\big) = o\Big(\exp(-M^{2-\alpha-\zeta})\Big) \qquad \text{as } M \to \infty.$$

Proof. We write $n(M) = \theta M^{1+\Delta-\frac{1}{\alpha}}$ and observe that $|S_{*,\Delta}| \overset{d}{=} \text{Binomial}\big(M, \mathbf{P}(h > n(M))\big)$. Let V_i's be iid copies of $\text{Bernoulli}\big(\mathbf{P}(h > n(M))\big)$. Observe that

$$\begin{aligned}
\mathbf{E}\Big[\sum_{i=1}^{M} V_i\Big] &= M\mathbf{P}(h > n(M)) \\
&\ge M \cdot \frac{1}{\big(n(M)\big)^{\alpha+\Delta}} \qquad \text{for any } M \text{ large enough due to Potter's bound (Proposition 2.6 of [17])} \\
&\ge \frac{1}{\theta^{\alpha+\Delta}} \cdot \frac{M}{M^{\alpha-1+\frac{\zeta}{3}}} \qquad \text{by the choice of } \Delta \text{ in (B.9) and the definition } n(M) = \theta M^{1+\Delta-\frac{1}{\alpha}} \\
&= \frac{1}{\theta^{\alpha+\Delta}} \cdot M^{2-\alpha-\frac{\zeta}{3}} \\
&\ge 2M^{2-\alpha-\frac{2\zeta}{3}} \qquad \text{for any } M \text{ large enough.}
\end{aligned}$$

Therefore, for such large M,

$$\begin{aligned}
\mathbf{P}\big(|S_{*,\Delta}| \le M^{2-\alpha-\zeta}\big) &\le \mathbf{P}\bigg(\sum_{i=1}^{M} V_i \le \frac{1}{2}\mathbf{E}\Big[\sum_{i=1}^{M} V_i\Big]\bigg) \\
&\le \exp\bigg(-\frac{1}{4}\mathbf{E}\Big[\sum_{i=1}^{M} V_i\Big]\bigg) \qquad \text{by Chernoff bound} \\
&\le \exp\bigg(-\frac{1}{2}M^{2-\alpha-\frac{2\zeta}{3}}\bigg) \\
&= o\Big(\exp(-M^{2-\alpha-\zeta})\Big).
\end{aligned}$$

This concludes the proof. ∎

Now, we are ready to prove Lemma 3.1.

Lemma (Lemma 3.1)**.** *Let Assumptions 1 and 2 hold with $\alpha \in (1,2)$. Given $\zeta \in (0, 2-\alpha)$, there exists $\gamma > 0$ such that*

$$\mathbf{P}\Big(|S_0| \le M^{2-\alpha-\zeta}\Big) = o\Big(\exp(-M^{\gamma})\Big).$$

Proof of Lemma 3.1. Let $\hat{i}$ and $S_{*,\Delta}$ be defined as in (B.6) and (B.8), respectively. Recall that $S_0^t = \{i \in [M] : (i, \hat{i}) \in E_t\}$ is the set of agents communicating with $\hat{i}$ at time t. Note that by the definition of the kernel $P(\cdot,\cdot)$ in (2.1), on event $\{h_{\hat{i}} > M^{\frac{1}{\alpha}-\Delta}\}$ we must have $S_{*,\Delta} \subseteq S_0$ for any $t \ge 1$. As a result, we have

$$\big\{|S_0| > M^{2-\alpha-\zeta}\big\} \supseteq \{h_{\hat{i}} > M^{\frac{1}{\alpha}-\Delta}\} \cap \{|S_{*,\Delta}| > M^{2-\alpha-\zeta}\},$$

and hence

$$\mathbf{P}\Big(|S_0| \le M^{2-\alpha-\zeta}\Big) \le \mathbf{P}\big(h_{\hat{i}} \le M^{\frac{1}{\alpha}-\Delta}\big) + \mathbf{P}\big(|S_{*,\Delta}| \le M^{2-\alpha-\zeta}\big).$$

Now, we pick $\Delta \in (0, \frac{1}{2\alpha})$ small enough such that condition (B.9) holds, and then pick $\gamma \in (0, 2-\alpha-\zeta)$ small enough such that condition (B.7) holds. By Lemma B.3, $\mathbf{P}\big(h_{\hat{i}} \le M^{\frac{1}{\alpha}-\Delta}\big) = o\Big(\exp(-M^{\gamma})\Big)$. By Lemma B.4, $\mathbf{P}\big(h_{\hat{i}} \le M^{\frac{1}{\alpha}-\Delta}\big) + \mathbf{P}\big(|S_{*,\Delta}| \le M^{2-\alpha-\zeta}\big) = o\Big(\exp(-M^{2-\alpha-\zeta})\Big) = o\Big(\exp(-M^{\gamma})\Big)$, where the last step follows from our choice of $\gamma \in (0, 2-\alpha-\zeta)$. This concludes the proof. □

Now, we provide the proof of Lemma 3.2. For any $p \in (0,1)$, we say that X is Geom(p) if

$$\mathbf{P}(X > k) = (1-p)^k \qquad \forall k = 1, 2, \ldots.$$

Lemma (Lemma 3.2)**.** *Let Assumptions 1 and 2 hold. Let $\zeta \in (0, 1-\frac{1}{\alpha})$. There exists $\gamma > 0$ such that*

$$\mathbf{P}\big(h_{\hat{i}} < M^{\frac{1}{\alpha}-\frac{\zeta}{2}}\big) = o\big(\exp(-M^{\gamma})\big). \tag{B.10}$$

Furthermore, there exists $M_0 > 0$ such that

$$\mathbf{P}\left(\sup_{t \le T} t - \tau(t) > \log T \;\middle|\; h_{\hat{i}} \ge M^{\frac{1}{\alpha}-\frac{\zeta}{2}}\right) \le \frac{1}{MT} \qquad \forall M \ge M_0,\ T \ge 1, \tag{B.11}$$

where $\tau(t) = \max\{u \le t :\ |S_0^u| > M^{\frac{1}{\alpha}-\zeta}\}$.

Proof of Lemma 3.2. Claim (B.10) is exactly the content of Lemma B.3. The rest of the proof is devoted to establishing the claim (B.11). To proceed, let $T_0 = 0$, and

$$T_k \triangleq \min\big\{t > T_{k-1} :\ |S_0^t| > M^{\frac{1}{\alpha}-\zeta}\big\}$$

be the k^{th} time the hub around $\hat{i}$ is large (w.r.t. threshold $M^{\frac{1}{\alpha}-\epsilon}$). Let $V_k \triangleq T_k - T_{k-1}$ be the k^{th} time gap between two large hubs. Note that

$$\sup_{t \le T} t - \tau(t) \le \max\left\{V_k - 1 :\ k \ge 1,\ \sum_{i=1}^{k-1} V_i \le T\right\} \le \max_{k \le T} V_k - 1. \tag{B.12}$$

To proceed, note that $\sup_{t \ge 1} \mathbf{P}\big(|S_0^t| \le M^{\frac{1}{\alpha}-\zeta} \;\big|\; h_{\hat{i}} \ge M^{\frac{1}{\alpha}-\frac{\zeta}{2}}\big) = \mathbf{P}\big(|S_0^1| \le M^{\frac{1}{\alpha}-\zeta} \;\big|\; h_{\hat{i}} \ge M^{\frac{1}{\alpha}-\frac{\zeta}{2}}\big)$, since the sequence S_0^t's are independent across t when conditioned on the value of $h_{\hat{i}}$. Now, we fix some $\tilde{\gamma} \in (0, \frac{1}{\alpha}-\zeta)$. By Lemma B.2, there exists M_0 such that

$$\sup_{t \ge 1} \mathbf{P}\Big(|S_0^t| \le M^{\frac{1}{\alpha}-\zeta} \;\Big|\; h_{\hat{i}} \ge M^{\frac{1}{\alpha}-\frac{\zeta}{2}}\Big) \le \exp(-M^{\tilde{\gamma}}) \qquad \forall M \ge M_0.$$

As a result, there exists a coupling between $(V_i)_{i \le T}$ and $(\tilde{V}_i)_{i \le T}$, which are iid copies of Geom$\big(1 - \exp(-M^{\tilde{\gamma}})\big)$, such that

$$(V_1, \ldots, V_T) | \{h_{\hat{i}} \ge M^{\frac{1}{\alpha}}\} \underset{\text{s.t.}}{\le} (\tilde{V}_1, \ldots, \tilde{V}_T),$$

Together with the upper bound in (B.12), we yield (for any $M \ge M_0$)

$$\mathbf{P}\left(\sup_{t \le T} t - \tau(t) > \log T \;\middle|\; h_{\hat{i}} \ge M^{\frac{1}{\alpha}-\frac{\zeta}{2}}\right) \le T \cdot \mathbf{P}\big(\tilde{V}_1 - 1 > \log T\big) = T \cdot \big(\exp(-M^{\tilde{\gamma}})\big)^{1+\log T}.$$

Note that $\exp(-M^{\tilde{\gamma}} = o(M^{-2})$. By picking a larger M_0 if necessary, we can ensure that $\exp(-M^{\tilde{\gamma}} \leq M^{-2}$ for each $M \geq M_0$, and hence

$$\mathbf{P}\left(\sup_{t\leq T} t - \tau(t) > \log T \;\middle|\; h_{\hat{i}} \geq M^{\frac{1}{\alpha}-\frac{\zeta}{2}}\right) \leq T/M^{2\cdot(1+\log T)} \leq \frac{1}{M}\cdot\frac{T}{M^{2\log T}}.$$

Lastly, by picking an even larger M_0 if needed, we ensure that $M_0 \geq e$, so for each $M \geq M_0$ and $T \geq 1$, we have

$$\mathbf{P}\left(\sup_{t\leq T} t - \tau(t) > \log T \;\middle|\; h_{\hat{i}} \geq M^{\frac{1}{\alpha}-\frac{\zeta}{2}}\right) \leq \frac{1}{M}\cdot\frac{T}{e^{2\log T}} = \frac{1}{M}\cdot\frac{T}{T^2} = \frac{1}{MT},$$

which concludes the proof. □

B.2 Proof of Lemma 3.3

Central to our proof is the following stochastic dominance argument regarding the graphs $(G_t)_{t\geq 1}$. Specifically, given some non-empty subset of clients $S \subseteq [M]$, we recall the definitions of $\bar{S}^0 = S$ and

$$\bar{S}^t \triangleq \{i \in [M] :\ i \in \bar{S}^{t-1};\ \text{or } \exists j \in \bar{S}^{t-1} \text{ s.t. } (i,j) \in E_t\},$$

which represents the collection of clients that have received the message at time t, which was sent from S at time 1 and is passed to neighbors over graph G_u at each time $u \leq t$. Let $K \triangleq |S|$ be the count of nodes in S and, without loss of generality, write $S = \{i_1, i_2, \ldots, i_K\}$. Let

$$\bar{V}^1_{i_1} \triangleq \{i \in [M] :\ i \notin S,\ (i, i_1) \in E_1\}.$$

be the set of nodes that are outside of S but can be reached from i_1 within one step (i.e., they are neighbors of i_1 over the graph G_1 at time $t = 1$). Analogously, let

$$\bar{V}^1_{i_k} \triangleq \{i \in [M] :\ i \notin S,\ (i, i_k) \in E_1\} \setminus \left(\bigcup_{j\in[k-1]} \bar{V}^1_{i_j}\right) \qquad \forall k = 2, 3, \ldots, K$$

be the set of nodes that are outsides of S and can be reached by i_k (but not any i_j with $j \in [k-1]$) within one step. By definition, we have

$$\bar{V}^1_{i_j} \cap \bar{V}^1_{i_k} = \emptyset\ \forall j \neq k, \qquad \bar{S}^1 = \bigcup_{k\in[K]} \Big(\{i_k\} \cup \bar{V}^1_{i_k}\Big).$$

We first consider the case where $K < M/2$, and we are able to uniformly randomly pick $\lceil M/2 \rceil$ nodes that are outside of S. By only checking whether these nodes are connected to i_1, and due to the lower bound $c_h > 0$ in Assumption 2, as well as the definition of the kernel $P(h, h')$ in (2.1), we have

$$Z_M(c_h) \underset{\text{s.t.}}{\leq} |\bar{V}^1_{i_1}| \qquad \text{where } Z_M(c_h) \overset{d}{=} \text{Binomial}\Big(\lceil M/2 \rceil, \frac{c_h^2}{\theta M}\Big). \tag{B.13}$$

Again, for any random variables X and Y, we use $X \underset{\text{s.t.}}{\leq} Y$ to denote stochastic dominance between X and Y, in the sense that $\mathbf{P}(X \geq x) \leq \mathbf{P}(Y \geq x)$ holds for any $x \in \mathbb{R}$. Furthermore, consider the following inductive procedure for each $k = 2, 3, \ldots, K$: On the event

$$\left\{|S| + \sum_{j\in[k-1]} |\bar{V}^1_{i_j}| < M/2\right\}, \tag{B.14}$$

we are, again, able to uniformly randomly pick $\lceil M/2 \rceil$ nodes that are still outside of S and $\bigcup_{j\in[k-1]} \bar{V}_{i_j}^1$. Let $Z_M^{(k)}(c_h)$ be iid copies of $Z_M(c_h)$. By repeating the arguments above, on the event defined in (B.14), we have

$$Z_M^{(k)}(c_h) \underset{\text{s.t.}}{\leq} |\bar{V}_{i_k}^1| \qquad \forall k = 1, 2, \ldots, K. \tag{B.15}$$

Define a branching process (i.e., Galton-Watson process) $\left(X_t^S\right)_{t\geq 0}$ by

$$X_0^S = |S|, \qquad X_t^S = \sum_{i=1}^{X_{t-1}^S} \left(1 + Z_M^{(t,i)}(c_h)\right) \ \ \forall t \geq 1, \tag{B.16}$$

where $Z_M^{(n,i)}(c_h)$ are iid copies of $Z_M(c_h)$. By the arguments in (B.13)–(B.15),

$$\min\left\{X_t^S, \frac{M}{2}\right\} \underset{\text{s.t.}}{\leq} \min\left\{|\bar{S}^t|, \frac{M}{2}\right\} \qquad \forall t = 0, 1, 2, \ldots. \tag{B.17}$$

This coupling is crucial to our proof below. Specifically, we fix a few constants. Let

$$\rho_h \triangleq \frac{c_h^2}{2\theta}. \tag{B.18}$$

where $\theta = \mathbf{E}h$, and $c_h > 0$ is the constant lower bound for the law of h stated in Assumption 2. Next, pick $\gamma_h \in (0, 1)$ such that

$$(1 + \rho_h)\gamma_h \triangleq 1 + \frac{\rho_h}{2}. \tag{B.19}$$

Note that ρ_h and γ_h only depend on $\theta = \mathbf{E}h$ and the constant c_h in Assumption 2. That is, these constants only depend on the law of h, and do not vary with any other parameters. Let

$$T_M \triangleq \min\{t \geq 1 : \ |\bar{S}^t| \geq M/2\} \tag{B.20}$$

be the first time that at least half of the nodes have received the message sent out from S.

Lemma B.5. *Under Assumption 2, it holds for any non-empty $S \subseteq \{1, 2, \ldots, M\}$ that*

$$\mathbf{P}\left(T_M > \lceil \tilde{t}_M \rceil\right) \leq \frac{2}{2 + \rho_h},$$

where

$$\tilde{t}_M \triangleq \frac{\log M - \log\left(2(1 - \sqrt{\gamma_h})\right)}{\log\left(1 + \lceil \frac{M}{2} \rceil \cdot \frac{c_h^2}{\theta M}\right)}. \tag{B.21}$$

Proof. For the branching process defined in (B.16), we use $(X_n)_{n\geq 0}$ to denote the process under the initial value $X_0 = 1$. For any $t \geq 1$ such that $\mathbf{E}X_t > \frac{M}{2}$, observe that

$$\begin{aligned}
\mathbf{P}(T_M > t) &= \mathbf{P}(|\bar{S}^t| < M/2) \\
&\leq \mathbf{P}(X_t^S < M/2) \qquad \text{by (B.17)} \\
&\leq \mathbf{P}(X_t < M/2) \leq \mathbf{P}\left(|X_t - \mathbf{E}X_t| \geq \left|\mathbf{E}X_t - \frac{M}{2}\right|\right) \\
&\leq \frac{\operatorname{var}(X_t)}{\left|\mathbf{E}X_t - \frac{M}{2}\right|^2} \qquad \text{by Chebyshev's inequality.}
\end{aligned} \tag{B.22}$$

Let

$$x \triangleq \lceil \frac{M}{2} \rceil \cdot \frac{c_h^2}{\theta M}, \qquad m \triangleq 1 + x, \qquad \sigma \triangleq \sqrt{x \cdot \left(1 - \frac{c_h^2}{\theta M}\right)} \leq \sqrt{x}. \tag{B.23}$$

By Theorem 5.1 of [11],

$$\mathbf{E}X_t = m^t, \qquad \mathrm{var}(X_t) \leq \frac{\sigma^2 m^{2t}}{m^2 - m} \qquad \forall t \geq 1.$$

In particular, at $t = \lceil \tilde{t}_M \rceil$ (see (B.21)), we have

$$\begin{aligned}
\log \mathbf{E}X_{\lceil \tilde{t}_M \rceil} &\geq \frac{\log M - \log\left(2(1 - \sqrt{\gamma_h})\right)}{\log(1+x)} \cdot \log(1+x) = \log\left(\frac{M}{2(1-\sqrt{\gamma_h})}\right) \\
\implies \mathbf{E}X_{\lceil \tilde{t}_M \rceil} &\geq \frac{M}{2} \cdot \frac{1}{1 - \sqrt{\gamma_h}} \\
\implies \mathbf{E}X_{\lceil \tilde{t}_M \rceil} - \frac{M}{2} &\geq m^{\lceil \tilde{t}_M \rceil} \cdot \left[1 - (1 - \sqrt{\gamma_h})\right] = m^{\lceil \tilde{t}_M \rceil} \cdot \sqrt{\gamma_h} \\
\implies \left| \mathbf{E}X_{\lceil \tilde{t}_M \rceil} - \frac{M}{2} \right|^2 &\geq \gamma_h \cdot m^{2\lceil \tilde{t}_M \rceil}.
\end{aligned}$$

Plugging these bounds back into (B.22), we yield (under $t = \lceil \tilde{t}_M \rceil$)

$$\begin{aligned}
\mathbf{P}(T_M > \lceil \tilde{t}_M \rceil) &\leq \frac{\sigma^2 m^{2\lceil \tilde{t}_M \rceil}}{m^2 - m} \cdot \frac{1}{\gamma_h \cdot m^{2\lceil \tilde{t}_M \rceil}} = \frac{\sigma^2}{m(m-1)} \cdot \frac{1}{\gamma_h} \\
&\leq \frac{x}{x(1+x)} \cdot \frac{1}{\gamma_h} = \frac{1}{1+x} \cdot \frac{1}{\gamma_h} \\
&\leq \frac{1}{1+\rho_h} \cdot \frac{1}{\gamma_h} \qquad \text{by definition of } x \text{ in (B.23)} \\
&= \frac{2}{2+\rho_h} \qquad \text{by the definition of } \gamma_h \text{ in (B.19)}
\end{aligned}$$

and conclude the proof. □

Next, we recall a bound for the tail cdf of geometric random variables. Straightforward as it is, the bound is useful for our subsequent analysis. For any $p \in (0,1)$, we say that X is Geom(p) if

$$\mathbf{P}(X > k) = (1-p)^k \qquad \forall k = 1, 2, \ldots.$$

Lemma B.6 (Lemma G.3 of [24]). *Let $a : (0,\infty) \to (0,\infty)$, $b : (0,\infty) \to (0,\infty)$ be two functions such that $\lim_{\epsilon \downarrow 0} a(\epsilon) = 0, \lim_{\epsilon \downarrow 0} b(\epsilon) = 0$. Let $\{U(\epsilon) : \epsilon > 0\}$ be a family of geometric RVs with success rate $a(\epsilon)$, i.e. $\mathbf{P}(U(\epsilon) > k) = (1 - a(\epsilon))^k$ for $k \geq 1$. For any $c > 1$, there exists $\epsilon_0 > 0$ such that*

$$\exp\Big(-\frac{c \cdot a(\epsilon)}{b(\epsilon)}\Big) \leq \mathbf{P}\Big(U(\epsilon) > \frac{1}{b(\epsilon)}\Big) \leq \exp\Big(-\frac{a(\epsilon)}{c \cdot b(\epsilon)}\Big) \quad \forall \epsilon \in (0, \epsilon_0).$$

As an immediate consequence of Lemma B.5, the next result provides upper bounds for the information delay between a given set S and any node j. Note that $t_M = O(\log M)$.

Lemma B.7. *Under Assumption 2,*

$$\inf_{j \in [M]} \inf_{S \subseteq [M]:\ S \neq \emptyset} \mathbf{P}\big(j \in \bar{S}^{t_M}\big) \geq \frac{\rho_h}{2+\rho_h} \cdot \Big(1 - \exp(-\rho_h)\Big) - o(1) \qquad \text{as } M \to \infty,$$

where

$$t_M \triangleq \lceil 2 \cdot \frac{\log M - \log\left(2(1-\sqrt{\gamma_h})\right)}{\log(1+\rho_h)} \rceil.$$

Proof. Recall the definition of $\tilde{t}_M$ in (B.21). For any M large enough, we have $\lceil \tilde{t}_M \rceil \le t_M - 1$. Henceforth in the proof, we only consider such large M. By Lemma B.5,

$$\mathbf{P}\big(|\bar{S}^{t_M-1}| \ge M/2\big) = \mathbf{P}\big(T_M \le t_M - 1\big) = 1 - \mathbf{P}\big(T_M > t_M - 1\big) \ge \frac{\rho_h}{2+\rho_h}.$$

Next, on event $\big\{|\bar{S}^{t_M-1}| \ge M/2\big\}$, given any $j \in [M]$ we either have $j \in \bar{S}^{t_M-1}$ or $j \notin \bar{S}^{t_M-1}$. In the latter case, let

$$V_j \triangleq \Big|\Big\{i \in \bar{S}^{t_M-1} :\ (i,j) \in E_{t_M}\Big\}\Big|.$$

First, if $V_j > 0$, we then have $j \in S^{\boldsymbol{h}}_{t_M}$. Furthermore, conditioned on event $\big\{|\bar{S}^{t_M-1}| \ge M/2\big\}$, the lower bound c_h in Assumption 2 leads to the stochastic dominance relation that

$$\mathrm{Binomial}\Big(\lceil M/2 \rceil,\ \frac{c_h^2}{\theta M}\Big) \underset{\text{s.t.}}{\le} V_j \Big| \Big\{|\bar{S}^{t_M-1}| \ge M/2,\ j \notin \bar{S}^{t_M-1}\Big\}.$$

Therefore,

$$\begin{aligned}
&\inf_{j\in[M]} \mathbf{P}\big(j \in \bar{S}^{t_M} \ \big|\ |\bar{S}^{t_M-1}| \ge M/2,\ j \notin \bar{S}^{t_M-1}\big) \\
&= \inf_{j\in[M]} \mathbf{P}\big(V_j > 0 \ \big|\ |\bar{S}^{t_M-1}| \ge M/2,\ j \notin \bar{S}^{t_M-1}\big) \\
&\ge \mathbf{P}\bigg(\mathrm{Binomial}\Big(\lceil M/2 \rceil,\ \frac{c_h^2}{\theta M}\Big) > 0\bigg) = 1 - \mathbf{P}\bigg(\mathrm{Geom}\Big(\frac{c_h^2}{\theta M}\Big) > \lceil M/2 \rceil\bigg) \\
&= 1 - \exp\bigg(-\frac{c_h^2}{\theta M} \cdot \frac{M}{2}\bigg) - o(1) \qquad (\text{as } M \to \infty) \text{ by Lemma B.6} \\
&= 1 - \exp(-\rho_h) - o(1).
\end{aligned}$$

In summary,

$$\begin{aligned}
&\inf_{j\in[M]} \mathbf{P}\big(j \in \bar{S}^{t_M}\big) \\
&\ge \inf_{j\in[M]} \mathbf{P}\big(j \in \bar{S}^{t_M},\ |\bar{S}^{t_M-1}| \ge M/2\big) \\
&\ge \inf_{j\in[M]} \mathbf{P}\big(j \in \bar{S}^{t_M} \ \big|\ |\bar{S}^{t_M-1}| \ge M/2,\ j \notin \bar{S}^{t_M-1}\big) \cdot \mathbf{P}(|\bar{S}^{t_M-1}| \ge M/2,\ j \notin \bar{S}^{t_M-1}) \\
&+ \inf_{j\in[M]} \mathbf{P}\big(j \in \bar{S}^{t_M} \ \big|\ |\bar{S}^{t_M-1}| \ge M/2,\ j \in \bar{S}^{t_M-1}\big) \cdot \mathbf{P}(|\bar{S}^{t_M-1}| \ge M/2, j \in \bar{S}^{t_M-1}) \\
&= \inf_{j\in[M]} \mathbf{P}\big(j \in \bar{S}^{t_M} \ \big|\ |\bar{S}^{t_M-1}| \ge M/2,\ j \notin \bar{S}^{t_M-1}\big) \cdot \mathbf{P}(|\bar{S}^{t_M-1}| \ge M/2,\ j \notin \bar{S}^{t_M-1}) \\
&\quad + \inf_{j\in[M]} 1 \cdot \mathbf{P}(|\bar{S}^{t_M-1}| \ge M/2, j \in \bar{S}^{t_M-1}) \\
&\ge \inf_{j\in[M]} \mathbf{P}\big(j \in \bar{S}^{t_M} \ \big|\ |\bar{S}^{t_M-1}| \ge M/2,\ j \notin \bar{S}^{t_M-1}\big) \\
&\quad \cdot \Big(\mathbf{P}(|\bar{S}^{t_M-1}| \ge M/2, j \notin \bar{S}^{t_M-1}) + \mathbf{P}(|\bar{S}^{t_M-1}| \ge M/2, j \in \bar{S}^{t_M-1})\Big) \\
&= \inf_{j\in[M]} \mathbf{P}\big(j \in \bar{S}^{t_M} \ \big|\ |\bar{S}^{t_M-1}| \ge M/2,\ j \notin \bar{S}^{t_M-1}\big) \cdot \mathbf{P}(|\bar{S}^{t_M-1}| \ge M/2) \\
&\ge \Big(1 - \exp(-\rho_h) - o(1)\Big) \cdot \frac{\rho_h}{2+\rho_h}.
\end{aligned}$$

This concludes the proof. ∎

Now, we are ready to prove Lemma 3.3.

Lemma (Lemma 3.3)**.** *Under Assumption 2, there exists some $\kappa \in (0,\infty)$ such that for any $\gamma > 0$, $M \geq 1$, and any non-emtpy $S \subseteq [M]$,*

$$\mathbf{P}\big(j \notin \bar{S}^{\gamma\cdot\kappa\cdot(\log M)^2} \text{ for some } j \in [M]\big) \leq M^{-\gamma}. \tag{B.24}$$

Proof of Lemma 3.3. Note that we have a trivial upper bound 1 in the RHS of (B.24) under $M = 1$, so the claim holds trivially for $M = 1$. Henceforth in this proof we only consider $M \geq 2$. Let

$$\begin{aligned} q_h &\triangleq 1 - \frac{\rho_h}{4+2\rho_h}\cdot\big(1-\exp(\rho_h)\big) \in (0,1), \\ t_M &\triangleq \inf\left\{t \geq 1 : \inf_{j\in[M]}\inf_{S\subseteq[M]:\ S\neq\emptyset} \mathbf{P}(j \in \bar{S}^t) \geq 1 - q_h\right\}. \end{aligned} \tag{B.25}$$

First of all, by Lemma B.7, for all M large enough we have

$$t_M \leq \lceil 2\cdot\frac{\log M - \log\big(2(1-\sqrt{\gamma_h})\big)}{\log(1+\rho_h)}\rceil,$$

which confirms that $t_M = O(\log M)$. As a result, there exists $\tilde{\kappa} \in (0,\infty)$ such that $t_M \leq \tilde{\kappa}\log M$ for each $M \geq 2$. Then, observe that by Markov property, it holds for any $k \geq 1$ that

$$\begin{aligned} &\mathbf{P}\big(j \notin \bar{S}^{k\cdot\tilde{\kappa}\log M} \text{ for some } j \in [M]\big) \\ &\leq M\cdot\sup_{j\in[M]}\mathbf{P}\big(j \notin \bar{S}^{k\cdot\tilde{\kappa}\log M}\big) \\ &\leq M\cdot\prod_{l=1}^{k}\sup_{j\in[M]}\mathbf{P}\big(j \notin \bar{S}^{l\cdot\tilde{\kappa}\log M} \mid j \notin \bar{S}^{(l-1)\cdot\tilde{\kappa}\log M}\big) \\ &\leq M\cdot\left(1 - \inf_{j\in[M]}\inf_{S\subseteq[M]:\ S\neq\emptyset}\mathbf{P}\big(j \in \bar{S}^{\tilde{\kappa}\log M}\big)\right)^k \\ &\leq M\cdot\left(1 - \inf_{j\in[M]}\inf_{S\subseteq[M]:\ S\neq\emptyset}\mathbf{P}\big(j \in \bar{S}^{t_M}\big)\right)^k \qquad \text{due to } t_M \leq \tilde{\kappa}\log M \\ &\leq M\cdot q_h^k \qquad \text{by the definition in (B.25).} \end{aligned} \tag{B.26}$$

Lastly, given any $\gamma > 0$, by setting

$$k = \frac{\gamma+1}{\log(1/q_h)}\cdot\log M, \qquad \kappa = \frac{\gamma+1}{\log(1/q_h)}\cdot\tilde{\kappa}$$

in the display (B.26), we conclude the proof of claim (B.24). □

B.3 Reward Information Delay

We would like to add that the delay on spare graphs also leads to information asynchronization among clients, referred to as information delay. Such delays necessitate careful analysis in order to design an effective algorithm and address the challenges due to heavy-tailed observations and delayed information over sparse communication. To this end, we establish the following result regarding the delay in client information and prove that it is possible for the clients to stay synchronized within reasonable thresholds.

Lemma B.8. *Let $t_{m,j} = \max_{s\leq t}\{(m,j) \in E_s, c_m \neq c_j\}$ and $p = \frac{\eta}{TM}$. If $n_{m,i}(t) > 2\kappa(\log M)^2\log T$ for any client m and any arm i, then we obtain that with probability at least $1-p$, for $j \notin S_0$, $\min_m n_{m,i}(t_{m,j}) \geq \frac{1}{2}\min_m n_{m,i}(t)$ (B.27), $\min_m n_{m,i}(t) \geq \frac{1}{2}n_{m,i}(t)$ (B.28), and $N_{j,i}(t) = n_{m,i}(t_{m,j}) \geq n_{m,i}(t) - \kappa(\log M)^2\log T$ (B.29).*

B.4 Proof of Lemma B.8

Lemma (Lemma B.8)**.** *Let us assume that $p = \frac{\eta}{TM}$. Let us further assume that $L > 2\kappa(\log M)^2 \log T$ where L is the length of the burn-in period. Then we obtain with probability at least $1-p$, for $j \notin S_0$,*

$$\min_m n_{m,i}(t_{m,j}) \geq \frac{1}{2} \min_m n_{m,i}(t) \tag{B.30}$$

and

$$\min_m n_{m,i}(t) \geq \frac{1}{2} n_{m,i}(t) \tag{B.31}$$

and

$$N_{j,i}(t) = n_{m,i}(t_{m,j}) \geq n_{m,i}(t) - \kappa(\log M)^2 \log T \tag{B.32}$$

Proof of Lemma B.8. We demonstrate the proof steps as follows.

Based on Lemma in Section 3, we obtain that there exists $\kappa \in (0, \infty)$ such that given any $\gamma > 0$,

$$\mathbf{P}\big(j \notin \bar{S}^{\boldsymbol{h}}_{\gamma \cdot \kappa \cdot (\log M)^2} \text{ for some } j \in [M]\big) \leq M^{-\gamma}.$$

If we specify $\gamma = 2 \log T$, we obtain that after $\kappa \cdot \log T \cdot (\log M)^2$ steps, with probability at least $1 - M^{-2\log T}$, i.e. $1 - O(\frac{1}{T^2})$, all clients communicate. Equivalently, by the definition of $t_{m,j}$, we derive with probability $P_0 = 1 - O(\frac{1}{T^2})$,

$$t - t_{m,j} \leq \kappa \cdot \log T \cdot (\log M)^2.$$

Then we consider the above result for any client j and any t, and derive that

$$\begin{aligned}
&P(\forall m, t, t - t_{m,j} \leq \kappa \cdot \log T \cdot (\log M)^2) \\
&= 1 - P(\cup_{m,t}\{t - t_{m,j} \leq \kappa \cdot \log T \cdot (\log M)^2\}) \\
&\geq 1 - \sum_m \sum_t P(t - t_{m,j} \leq \kappa \cdot \log T \cdot (\log M)^2) \\
&= 1 - MTP_0 = 1 - \frac{1}{T}
\end{aligned} \tag{B.33}$$

by the Bonferroni's inequality.

As a result, we obtain that with probability at least $1 - \frac{1}{T}$

$$\begin{aligned}
&n_{m,i}(t_{m,j}) \\
&= n_{m,i}(t - (t - t_{m,j})) \\
&\geq n_{m,i}(t - \kappa \cdot \log T \cdot (\log M)^2)) \\
&\geq n_{m,i}(t) - \kappa \cdot \log T \cdot (\log M)^2)
\end{aligned}$$

which concludes part of the statement, i.e. Equation (B.32).

Meanwhile, we note that the clients follow the information (possibly delayed) from the hub to decide on which arm to pull and clients in the hub use the same strategy all the time as they share the same non-delayed information. Notably, such delay is upper bounded by $\kappa \cdot \log T \cdot (\log M)^2)$, which implies that for any m

$$\min_m n_{m,i}(t) \geq n_{m,i}(t) - \kappa \cdot \log T \cdot (\log M)^2)$$

By the choice of L, we have that

$$n_{m,i}(t) \geq n_{m,i}(L) \geq \frac{L}{K}$$

$$\geq 2\kappa \cdot \log T \cdot (\log M)^2)$$

which immediately implies that

$$\begin{aligned}\min_m n_{m,i}(t) &\geq n_{m,i}(t) - \kappa \cdot \log T \cdot (\log M)^2) \\ &\geq n_{m,i}(t) - \frac{1}{2} n_{m,i}(t) \\ &= \frac{1}{2} n_{m,i}(t).\end{aligned}$$

This completes the proof of Equation (B.31) in the statement.

Additionally, we note that

$$\begin{aligned}&\min_n n_{m,i}(t_{m,j}) \\ &\geq \frac{1}{2} n_{m,i}(t) \\ &\geq \frac{1}{2} \min_m n_{m,i}(t)\end{aligned}$$

where the first inequality holds true using Equation (B.31) and the second inequality uses the fact that $n_{m,i}(t) \geq \min_m n_{m,i}(t)$.

This completes the proof of Equation (B.32) and thus the entire proof of the statement. □

C Proof of Theorems

C.1 Proof of Theorem 4.1

Proof. First, we characterize the deviation of $\hat{\mu}_i^m(t)$ from the underlying groundtruth, the global mean value μ_i, through mathematical induction, given the construction of $\hat{\mu}_i^m$.

We would like to highlight that we do not characterize the variance of the estimators, since we are in a heavy-tailed reward regime, where the estimators may not necessarily have finite variance. In fact, we show that the heavy-tailed reward estimator can directly be bounded by the heavy-tail dynamics using median-of-means.

The next lemma is a two-sided version of Lemma 2 of [4] and provides the concentration inequality for the median-of-means estimator.

Lemma C.1 (Lemma 2 of [4]). *Let $\delta \in (0,1)$ and $\epsilon \in (0,1]$. Let X_n's be iid copies of X with $\mathbf{E}X = \mu$, and $\mathbf{E}|X-\mu|^{1+\epsilon} \leq v$. Let $k = \lfloor 8\log(e^{1/8}/\delta) \wedge n/2 \rfloor$ and $N = \lfloor n/k \rfloor$. Let*

$$\hat{\mu}_j = \frac{1}{N} \sum_{t=(j-1)N+1}^{jN} X_t \qquad \forall j = 1, 2, \ldots, k,$$

and let $\widehat{\mu}_M$ be the median of $(\hat{\mu}_j)_{j=1,2,\ldots,k}$. Then, with probability at least $1-2\delta$,

$$|\widehat{\mu}_M - \mu| \leq (12v)^{\frac{1}{1+\epsilon}} \left(\frac{16\log(e^{1/8}\delta^{-1})}{n} \right)^{\frac{\epsilon}{1+\epsilon}}.$$

By using Lemma C.4 with respect to the rewards from clients, denoted by set $rw_t = \{s \leq t : \{r_i^j(s)\}_{j \in \mathcal{N}_{m,t}(s), a_j^s = i}\}$ where m is a hub, we derive that the global estimator at the hub $m \in S_0$ meets the following.

Formally, we have that

$$|\hat{\mu}_i^m(t) - \mu^i| \leq 2C\rho^{\frac{1}{1+\epsilon}} \left(\frac{c\log(1/\delta)}{|rw_t|} \right)^{\frac{\epsilon}{1+\epsilon}} \quad \text{with probability at least } 1-\delta \qquad \forall n \geq 1.$$

where $\hat{\mu}_i^m(t)$ is the median of the means constructed as illustrated in Lemma C.4 based on Algorithm 4.

It is worth noting that by definition, the size of rw_t, which we denote as $|rw_t|$, is equivalent to $\sum_{j\in S_0} n_{j,i}(t) \geq |S_0| \min_j n_{j,i}(t)$.

Meanwhile, by our result on information delay as established in Section 3.2, we obtain that when $L \geq 2\kappa(\log M)^2 \log T$,

$$\min_m n_{m,i}(t) \geq \frac{1}{2} n_{m,i}(t)$$

which immediately implies that

$$\begin{aligned} |rw_t| &= \sum_{j\in S_0} n_{j,i}(t) \\ &\geq |S_0| \min_j n_{j,i}(t) \\ &\geq \frac{|S_0|}{2} n_{m,i}(t) \end{aligned}$$

Subsequently, we derive that the following concentration inequality

$$\begin{aligned} &|\hat{\mu}_i^m(t) - \mu^i| \\ &\leq 2C\rho^{\frac{1}{1+\epsilon}} \left(\frac{c\log(1/\delta)}{|rw_t|}\right)^{\frac{\epsilon}{1+\epsilon}} \\ &\leq 2C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} \text{ with probability at least } 1-\delta \qquad \forall n \geq 1. \end{aligned}$$

Then we consider the following regret decomposition that allows us to leverage the result from the above concentration inequality.

We recall that the optimal arm is

$$i^* = \arg\max_i \mu_i^m = \arg\max_i \mu_i.$$

By decomposing R_T, we obtain that

$$\begin{aligned} R_T &= \frac{1}{M}(\max_i \sum_{t=1}^{T}\sum_{m=1}^{M} \mu_i^m - \sum_{t=1}^{T}\sum_{m=1}^{M} \mu_{a_t^m}^m) \\ &= \sum_{t=1}^{T} \frac{1}{M}\sum_{m=1}^{M} \mu_{i^*}^m - \sum_{t=1}^{T}\frac{1}{M}\sum_{m=1}^{M} \mu_{a_t^m}^m \\ &\leq \sum_{t=1}^{L} |\frac{1}{M}\sum_{m=1}^{M} \mu_{i^*}^m - \frac{1}{M}\sum_{m=1}^{M} \mu_{a_t^m}^m| + \sum_{t=L+1}^{T} (\frac{1}{M}\sum_{m=1}^{M} \mu_{i^*}^m - \frac{1}{M}\sum_{m=1}^{M} \mu_{a_t^m}^m) \\ &\leq L + \sum_{t=L+1}^{T} (\frac{1}{M}\sum_{m=1}^{M} \mu_{i^*}^m - \frac{1}{M}\sum_{m=1}^{M} \mu_{a_t^m}^m) \\ &= L + \sum_{t=L+1}^{T} (\mu_{i^*} - \frac{1}{M}\sum_{m=1}^{M} \mu_{a_t^m}^m) \\ &= L + ((T-L)\cdot \mu_{i^*} - \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{K} n_{m,i}(T)\mu_i^m) \end{aligned}$$

where the first inequality uses that $|a| \geq a$ for any a and the second inequality holds by noting that $0 < \mu_i^j < 1$

We consider the number of pulls of arms resulting from the UCB strategies as follows.

We assert that the factors causing the selection of a sub-optimal arm i are explicitly defined by the decision rule of Algorithm 1. Specifically, the outcome $a_t^m = i$ occurs when any of the following conditions is satisfied:

- Case 1: $\tilde{\mu}_i^m - \mu_i > C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0| \cdot n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$,
- Case 2: $-\tilde{\mu}_{i^*}^m + \mu_{i^*} > C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0| \cdot n_{m,i^*}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$,
- Case 3: $\mu_{i^*} - \mu_i < 2C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0| \cdot n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$.

Subsequently, we obtain that the number of arm pulls of arm i for client m can be upper bounded as follows:

$$
\begin{aligned}
n_{m,i}(T) &\leq l + \sum_{t=L+1}^{T} 1_{\{a_t^m = i, n_{m,i}(t) > l\}} \\
&\leq l + \sum_{t=L+1}^{T} 1_{\{\tilde{\mu}_i^m - C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0| \cdot n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} > \mu_i, n_{m,i}(t-1) \geq l\}} \\
&\quad + \sum_{t=L+1}^{T} 1_{\{\tilde{\mu}_{i^*}^m + C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0| \cdot n_{m,i^*}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} < \mu_{i^*}, n_{m,i}(t-1) \geq l\}} \\
&\quad + \sum_{t=L+1}^{T} 1_{\{\mu_i + 2C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0| \cdot n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} > \mu_{i^*}, n_{m,i}(t-1) \geq l\}}.
\end{aligned}
$$

By taking expected values over $n_{m,i}(t)$ conditional on $A_{\zeta,\delta}$, we derive

$$
\begin{aligned}
&E[n_{m,i}(T)|A_{\zeta,\delta}] \\
&= l + \sum_{t=L+1}^{T} P(\tilde{\mu}_i^m - C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0| \cdot n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} > \mu_i, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \\
&\quad + \sum_{t=L+1}^{T} P(\tilde{\mu}_{i^*}^m + C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0| \cdot n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} < \mu_{i^*}, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \\
&\quad + \sum_{t=L+1}^{T} P(\mu_i + 2C\rho^{\frac{1}{1+\epsilon}} \left(\frac{2c\log(1/\delta)}{|S_0| \cdot n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} > \mu_{i^*}, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \\
&= l + \sum_{t=L+1}^{T} P(Case1, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \qquad \text{(C.1)} \\
&\quad + \sum_{t=L+1}^{T} P(Case2, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) + \sum_{t=L+1}^{T} P(Case3, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \qquad \text{(C.2)}
\end{aligned}
$$

where $l = \frac{2c\log T}{|S_0|(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}}$ with $\Delta_i = \mu_{i^*} - \mu_i$..

For the last term in (C.12), we have

$$\sum_{t=L+1}^{T} P(Case4: \mu_i + 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} > \mu_{i^*}, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})) = 0 \tag{C.3}$$

since the choice of l satisfies $l \geq \frac{2c\log T}{|S_0|(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}}$ with $\Delta_i = \mu_{i^*} - \mu_i$.

We start with the two terms, and subsequently obtain that on event $A_{\zeta,\delta}$

$$\begin{aligned}
&\sum_{t=L+1}^{T} P(Case2, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) + \sum_{t=1}^{T} P(Case3, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \\
&\leq \sum_{t=L+1}^{T} P(\tilde{\mu}_{m,i} - \mu_i > C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} | A_{\zeta,\delta}) + \qquad (C.4) \\
&\quad \sum_{t=1}^{T} P(-\tilde{\mu}_{m,i^*} + \mu_{i^*} > C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} | A_{\zeta,\delta}) \\
&\leq \sum_{t=1}^{T}(\frac{1}{t^2}) + \sum_{t=1}^{T}(\frac{1}{t^2}) \leq \frac{\pi^2}{3} \qquad (C.5)
\end{aligned}$$

where the first inequality utilizes the property of the probability measure when removing the event $n_{m,i}(t-1) \geq l$ and the second inequality holds by the aforementioned concentration inequality based on median-of-means.

As a result, by the above decomposition, we derive that

$$\begin{aligned}
&E[n_{m,i}(t) | A_{\zeta,\delta}] \\
&\leq l + \sum_{t=L+1}^{T} P(Case1, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \\
&\qquad + \sum_{t=L+1}^{T} P(Case2, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) + \sum_{t=L+1}^{T} P(Case3, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \\
&\leq \frac{2c\log T}{|S_0|(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} + \frac{\pi^2}{3}
\end{aligned}$$

Consequently, we consider the following upper bound on R_T by the previous decomposition, which gives us that

$$\begin{aligned}
R_T &\leq L + ((T-L)\cdot \mu_{i^*} - \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{K} n_{m,i}(T)\mu_i^m) \\
&\leq 2\kappa(\log M)^2 \log T + \sum_m \sum_i n_{m,i}(t)\Delta_i
\end{aligned}$$

and

$$\begin{aligned}
&E[R_T | A_{\zeta,\delta}] \\
&\leq L + (\sum_m (\frac{2c\log T}{|S_0|(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} + \frac{\pi^2}{3})) \cdot \Delta_i
\end{aligned}$$

$$\leq L + M(\frac{2c\Delta_i \log T}{|S_0|(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} + \frac{\pi^2}{3}\Delta_i) = O(\frac{M}{|S_0|}\log T)$$

Meanwhile, by the definition of $A_{\zeta,\delta}$, we obtain that $S_0 \geq M^{2-\alpha-\zeta}$. This completes the first part of the statement.

Then by direct computation with $S_0 \geq M^{2-\alpha-\zeta}$, we derive the upper bound on R_T with respect to M and T, which is

$$E[R_T|A_{\zeta,\delta}] \leq O(\frac{M}{|S_0|}\log T) \leq O(M^{\alpha-1+\zeta}\log T)$$

which concludes the second part of the statement and thus completes proof of Theorem 4.1. □

C.2 Proof of Theorem 4.2

Proof. We would like to highlight that the proof of Theorem 4.2 overlaps with the proof of Theorem 4.1, since they both consider Algorithm 2 in a homogeneous setting. For completeness, we present the full proof here, and may repeat some steps in the previous proof.

First, we establish the statistical property of the estimator $\hat{\mu}_i^m(t)$. By using Lemma C.4 with respect to the rewards from clients, denoted by set $rw_t = \{s \leq t : \{r_i^j(s)\}_{j\in\mathcal{N}_{m,t}(s),a_j^s=i}\}$ where m is a hub, we derive that the global estimator at the hub $m \in S_0$ meets the following.

Formally, we have that

$$|\hat{\mu}_i^m(t) - \mu^i| \leq (12v)^{\frac{1}{1+\epsilon}}\left(\frac{16\log(e^{1/8}\delta^{-1})}{|rw_t|}\right)^{\frac{\epsilon}{1+\epsilon}} \quad \text{with probability at least } 1-\delta \qquad \forall n \geq 1,$$

where $\hat{\mu}_i^m(t)$ is the median of the means constructed as illustrated in Lemma C.4.

It is worth noting that by definition, the size of rw_t^i, which we denote as $|rw_t^i|$, is the total number of arm pulls of arm i by time t with respect to all clients in the hub.

Meanwhile, we have that the total number of the hub size is the total number of arm pulls of all arms with respect to all clients in the club.

Given the modifications to the algorithm where we add the burn-in period ($L \geq 2\kappa K \log M \log T$), we have the following claim. We consider the following modification to the information transmission. When $S_0^t < a$ (possibly connecting M with T by considering large-scale systems), the transmission between the hub and the non-hub is paused, which implies that $rw_t^i \leq rw_{t-1}^i + 1$. Consider $\tau(t) = \max\{s \leq t : S_0^s \geq a\}$ where $a = M^{\frac{1}{\alpha}-\zeta}$. Equivalently, we have that $rw_t^i \geq a \cdot n_{m,i}(\tau)$. Meanwhile, it is worth noting that the difference between t and τ can be upper bounded by $\kappa \log M \log T$, noting that $|S_0^1|, |S_0^2|, \ldots, |S_0^t|$ are i.i.d random variables, and thus the condition of τ follows a geometric distribution with an exponential decay based on Lemma 3.2 in Section 3.2, with probability $1 - \frac{1}{MT}$ conditional on event $A_{\alpha,\epsilon}$. Consequently, we derive that when $t > L$, on event $A_{\alpha,\epsilon}$ with probability $1 - \frac{1}{MT}$

$$\begin{aligned} rw_t^i &\geq a \cdot n_{m,i}(\tau) \\ &\geq a \cdot (n_{m,i}(t) - \kappa \log M \log T) \\ &\geq a \cdot \frac{1}{2} n_{m,i}(t) \\ &\doteq |S_0| \cdot \frac{1}{2} n_{m,i}(t) \end{aligned}$$

by noting that we have $L \geq 2\kappa K \log M \log T$, and as such $n_{m,i}(t) \geq 2\kappa \log M \log T$ for any arm i.

We define event $A_{\zeta,\delta} = A_{\zeta,\delta} \cap A_{\alpha,\epsilon}$.

Subsequently, we derive that

$$\begin{aligned}
&P(A_{\zeta,\delta}) \\
&\geq 1-(1-P(A_{\zeta,\delta})+1-P(A_{\alpha,\epsilon})) \\
&\geq 1-\frac{1}{M^2}-\frac{\eta}{TM}
\end{aligned}$$

Subsequently, we derive that the following concentration inequality

$$\begin{aligned}
&|\hat{\mu}_i^m(t)-\mu^i| \\
&\leq 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{c\log(1/\delta)}{|rw_t|}\right)^{\frac{\epsilon}{1+\epsilon}} \\
&\leq 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} \text{ with probability at least } 1-\delta \qquad \forall n\geq 1.
\end{aligned}$$

We assert that the factors causing the selection of a sub-optimal arm i are explicitly defined by the decision rule of Algorithm 1. Specifically, the outcome $a_t^m = i$ occurs when any of the following conditions is satisfied:

- Case 1: $\tilde{\mu}_i^m - \mu_i > C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}$,
- Case 2: $-\tilde{\mu}_{i^*}^m + \mu_{i^*} > C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i^*}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}$,
- Case 3: $\mu_{i^*} - \mu_i < 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}$.

Subsequently, we obtain that the number of arm pulls of arm i for client m can be upper bounded as follows:

$$\begin{aligned}
n_{m,i}(T) &\leq l+\sum_{t=L+1}^{T} 1_{\{a_t^m=i, n_{m,i}(t)>l\}} \\
&\leq l+\sum_{t=L+1}^{T} 1_{\{\tilde{\mu}_i^m - C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} > \mu_i, n_{m,i}(t-1)\geq l\}} \\
&\quad + \sum_{t=L+1}^{T} 1_{\{\tilde{\mu}_{i^*}^m + C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i^*}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} < \mu_{i^*}, n_{m,i}(t-1)\geq l\}} \\
&\quad + \sum_{t=L+1}^{T} 1_{\{\mu_i + 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} > \mu_{i^*}, n_{m,i}(t-1)\geq l\}}.
\end{aligned}$$

By taking expected values over $n_{m,i}(t)$ conditional on $A_{\zeta,\delta}$, we derive

$$\begin{aligned}
&E[n_{m,i}(T)|A_{\zeta,\delta}] \\
&= l+\sum_{t=L+1}^{T} P(\tilde{\mu}_i^m - C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} > \mu_i, n_{m,i}(t-1)\geq l|A_{\zeta,\delta}) \\
&\quad + \sum_{t=L+1}^{T} P(\tilde{\mu}_{i^*}^m + C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} < \mu_{i^*}, n_{m,i}(t-1)\geq l|A_{\zeta,\delta})
\end{aligned}$$

$$+\sum_{t=L+1}^{T} P(\mu_i + 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} > \mu_{i^*}, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})$$

$$= l + \sum_{t=L+1}^{T} P(Case1, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \tag{C.6}$$

$$+\sum_{t=L+1}^{T} P(Case2, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) + \sum_{t=L+1}^{T} P(Case3, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \tag{C.7}$$

where $l =$ with $\Delta_i = \mu_{i^*} - \mu_i$..

For the last term in (C.12), we have

$$\sum_{t=L+1}^{T} P(Case3 : \mu_i + 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} > \mu_{i^*}, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})) = 0 \tag{C.8}$$

which also implies that

$$\sum_{t=L+1}^{T} P(Case4 : 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} > \Delta_i, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})) = 0 \tag{C.9}$$

since the choice of l satisfies $l \geq \frac{2cN\log T}{(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}}$ with the choice of $|S_0| = M^{\frac{1}{\alpha}-\zeta}$.

Next, we consider the first two terms, and as in the proof of Theorem 4.1, we subsequently obtain that on event $A_{\zeta,\delta}$

$$\sum_{t=L+1}^{T} P(Case2, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) + \sum_{t=1}^{T} P(Case3, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})$$

$$\leq \sum_{t=L+1}^{T} P(\tilde{\mu}_{m,i} - \mu_i > C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} | A_{\zeta,\delta}) + \tag{C.10}$$

$$\sum_{t=1}^{T} P(-\tilde{\mu}_{m,i^*} + \mu_{i^*} > C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(1/\delta)}{|S_0|\cdot n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} | A_{\zeta,\delta})$$

$$\leq \sum_{t=1}^{T}(\frac{1}{t^2}) + \sum_{t=1}^{T}(\frac{1}{t^2}) \leq \frac{\pi^2}{3} \tag{C.11}$$

where the first inequality utilizes the property of the probability measure when removing the event $n_{m,i}(t-1) \geq l$ and the second inequality holds by the aforementioned concentration inequality based on median-of-means.

As a result, by the above decomposition, we derive that

$$E[n_{m,i}(t) | A_{\zeta,\delta}]$$

$$\leq l + \sum_{t=L+1}^{T} P(Case1, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})$$

$$+\sum_{t=L+1}^{T} P(Case2, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) + \sum_{t=L+1}^{T} P(Case3, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})$$

$$\leq \frac{2c\log T}{|S_0|(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} + \frac{\pi^2}{3}$$

Consequently, we consider the following upper bound on R_T by the previous decomposition, which gives us that

$$\begin{aligned} R_T &\leq L + ((T-L)\cdot \mu_{i^*} - \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{K} n_{m,i}(T)\mu_i^m) \\ &\leq 2\kappa(\log M)^2 \log T + \sum_m \sum_i n_{m,i}(t)\Delta_i \end{aligned}$$

and

$$\begin{aligned} &E[R_T|A_{\epsilon,\delta,\zeta}] \\ &\leq L + (\sum_m (\frac{2c\log T}{|S_0|(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} + \frac{\pi^2}{3}))\cdot\Delta_i \\ &\leq L + M(\frac{2c\Delta_i \log T}{|S_0|(\frac{\Delta_i}{2C\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} + \frac{\pi^2}{3}\Delta_i) = O(\frac{M}{|S_0|}\log T) \end{aligned}$$

where $L \geq 2\kappa K(\log M)^2 \log T$. This completes the first part of the proof.

Next we consider the value of $|S_0|$, by the definition of a, we obtain that $S_0 \geq M^{\frac{1}{\alpha}-\zeta}$. Then by direct computation with $S_0 \geq M^{\frac{1}{\alpha}-\zeta}$, we derive the upper bound on R_T with respect to M and T, which is

$$E[R_T|A_{\epsilon,\delta,\zeta}] \leq O(\frac{M}{|S_0|}\log T) \leq O(M^{1-\frac{1}{\alpha}+\zeta}\log T)$$

which concludes the second part of the proof and thus completes proof of Theorem 4.2.

□

Corollary C.2 (Full information setting)**.** *Let us assume the assumptions in Theorem 4.1 hold, except that we are in a full information setting instead of a partial feedback (bandit) setting. This means that the rewards of all arms are observable at each time step, rather than only the reward of the pulled arm. Under this setting, the same regret bound holds.*

C.3 Proof of Corollary C.2

Proof. Let us consider a full information setting where the clients observe the rewards of all arms.

It is equivalent to implying that $\sum_{t=1}^{T}|S_0^t| = rw_t^i$, by the fact that every time if there is a client in S_0^t, then there is a new sample about arm i that is shared within the hub.

Subsequently, the entire analysis in the proof of Theorem 4.1 comes through, which concludes the proof. □

C.4 Proof of Theorem 5.1

Proof. We would like to emphasize that the graph property we establish for heavy-tailed graphs does not depend on the rewards, and as such, the statements in Section 3 holds. This also implies that, compared to the homogeneous case, the key difference is in the reward aggregation, which we demonstrate in the following.

Again, based on Section 3.2, we have the following result regarding the information delay.

Proposition C.3. *Let us assume that $p = \frac{\eta}{TM}$. Let us further assume that $L > 2\kappa(\log M)^2 \log T$ where L is the length of the burn-in period. Then we obtain with probability at least $1-p$,for $j \notin S_0$,*

$$\min_m n_{m,i}(h_{m,j}^t) \geq \frac{1}{2}\min_m n_{m,i}(t)$$

and

$$\min_m n_{m,i}(t) \geq \frac{1}{2} n_{m,i}(t)$$

and

$$N_{j,i}(t) = n_{m,i}(h^t_{m,j}) \geq n_{m,i}(t) - \kappa(\log M)^2 \log T.$$

With this, we proceed to establish the statistical property of the global estimator $\tilde{\mu}^m_i(t)$.

In the heterogeneous setting, the reward distribution is the same as in the homogeneous setting, which, however, differs in clients even for the same arm. That being said, the following lemma still holds

Lemma C.4 (Lemma 2 of [4])**.** *Let $\delta \in (0,1)$ and $\epsilon \in (0,1]$. Let X_n's be iid copies of X with $\mathbf{E}X = \mu$, and $\mathbf{E}|X-\mu|^{1+\epsilon} \leq v$. Let $k = \lfloor 8\log(e^{1/8}/\delta) \wedge n/2 \rfloor$ and $N = \lfloor n/k \rfloor$. Let*

$$\hat{\mu}_j = \frac{1}{N} \sum_{t=(j-1)N+1}^{jN} X_t \qquad \forall j = 1, 2, \ldots, k,$$

and let $\widehat{\mu}_M$ be the median of $(\hat{\mu}_j)_{j=1,2,\ldots,k}$. Then, with probability at least $1-2\delta$,

$$|\widehat{\mu}_M - \mu| \leq (12v)^{\frac{1}{1+\epsilon}} \left(\frac{16\log(e^{1/8}\delta^{-1})}{n} \right)^{\frac{\epsilon}{1+\epsilon}}.$$

It is worth noting that by using the above lemma, we have the following concentration inequality for the local estimator $\hat{\mu}$ at each client, with $n_{m,i}(t)$ samples,

$$|\hat{\mu}^i_m(n_{m,i}(t); k) - \mu^i_m| \leq (12v)^{\frac{1}{1+\epsilon}} \left(\frac{16\log(e^{1/8}\delta^{-1})}{n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} \quad \text{with probability at least } 1-\delta \qquad \forall n \geq 1.$$

Now we proceed to establish the concentration inequality of the global estimator $\tilde{\mu}^m_i(t)$, which is constructed by

$$\tilde{\mu}^m_i(t+1) = \sum_{j=1}^M P'_t(m,j)\tilde{\mu}^m_{i,j}(h^t_{m,j}) + d_{m,t} \sum_{j \in N_m(t)} \hat{\mu}^m_{i,j}(t) + d_{m,t} \sum_{j \notin N_m(t)} \hat{\mu}^m_{i,j}(h^t_{m,j})$$
$$\text{with } d_{m,t} = \frac{1 - \sum_{j=1}^M P'_t(m,j)}{M}$$

It is worth noting that this global estimator is the weighted average of the local estimators, where the weights can be carefully designed. Precisely, we specify $P'_t(m,j) = \frac{N - M2^{\frac{1}{\epsilon+1}}}{MN2^{\frac{1}{\epsilon+1}}}$ and use mathematical induction to show the following concentration inequality on $\tilde{\mu}^m_i$.

Lemma C.5. *Let us assume that Assumption 1 and 2 hold. Let us assume that $p = \frac{\eta}{TM}$ and $P'_t(m,j) = \frac{N - M2^{\frac{1}{\epsilon+1}}}{MN2^{\frac{1}{\epsilon+1}}}$. Let us further assume that $L > 2\kappa(\log M)^2 \log T$ where L is the length of the burn-in period. Then for any m, i and $t > L$, $\tilde{\mu}_{m,i}(t)$ satisfies that if $n_{m,i}(t) \geq 2(K^2 + KM + M)$, then we have the following hold*

$$P(\tilde{\mu}_{m,i}(t) - \mu_i \geq 2\rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}) \leq \frac{1}{t^2},$$
$$P(\mu_i - \tilde{\mu}_{m,i}(t) \geq 2\rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}) \leq \frac{1}{t^2}.$$

Proof of Lemma C.5. We prove the conclusion through mathematical induction.

First, at the end of the burn-in period, we have that based on Lemma C.4, we derive that when $t = L$

$$\begin{aligned}
&|\tilde{\mu}_i^m(t) - \mu_i| \\
&= |\hat{\mu}_i^m(t) - \mu_i| \leq 2C\rho^{\frac{1}{1+\epsilon}} \left(\frac{c \log(t)}{|rw_t|} \right)^{\frac{\epsilon}{1+\epsilon}} \\
&\text{with probability at least } 1 - \frac{1}{t^2} \qquad \forall n \geq 1.
\end{aligned}$$

where rw_t is the total number of samples being used in computing $\tilde{\mu}_i^m(t)$. In our case, we have $rw_t = \frac{L}{K} = n_{m,i}(t) \geq \min_{m,i} n_{m,i}(t)$. Then we derive that with probability at least $1 - \frac{1}{t^2}$,

$$\begin{aligned}
&|\tilde{\mu}_i^m(t) - \mu_i| \\
&\leq 2C\rho^{\frac{1}{1+\epsilon}} \left(\frac{c \log(t)}{\min_{m,i} n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} \\
&\leq 2\rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc \log(t)}{\min_{m} n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}
\end{aligned}$$

by definition, which proves the statement for $t = L$.

Now let us assume that for any $s \leq t$, we have

$$P(\tilde{\mu}_{m,i}(s) - \mu_i \geq 2\rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc \log(s)}{\min_{m} n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}) \leq \frac{1}{s^2}.$$

At time $t+1$, we have that with $P'_t(m,j) = \frac{N - M2^{\frac{1}{\epsilon+1}}}{MN2^{\frac{1}{\epsilon+1}}}$

$$\begin{aligned}
\tilde{\mu}_i^m(t+1) &= \sum_{j=1}^{M} P'_t(m,j) \tilde{\mu}_{i,j}^m(t_{m,j}) + d_{m,t} \sum_{j \in N_m(t)} \hat{\mu}_{i,j}^m(t) + d_{m,t} \sum_{j \notin N_m(t)} \hat{\mu}_{i,j}^m(t_{m,j}) \\
&\text{with } d_{m,t} = \frac{1 - \sum_{j=1}^{M} P'_t(m,j)}{M}
\end{aligned}$$

and thus the difference between $\tilde{\mu}_i^m(t)$ and μ_i is bounded by

$$\begin{aligned}
&|\tilde{\mu}_i^m(t+1) - \mu_i| \\
&= |\sum_{j=1}^{M} P'_t(m,j) \tilde{\mu}_{i,j}^m(t_{m,j}) + d_{m,t} \sum_{j \in N_m(t)} \hat{\mu}_{i,j}^m(t) + d_{m,t} \sum_{j \notin N_m(t)} \hat{\mu}_{i,j}^m(t_{m,j}) - \mu_i| \\
&= |\sum_{j=1}^{M} P'_t(m,j) (\tilde{\mu}_{i,j}^m(t_{m,j}) - \mu_i) + d_{m,t} \sum_{j \in N_m(t)} (\hat{\mu}_{i,j}^m(t) - \mu_i) + d_{m,t} \sum_{j \notin N_m(t)} (\hat{\mu}_{i,j}^m(t_{m,j}) - \mu_i)| \\
&\leq \sum_{j=1}^{M} P'_t(m,j) |\tilde{\mu}_{i,j}^m(t_{m,j}) - \mu_i| + \\
&\quad d_{m,t} \sum_{j \in N_m(t)} |(\hat{\mu}_{i,j}^m(t) - \mu_i)| + d_{m,t} \sum_{j \notin N_m(t)} |\hat{\mu}_{i,j}^m(t_{m,j}) - \mu_i| \\
&\leq \sum_{j=1}^{M} P'_t(m,j) \cdot 2\rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc \log(s)}{\min_{m} n_{m,i}(t_{m,j})} \right)^{\frac{\epsilon}{1+\epsilon}} +
\end{aligned}$$

$$
\begin{aligned}
&d_{m,t}\sum_{j\in N_m(t)} 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{c\log(t)}{n_{m,j}(t_{m,j})}\right)^{\frac{\epsilon}{1+\epsilon}}\\
&+d_{m,t}\sum_{j\notin N_m(t)} 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{c\log(t)}{n_{m,j}(t_{m,j})}\right)^{\frac{\epsilon}{1+\epsilon}}\\
\le&\sum_{j=1}^{M} P'_t(m,j)\cdot 2\rho^{\frac{1}{1+\epsilon}}\left(\frac{2Nc\log(s)}{\min_m n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}+\\
&d_{m,t}\sum_{j\in N_m(t)} 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(t)}{\min_m n_{m,j}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}\\
&+d_{m,t}\sum_{j\notin N_m(t)} 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{c\log(t)}{2\min_m n_{m,j}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}\\
=&\frac{N-M2^{\frac{1}{\epsilon+1}}}{N2^{\frac{1}{\epsilon+1}}}2\rho^{\frac{1}{1+\epsilon}}\left(\frac{2Nc\log(s)}{\min_m n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}\\
&+M\frac{1-\frac{N-M2^{\frac{1}{\epsilon+1}}}{N2^{\frac{1}{\epsilon+1}}}}{M}2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(t)}{\min_m n_{m,j}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}\\
=&\frac{N-M2^{\frac{1}{\epsilon+1}}}{N2^{\frac{1}{\epsilon+1}}}2\rho^{\frac{1}{1+\epsilon}}\left(\frac{2Nc\log(s)}{\min_m n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}+\\
&(1-\frac{N-M2^{\frac{1}{\epsilon+1}}}{N2^{\frac{1}{\epsilon+1}}})\cdot 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(t)}{\min_m n_{m,j}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}
\end{aligned}
$$

where the second inequality uses the supposition in the mathematical induction, and the concentration inequality for the estimators $\hat{\mu}_i^m$, the third inequality uses Lemma B.8, and the last inequality uses the definition of $P'_t(m,j)$ and $d_{m,t}$.

It is worth noting that when $N > C^{\frac{1+\epsilon}{\epsilon}}$, we have

$$
\begin{aligned}
&2\rho^{\frac{1}{1+\epsilon}}\left(\frac{2Nc\log(s)}{\min_m n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}\\
\ge&\, 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(t)}{\min_m n_{m,j}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}
\end{aligned}
$$

Subsequently, we obtain that

$$
\begin{aligned}
&|\tilde{\mu}_i^m(t+1)-\mu_i|\\
\le&\frac{N-M2^{\frac{1}{\epsilon+1}}}{N2^{\frac{1}{\epsilon+1}}}2\rho^{\frac{1}{1+\epsilon}}\left(\frac{2Nc\log(s)}{\min_m n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}+\\
&(1-\frac{N-M2^{\frac{1}{\epsilon+1}}}{N2^{\frac{1}{\epsilon+1}}})\cdot 2C\rho^{\frac{1}{1+\epsilon}}\left(\frac{2c\log(t)}{\min_m n_{m,j}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}\\
\le&\frac{N-M2^{\frac{1}{\epsilon+1}}}{N2^{\frac{1}{\epsilon+1}}}2\rho^{\frac{1}{1+\epsilon}}\left(\frac{2Nc\log(s)}{\min_m n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}+\\
&(1-\frac{N-M2^{\frac{1}{\epsilon+1}}}{N2^{\frac{1}{\epsilon+1}}})\cdot 2\rho^{\frac{1}{1+\epsilon}}\left(\frac{2Nc\log(t)}{\min_m n_{m,j}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}\\
=&\,2\rho^{\frac{1}{1+\epsilon}}\left(\frac{2Nc\log(t)}{\min_m n_{m,j}(t)}\right)^{\frac{\epsilon}{1+\epsilon}}
\end{aligned}
$$

which subsequently proves the result for $s = t + 1$.

Consequently, we finish the mathematical induction, and conclude the proof of the claim. □

Subsequently, we derive the following upper bound on the number of arm pulls following a different argument compared to the one in the proof of Theorem 4.1, due to the difference in the arm pulling strategy in the algorithm.

We consider the variant of the UCB strategy used in Algorithm 1, and show that there can only be the following 4 possible scenarios when $a_t^m = i$, which means that arm i is pulled by client m:

- Case 1: $n_{m,i}(t) \leq N_{m,i}(t) - 2\kappa \log M \log T$,
- Case 2: $\tilde{\mu}_{m,i} - \mu_i > \rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$,
- Case 3: $-\tilde{\mu}_{m,i^*} + \mu_{i^*} > \rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$,
- Case 4: $\mu_{i^*} - \mu_i < \rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$.

Translating the scenarios into the value of the number of pulls $n_{m,i}(t)$, we obtain

$$
\begin{aligned}
n_{m,i}(T) &\leq l + \sum_{t=L+1}^{T} 1_{\{a_t^m = i, n_{m,i}(t) > l\}} \\
&\leq l + \sum_{t=L+1}^{T} 1_{\{\tilde{\mu}_i^m - \rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} > \mu_i, n_{m,i}(t-1) \geq l\}} \\
&\quad + \sum_{t=L+1}^{T} 1_{\{\tilde{\mu}_{i^*}^m + \rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} < \mu_{i^*}, n_{m,i}(t-1) \geq l\}} \\
&\quad + \sum_{t=L+1}^{T} 1_{\{n_{m,i}(t) < \mathcal{N}_{m,i}(t) - K, a_t^m = i, n_{m,i}(t-1) \geq l\}} \\
&\quad + \sum_{t=L+1}^{T} 1_{\{\mu_i + 2\rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} > \mu_{i^*}, n_{m,i}(t-1) \geq l\}}.
\end{aligned}
$$

We then take the expectation of $n_{m,i}(t)$ conditional on event $A_{\zeta,\delta}$, and derive

$$
\begin{aligned}
&E[n_{m,i}(T)|A_{\zeta,\delta}] \\
&= l + \sum_{t=L+1}^{T} P(\tilde{\mu}_i^m - \rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} > \mu_i, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \\
&\quad + \sum_{t=L+1}^{T} P(\tilde{\mu}_{i^*}^m + \rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} < \mu_{i^*}, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \\
&\quad + \sum_{t=L+1}^{T} P(n_{m,i}(t) < \mathcal{N}_{m,i}(t) - K, a_t^m = i, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \\
&\quad + \sum_{t=L+1}^{T} P(\mu_i + 2\rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)} \right)^{\frac{\epsilon}{1+\epsilon}} > \mu_{i^*}, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})
\end{aligned}
$$

$$= l + \sum_{t=L+1}^{T} P(Case2, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) + \sum_{t=L+1}^{T} P(Case3, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})$$
$$+ \sum_{t=L+1}^{T} P(Case1, a_t^m = i, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) + \sum_{t=L+1}^{T} P(Case4, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) \quad \text{(C.12)}$$

where $l = \max\{\frac{2cN\log T}{(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}}, 2\kappa(\log M)^2 \log T\}$.

We bound the last term in (C.12) by

$$\sum_{t=L+1}^{T} P(Case4 : \mu_i + 2\rho^{\frac{1}{1+\epsilon}} \left(\frac{2Nc\log(t)}{\min_m n_{m,i}(t)}\right)^{\frac{\epsilon}{1+\epsilon}} > \mu_{i^*}, n_{m,i}(t-1) \geq l) = 0 \quad \text{(C.13)}$$

by the fact that $l \geq \frac{2cN\log T}{(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}}$ with $\Delta_i = \mu_{i^*} - \mu_i$.

Considering the first two terms, we obtain that on $A_{\zeta,\delta}$

$$\sum_{t=L+1}^{T} P(Case2, n_{m,i}(t-1) \geq l | A_{\zeta,\delta}) + \sum_{t=1}^{T} P(Case3, n_{m,i}(t-1) \geq l | A_{\zeta,\delta})$$
$$\leq \sum_{t=L+1}^{T} P(\tilde{\mu}_{m,i} - \mu_i > \frac{2cN\log T}{(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} | A_{\zeta,\delta})$$
$$+ \sum_{t=1}^{T} P(-\tilde{\mu}_{m,i^*} + \mu_{i^*} > \frac{2cN\log T}{(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}} | A_{\zeta,\delta})$$
$$\leq \sum_{t=1}^{T} (\frac{1}{t^2}) + \sum_{t=1}^{T} (\frac{1}{t^2}) \leq \frac{\pi^2}{3} \quad \text{(C.14)}$$

where the first inequality is true when $n_{m,i}(t-1) \geq l$ and the second inequality results from Lemma C.5.

For Case 1, we note that Lemma B.8 implies that

$$n_{m,i}(t) > N_{m,i}(t) - 2\kappa(\log M)^2 \log T$$

with the definition of $N_{m,i}(t+1) = \max\{n_{m,i}(t+1), N_{j,i}(t), j \in \mathcal{N}_m(t)\}$.

Based on the observation that the difference between $N_{m,i}(t)$ and $n_{m,i}(t)$ is at most $2\kappa(\log M)^2 \log T$, we next show the exact time steps we need to explore in order to make sure $-n_{m,i}(t) + N_{m,i}(t)$ to be smaller than $2\kappa(\log M)^2 \log T$.

At time step t, if Case 1 holds for client m, then $n_{m,i}(t+1)$ increases by 1 based on $n_{m,i}(t)$. The following discussion characterizes the change in $N_{m,i}(t+1)$. For client m, if $n_{m,i}(t) \leq \mathcal{N}_{m,i}(t) - 2\kappa(\log M)^2 \log T$, the value of $N_{m,i}(t+1)$ remains unchanged, as defined by $N_{m,i}(t+1) = \max\{n_{m,i}(t+1), N_{j,i}(t) : j \in \mathcal{N}_m(t)\}$.

Additionally, for any client $j \in \mathcal{N}_m(t)$ such that $n_{j,i}(t) < \mathcal{N}_{j,i}(t) - 2\kappa(\log M)^2 \log T$, the value $\mathcal{N}_{j,i}(t+1)$ is not affected since $n_{j,i}(t+1) \leq n_{j,i}(t)+1$. Consequently, such clients do not influence the value of $N_{m,i}(t+1)$, which remains defined as $N_{m,i}(t+1) = \max\{n_{m,i}(t+1), N_{j,i}(t) : j \in \mathcal{N}_m(t)\}$. Now, Let us consider a client $j \in \mathcal{N}_m(t)$ with $n_{j,i}(t) > \mathcal{N}_{j,i}(t) - 2\kappa(\log M)^2 \log T$. If such a client does not sample arm i, the value of $N_{j,i}(t)$ remains unchanged, leading to a decrease of 1 in the difference $-n_{m,i}(t) + N_{m,i}(t)$. On the other hand, if this client samples arm i, $N_{m,i}(t)$ increases by 1, keeping the difference between $n_{m,i}(t)$ and $N_{m,i}(t)$ unchanged. However, this scenario falls under Cases 2 and 3, whose total duration has already been upper-bounded by $\frac{\pi^2}{3}$, as shown in (C.14).

Subsequently, we establish that the time frame of the exploration does not exceed $2\kappa(\log M)^2 \log T + \frac{\pi^2}{3}$, i.e.

$$\begin{aligned}
&\sum_{t=1}^{T} P(Case1, a_t^m = i, n_{m,i}(t-1) \geq l | A) \\
&\leq 2\kappa(\log M)^2 \log T + \frac{\pi^2}{3}.
\end{aligned} \tag{C.15}$$

Consequently, we establish that

$$\begin{aligned}
&E[n_{m,i}(T)|A_{\zeta,\delta}] \\
&\leq l + \frac{\pi^2}{3} + 2\kappa(\log M)^2 \log T + \frac{\pi^2}{3} + 0 \\
&= l + \frac{2\pi^2}{3} + 2\kappa(\log M)^2 \log T \\
&= \max\{\frac{2cN\log T}{(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}}, 2(K^2+MK+M)\} + \frac{2\pi^2}{3} + 2\kappa(\log M)^2 \log T
\end{aligned}$$

where the inequality results from (C.12), (C.13), (C.14), and (C.15).

Then again, we consider the aforementioned regret decomposition (which holds by definition and thus does not rely on any reward property), which gives us that

$$\begin{aligned}
R_T &\leq L + ((T-L)\cdot \mu_{i^*} - \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{K} n_{m,i}(T)\mu_i^m) \\
&\leq 2\kappa(\log M)^2 \log T + \sum_m \sum_i n_{m,i}(t)\Delta_i.
\end{aligned}$$

By taking the expected value of R_T given event $A_{\zeta,\delta}$, we derive that

$$\begin{aligned}
&E[R_T|A_{\zeta,\delta}] \\
&\leq 2\kappa(\log M)^2 \log T + \sum_m \sum_i E[n_{m,i}(t)|A_{\epsilon,\delta}]\Delta_i \\
&\leq 2\kappa(\log M)^2 \log T + \\
&\quad \sum_m \sum_i \Delta_i \cdot (\max\{\frac{2cN\log T}{(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}}, 2(K^2+MK+M)\} + \frac{2\pi^2}{3} + 2\kappa(\log M)^2 \log T) \\
&\leq 2\kappa(\log M)^2 \log T + \\
&\quad \sum_i M\Delta_i \cdot (\max\{\frac{2cN\log T}{(\frac{\Delta_i}{2\rho^{\frac{1}{1+\epsilon}}})^{\frac{1+\epsilon}{\epsilon}}}, 2\kappa\log M\log T\} + \frac{2\pi^2}{3} + 2\kappa(\log M)^2 \log T).
\end{aligned}$$

This concludes the proof of Theorem 5.1. □