Federated Combinatorial Causal Bandits with Heterogeneous Causal Influences

Zheshun Wu[1,2] **Wei Chen**[3] **Zenglin Xu**[4,5,*] **Fang Kong**[1,*]

[1]Southern University of Science and Technology, [2]Harbin Institute of Technology (Shenzhen), [3]Microsoft Research Asia
[4]Fudan University, [5]Shanghai Academy of AI for Science
wuzhsh23@gmail.com, weic@microsoft.com, *zenglin@gmail.com, *kongf@sustech.edu.cn

Abstract

We explore the problem of federated combinatorial causal bandits (FedCCB), where multiple agents collaboratively select variables for intervention and gather feedback. The primary objective of FedCCB is to identify optimal interventions for each agent while minimizing the total cumulative regret associated with the target nodes. A key challenge in FedCCB stems from the inherent heterogeneity of local causal models, which often exhibit diverse causal influences. To address this challenge, we propose a novel Federated Subset-Clustered Bandit (FedSCuB) method. This method groups agents to tackle heterogeneity based on the similarity of their causal relationships with specified subsets of variables. FedSCuB incorporates an intervention-based exploration strategy to collect partial observations necessary for clustering, as well as an alternating minimization method to facilitate collaboration among agents within the same cluster. Theoretical analysis demonstrates that the proposed FedSCuB method achieves sub-linear regret and offers a better regret bound compared to baseline methods. Empirical evaluations on synthetic tasks further confirm the effectiveness and superiority of our method.

1 INTRODUCTION

The causal bandit (CB) problem, first introduced by Lattimore et al. [2016], provides a formal framework for sequential interventions in a Bayesian network of causally interacting variables [Yabe et al., 2018a, Lee and Bareinboim, 2018, Xiong and Chen, 2023]. This framework is defined using a causal graph $\mathcal{G} = \{\mathcal{V} = (\mathbf{X}, Y), \mathcal{E}\}$, which characterizes the causal relationships $\mathcal{E}$ among variables $\mathcal{V}$ in the network. While the structure of $\mathcal{G}$ is usually known to the learning agents, the underlying probability distribution of the causal model is typically unknown. The objective of the CB problem is to identify an optimal intervention that maximizes the output at a target node Y. CB has found applications in diverse domains, including drug discovery [Liu et al., 2020], computational advertising [Varici et al., 2023a, Lu et al., 2020], economic decision-making [Kanase et al., 2022, Nair et al., 2021], and so forth.

In many practical scenarios, learning agents need to intervene in multiple variables simultaneously. For instance, physicians administer multiple treatments to patients, such as prescribing medications and ordering laboratory tests, to maximize the effectiveness of rehabilitative treatment [Liu et al., 2020, Kim et al., 2023]. However, simply treating each possible combination of variables as a distinct action results in an exponential increase in computational cost, making the problem computationally intractable in high-dimensional settings. To address this challenge, the more general combinatorial CB (CCB) problem has been proposed [Feng and Chen, 2023, Feng et al., 2023]. CCB extends the CB framework by efficiently handling interventions involving combinations of variables.

Most existing studies on CCB focus on designing algorithms for a single agent, assuming centralized data availability [Feng and Chen, 2023, Xiong and Chen, 2023]. However, with the increasing amount of data being distributed across agents and rising privacy concerns, there is a growing need for decentralized approaches that enable CCB on the agents' side. To reduce individual regret, decentralized agents often hope to collaboratively learn the causal relationships. Federated learning (FL) offers a promising framework for such collaboration [McMahan et al., 2017, Kairouz et al., 2021, Li et al., 2020b, 2024, Wu et al., 2024] and it has recently garnered significant attention in the bandit literature [Wang et al., 2020, Huang et al., 2021, He et al., 2022]. With the growing demand for collaboratively learning causal relationships in scenarios such

*Corresponding author.

as multiple companies collaborating to identify the optimal advertising strategy that maximizes user click-through rates, FL shows potential to enhance CCB applications in distributed settings.

In this paper, we present a new federated CCB (FedCCB) framework. While multiple agents independently perform interventions on their local causal models, they can collaborate on parameter estimation to accelerate their individual learning processes. Motivated by the observation that different agents may share distinct causal influences such as user intentions diverge among advertising companies, we focus on the FedCCB setting with heterogeneous causal influences, i.e., the underlying probability distributions of local causal models differ across agents. To maximize the benefits of collaboration while avoiding misleading information from heterogeneous agents, we introduce a finer-grained cluster structure specifically designed for causal bandits. Unlike existing works on clustering bandits which typically rely on coarser clustering based on the entire variable set [Gentile et al., 2014, 2017, Liu et al., 2022, Blaser et al., 2024], our approach enables agents to cluster based on the similarity of causal relationships within specified subsets of variables. This more general structure allows for greater flexibility and has the potential to improve learning efficiency in the presence of diverse causal influences.

New challenges arise to deal with this general structure. Specifically, to perform homogeneity test-based clustering on a subset of variables, it is crucial to isolate the partial response attributable to the corresponding subset. However, this operation is generally infeasible in practical implementations. For example, an advertising platform can only observe the click-through rate of an ad placement, which is influenced by a combination of factors such as predictive algorithms and bidding strategies [Bottou et al., 2013]. Hence, we can only observe the overall response value, which encapsulates the combined contributions of all covariates. Moreover, even if the ground-truth clusters are accurately identified, collaboration between agents still remains challenging because agents cannot share sufficient statistics derived from partial observations that are unavailable to them.

To address these challenges, we propose a novel Federated Subset-Clustered Bandit (FedSCuB) method. For addressing the first challenge, we leverage the properties of causal bandits to propose an intervention-based independent exploration strategy. In the initial stage of our framework, all agents sequentially intervene on subsets of parent variables. This process enables agents to directly infer the partial response values contributed by the non-intervened covariates. The partial observations are then utilized to perform subset-level clustering estimation effectively. For tackling the second challenge, we adopt an alternating minimization optimization technique to isolate the influence of other variables when estimating the causal relationship from a specified subset. Agents can thus estimate the influence from each subset of variables and collaborate with others by sharing information derived from the subset-level analysis.

Our contributions can be summarized as follows:

- We take the first step in formalizing the problem of federated combinatorial causal bandits (FedCCB) with heterogeneous causal influences. To effectively capture distribution heterogeneity in FedCCB and enhance the collaborative effect, we introduce a finer-grained cluster structure among agents based on the specified subsets of variables.
- To tackle the challenge of unobservable partial observations arising from this fine-grained cluster structure, we propose a novel Federated Subset-Clustered Bandit (FedSCuB) method including an intervention-based independent exploration strategy for the clustering estimation phase, and an alternating minimization approach during the collaboration phase.
- We prove that our proposed FedSCuB achieves the regret of $\tilde{\mathcal{O}}(\sum_X |\boldsymbol{Pa}(X)|^2 \sum_l \sum_k \sqrt{T|C_k||\boldsymbol{Pa}^{(l)}(X)|})$, where $\boldsymbol{Pa}(X)$ denotes the parents of node X, $\boldsymbol{Pa}^{(l)}(X)$ denotes the l-th subset of $\boldsymbol{Pa}(X)$, and C_k denotes the k-th ground-truth cluster at the level of corresponding subset. This theoretical result demonstrates significant advantages over existing baselines. Furthermore, the empirical superiority of the proposed algorithm has been confirmed through a series of experiments.

2 PROBLEM FORMULATION

In this section, we formulate the problem setting of federated combinatorial causal bandits (FedCCB). Consider a learning system consisting of M agents with their associated causal models, and a central server responsible for coordinating the communication between agents for collaborative model estimation. In the following, we first provide an overview of a causal model. A causal model includes a causal graph, a cause-effect statistical model, and an interventional model respectively [Pearl, 2009].

A *causal graph* $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is a directed acyclic graph (DAG), where $\mathcal{V} = \mathbf{X} \cup \{Y\}$ denotes the set of nodes. Y is a special target node with no outgoing edges, and $\mathcal{E}$ is the set of directed edges. For simplicity, in this paper, we consider all variables in $\mathcal{V}$ to be continuous random variables such that $X \in [0, 1], \forall X \in \mathcal{V}$, and all of them can be observed by the learning agent. For a node X in the graph $\mathcal{G}$, we call its in-neighbor nodes in $\mathcal{G}$ the *parents* of X, denoted as $\boldsymbol{Pa}(X)$, and the values taken by these parent random variables are denoted as $\boldsymbol{pa}(X)$. We denote the size of $\mathcal{V}$ as N and the maximum in-degree in $\mathcal{G}$ as $D := \max_{X \in \mathcal{V}} |\boldsymbol{Pa}(X)|$ below.

Following the causal Bayesian model, the causal influence from the parents to X is modeled as the probability distribution $P(X|\boldsymbol{Pa}(X))$. In many real-world applications, although the causal relationships are usually consistent across different environments, the strength of these influences may vary. Based on this observation, we consider the distribution heterogeneity among agents. Specifically, we assume the topology of the underlying causal graph $\mathcal{G}$ is consistent among all agents $m \in [M]$ but the conditional distribution $P(X|\boldsymbol{Pa}(X))$ for each node $X \in \mathcal{V}$ can be different. In the following, the random variable corresponding to node $X \in \mathcal{V}$ in the m-th agent's local causal model is denoted as X_m, and the corresponding conditional distribution is denoted as $P(X_m|\boldsymbol{Pa}(X_m)), \forall m \in [M]$.

Notice that the non-parametric form of distribution needs an exponential number of quantities and is difficult to learn since this is a functional optimization problem. Hence, in this paper, we adopt the linear model as the parametric model [Varici et al., 2023a, Yan et al., 2024]. For each agent $m \in [M]$, we assume that the propagation relationship between X_m and its parents $\boldsymbol{Pa}(X_m)$ is defined as $X_m = \theta_{X_m} \cdot \boldsymbol{pa}(X_m) + \eta_{X_m}$, where θ_{X_m} is a unknown parameter vector in $[0,1]^{|\boldsymbol{Pa}(X)|}$. η_{X_m} is a zero-mean σ_X sub-Gaussian noise that ensures that $X_m \in [0,1]$ [Feng and Chen, 2023, Xiong and Chen, 2023], where $\sigma_X > 0$. We further note that this assumption can be generalized to settings with negative parameter values (e.g., $[-1,1]$), as such cases simply involve a linear transformation of $[0,1]$. Crucially, this extension preserves the theoretical foundations of both clustered bandits and combinatorial causal bandits.

Regarding the intervention model, we adopt the modeling framework introduced in Feng and Chen [2023], Xiong and Chen [2023]. Specifically, in this paper, an *intervention* in the causal model is to force a subset of variables $\mathbf{S} \subset \mathbf{X}$ to take some predetermined values $\mathbf{s}$, and to see its effect on the target variable Y. The intervention on $\mathbf{S}$ to Y is denoted as $\mathbb{E}[Y|do(\mathbf{S} = \mathbf{s})]$. In this paper, we assume that agents can set a variable $X \in \mathbf{X}$ to $x \in [0,1]$. The parametric model defined above has the monotonicity property. Therefore, setting a variable to 1 is always better setting it for other values in maximizing $\mathbb{E}[Y]$, so it is natural to set $\mathbf{S}$ to all ones, for which we simply denote as $\mathbb{E}[Y|do(\mathbf{S})]$.

Heterogeneous Local Causal Models. Existing works typically cluster agents based on the similarity of their entire parameter vectors [Blaser et al., 2024, Yang et al., 2024], which can be too coarse for CCB. Consider a FedCCB system where multiple online advertising platforms aim to identify their optimal ad placement strategies. One platform may share similar user intention with certain peers, and similar ad inventory with othersbut not necessarily both Varici et al. [2023a]. Clustering based on the similarity of the entire parameter vector, as in existing methods [Blaser et al., 2024, Li et al., 2021, Liu et al., 2022], would prevent these platforms from sharing information effectively. In contrast, we argue that platforms with similar ad inventory or user profiles can share information regarding these factors to speed up learning about their impact on ad placement. Building on these observations, we propose a finer-grained clustering approach: platforms can group variables into subsets based on broader factors like user attributes and ad infrastructure. Clustering then focuses on agent similarity within these subsets. Our approach is flexible and can also be extended to scenarios where such predefined subsets do not exist. In such cases, agents can cluster and share information at the level of individual variables. This finer-grained structure allows for more precise and effective collaboration, enhancing the efficiency of federated CCB in diverse settings.

Cluster Structure. For any node $X \in \mathcal{V}$ in the graph, agents have a pre-agreed partition that groups $\boldsymbol{Pa}(X)$ into L_X subsets, i.e., $\boldsymbol{Pa}(X) = \cup_{l\in[L_X]}\boldsymbol{Pa}^{(l)}(X)$, where $1 \leq |\boldsymbol{Pa}^{(l)}(X)| \leq |\boldsymbol{Pa}(X)|$ and $\sum_{l\in[L_X]} |\boldsymbol{Pa}^{(l)}(X)| = |\boldsymbol{Pa}(X)|$. For any node $X \in \mathcal{V}$, similar causal influences in $\{\theta_X^{(l)}\}_{l\in[L_X]}$ from distinct agents form clusters respectively. The clusters are represented as $\mathcal{C}_X^{(l)} = \{C_1, C_2, ..., C_{|\mathcal{C}_X^{(l)}|}\}$. We highlight that $\theta_X^{(l)}$ within the parameter vector θ_X corresponds to the subset $\boldsymbol{Pa}^{(l)}(X) \subset \boldsymbol{Pa}(X)$, for any subset $l \in [L_X]$. Notice that if the number of subsets meets $L_X = 1$, the cluster structure is defined in terms of the overall parameter vector, which was commonly considered in previous clustered bandits studies. On the other hand, we will conduct component-wise clustering if $L_X = |\boldsymbol{Pa}(X)|$. Consequently, for each node $X \in \mathcal{V}$ and each agent $m \in [M]$, we have $X_m = \sum_{l\in[L_X]} \theta_{X_m}^{(l)} \cdot \boldsymbol{pa}^{(l)}(X_m) + \eta_{X_m}$. The set of parents thus satisfy $\boldsymbol{Pa}(X) = \cup_{l\in L_X}\boldsymbol{Pa}^{(l)}(X), \forall X \in \mathcal{V}$ and we assume $\boldsymbol{Pa}^{(l)}(X) \cap \boldsymbol{Pa}^{(l')}(X) = \emptyset, \forall l \neq l' \in [L_X]$. In this paper, we assume that the partition of parents for any node is known to each agent, while the cluster structure w.r.t. these partitions are unknown to both agents and the central server. **This assumption can be readily relaxed to the setting of unknown partitions since the clustering can always be performed at the level of individual parameter components.** We illustrate this structure through a toy example, as depicted in Figure 1.

The goal of this paper is to discover the component-subset level cluster structure among agents to achieve effective collaborations. Consistent with Yang et al. [2024], the ground-truth clusters $\{C_k\}_{k\in[|\mathcal{C}_X^{(l)}|]}$ in $\mathcal{C}_X^{(l)}$ are considered to be disjoint, as the distinctions across various clusters are significant in practice [Gentile et al., 2014, Liu et al., 2022, Yang et al., 2024]. We denote the size of $\mathcal{C}_X^{(l)}$ by $K_X^{(l)}$ in the rest of this paper. In round t, for any node $X \in \mathcal{V}$ in agent m's causal model, we define $V_{X_m}(t)$ as the observation of its parents $\boldsymbol{pa}(X)$, and $X_m(t)$ as the corresponding propagating result of node X. Furthermore, we define $V_{X_m}^{(l)}(t)$ as the observation of the subset $l \in [L_X]$ of the parents $\boldsymbol{pa}(X)$ in

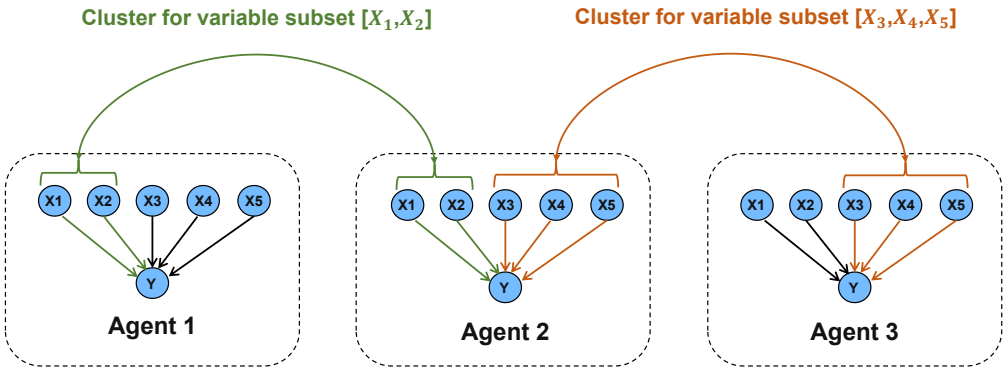


Figure 1: A FedCCB system with 3 agents, each associated with a causal graph sharing a common topology. The causal graph consists of 5 parent nodes, all pointing to the target node. This example demonstrates heterogeneous causal influences: agents 1 and 2 share identical edge weights for $\{X_1, X_2\}$ pointing to Y, while agents 2 and 3 share common edge weights for $\{X_3, X_4, X_5\}$ pointing to Y.

round t. The propagating result $X_m^{(l)}(t)$ contributed by this subset is defined as $X_m^{(l)}(t) = \theta_{X_m}^{(l)} \cdot V_{X_m}^{(l)}(t)$. We highlight that $X_m^{(l)}(t)$ is actually unobservable in practice.

We adopt the following assumptions widely used in existing clustering works [Gentile et al., 2014, 2017, Li et al., 2021, Blaser et al., 2024] to characterize our considered subset-level cluster structure.

Assumption 2.1 (Proximity within clusters)**.** For any node $X \in \mathcal{V}$ and any subset of parents $\boldsymbol{Pa}^{(l)}(X), l \in [L_X]$, if any two agents $m, m' \in [M]$ belong to a specific cluster in $\mathcal{C}_X^{(l)}$, we have $\|\theta_{X_m}^{(l)} - \theta_{X_{m'}}^{(l)}\| \leq \epsilon \leq \frac{1}{M\sqrt{DT}}$.

Assumption 2.2 (Separateness among clusters)**.** For any node $X \in \mathcal{V}$ and any subset of parents $\boldsymbol{Pa}^{(l)}(X), l \in [L_X]$, we have $\|\theta_{X_m}^{(l)} - \theta_{X_{m'}}^{(l)}\| \geq \gamma \geq 0, \forall m \in C_i, m' \in C_j$, for any two clusters $C_i, C_j \in \mathcal{C}_X^{(l)}$.

The FedCCB Setting. In practice, identifying cause-effect relationships is easier than measuring the exact causal influences between variables in causal bandits [Feng and Chen, 2023, Yabe et al., 2018b, Varici et al., 2023a]. This is because causal bandits commonly assume that agents can perform interventions on nodes of the causal graph. Even when operating under complete uncertainty about the topology, agents can acquire interventional data to discover it. Hence, we mainly focus on identifying the latter and assume that M agents exactly know the common topology structure of DAG, but do not know the distributions $\{P(X_m|\boldsymbol{Pa}(X_m))\}_{m\in[M]}$ for all the nodes in graph $\mathcal{G}$.

In a synchronized fashion of FedCCB, each agent $m \in [M]$ is active in each round $t \in [T]$ and selects at most K variables in $\mathbf{X}$ for intervention. Then the agent obtains the observed Y value as the reward, and observes all variable values in $\mathcal{V}$ as the feedback. The goal of agents is to maximize their respective cumulative reward from all T rounds. Hence, the performance of synchronous FedCCB can be measured by the total cumulative regret defined as

$$R(T) := \mathbb{E}\Big[\sum_{t\in[T]}\sum_{m\in[M]}\big(\mathbb{E}[Y_m|do(\mathbf{S}_m^*)] - \mathbb{E}[Y_m|do(\mathbf{S}_m(t))]\big)\Big], \tag{1}$$

where $\mathbf{S}_m^* \in \arg\max_{\mathbf{S}\subset\mathbf{X},|\mathbf{S}|=K} \mathbb{E}[Y_m|do(\mathbf{S})]$ and $\mathbf{S}_m(t)$ is the intervention set selected by agent m in round t. Taking the collaboration among advertisers as an example, we can consider that advertisers serve as agents that sequentially select and display ads to users based on historical performance. Here, the model parameters quantify the causal effects of interventions on recovery outcomes.

3 METHODOLOGY

In this section, we introduce our proposed FedSCuB method for solving the problem of FedCCB. This method is a three-phase clustered bandit algorithm. As discussed above, the main challenge of performing clustering at the level of component-subset is that we actually can not obtain the partial observations of causal variable contributed by subsets of parents. Consequently, we can not naively use the overall response value of node to estimate parameter vectors for conducting homogeneity test-based clustering proposed in Blaser et al. [2024].

To address this challenge, an intervention-based independent exploration method is provided in this study.

3.1 INDEPENDENT EXPLORATION PHASE

In this phase, agents independently execute pure exploration of length $T_0 = \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]} T_X^{(l)}$, where $T_X^{(l)}$ denotes the rounds needed for guaranteeing the accuracy

of the homogeneity test for the node $X \in \mathcal{V}$ and the subset $l \in [L_X]$ of its parents. Different with Blaser et al. [2024] requiring agents to select actions by uniformly sampling, we provide an intervention-based exploration method to obtain partial observations needed for clustering agents at the level of component subset.

During this exploration stage, each agent $m \in [M]$ should construct datasets $\mathcal{H}^{(l)}_{X_m}(t) = \{V^{(l)}_{X_m}(\tau), X^{(l)}_m(\tau)\}_{\tau\in[t]}$, where the key challenge is to assess the unobservable $X^{(l)}_m(\tau)$. The core idea of our method is to first observe the natural realization of X without any intervention, and then repeatedly observe it after intervening on the parent subset $\boldsymbol{Pa}^{(l)}(X)$. The difference between these two realizations provides the information we need. In this paper we assume that the intervention budget satisfies $K \geq \max_{X\in\mathcal{V},l\in[L_X]} |\boldsymbol{Pa}^{(l)}(X)|$, which is reasonable since L_X can be determined by the agents themselves arbitrarily. Specifically, agent m first observes the non-intervened realization of $X_m(t)$. Then the agent selects the intervention set $\mathbf{S}_m(t) = \boldsymbol{Pa}^{(l)}(X_m)$, and intervenes by setting all nodes in $\mathbf{S}_m(t)$ to zero. The difference between two realizations of X_m, defined as $X^{(l)}_m(t) := \theta^{(l)}_{X_m} \cdot \boldsymbol{Pa}^{(l)}(X_m) + \sum_{l'\neq l} \theta^{(l')}_{X_m} \cdot (\boldsymbol{Pa}^{(l')}(X_m) - \tilde{\boldsymbol{Pa}}^{(l')}(X_m)) + \eta_{X_m} - \tilde{\eta}_{X_m}$, has mean $\mathbb{E}\big[\theta^{(l)}_{X_m} \cdot \boldsymbol{Pa}^{(l)}(X_m)\big]$ and noise variance $\mathcal{O}(\sigma^2_X + L^2_X)$. Below we define $\tilde{\sigma}^2_X := \mathcal{O}(\sigma^2_X + L^2_X)$. Then agent updates its sufficient statistics $M^{(l)}_{X_m}(t) += V^{(l)}_{X_m}(t)(V^{(l)}_{X_m}(t))^\top$, $b^{(l)}_{X_m}(t) += V^{(l)}_{X_m}(t)X^{(l)}_m(t)$ and uploads these statistics to the central server after obtaining sufficient observations. We present the details of this phase in Appendix C.

3.2 CLUSTER ESTIMATION

In this paper, we follow the idea proposed in Blaser et al. [2024] of clustering similar parameters (ϵ defined in Assumption 2.1). The key distinction is that we further conduct the more fine-grained component-subset-wise clustering to broaden the scope of beneficial collaboration. We implement this by testing whether $\|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| \leq \epsilon$ via the homogeneity test provided in Blaser et al. [2024], Li et al. [2021]. This homogeneity test determines whether the parameters of linear regression models associated with two datasets are similar under the equal variance. Since $\{\theta^{(l)}_{X_m}\}_{m\in[M]}$ are not observable, this test has to be performed on their estimates, for which maximum likelihood estimator (MLE) is a prevalent choice [Chow, 1960].

Below we provide the intuition behind our cluster estimation method, while the full procedure is detailed in Appendix C. Based on uploaded sufficient statistics, the server can obtain the MLE $v^{(l)}_{X_m} = (M^{(l)}_{X_m}(t))^- b^{(l)}_{X_m}(t)$ of $\theta^{(l)}_{X_m}$, for any node and any subset of its parents. We apply the following test statistics: $s(\mathcal{H}^{(l)}_{X_m}(t), \mathcal{H}^{(l)}_{X_n}(t)) := \frac{\|\mathbf{V}^{(l)}_{X_m}(v^{(l)}_{X_m} - v^{(l)}_{X_m,X_n})\|^2 + \|\mathbf{V}^{(l)}_{X_n}(v^{(l)}_{X_n} - v^{(l)}_{X_m,X_n})\|^2}{\tilde{\sigma}^2_X}$, where $v^{(l)}_{X_m,X_n}$ denotes the estimator using data from $\mathcal{H}^{(l)}_{X_m}(t)$ and $\mathcal{H}^{(l)}_{X_n}(t)$. $\mathbf{V}^{(l)}_{X_m}$ denotes the design matrix, where each row of $\mathbf{V}^{(l)}_{X_m}$ is the covariate vector corresponding to subset $\boldsymbol{Pa}^{(l)}(X)$. When $s(\mathcal{H}^{(l)}_{X_m}(t), \mathcal{H}^{(l)}_{X_n}(t))$ is above a selected threshold $\nu^{(l)}_X$ (will be introduced in Theorem 4.1), it shows that the aggregated estimator deviates heavily from the individual estimators on two datasets. Hence, for any node and any subset of its parents, to determine the corresponding sets of agents that can collaborate, the central server performs this pairwise homogeneity test among each pair of agents. We present the workflow of cluster estimation phase in Algorithm 1.

Algorithm 1 Cluster Estimation

1: **for** $X \in \mathcal{V}$ and $l \in [L_X]$ **do**
2: **for** $m \neq n \in [M]$ **do**
3: **if** $s(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_X), \mathcal{H}^{(l)}_{X_n}(T^{(l)}_X)) \leq \nu^{(l)}_X$ **then**
4: add edge $e(m,n)$ to $\mathcal{G}^{(l)}_X$.
5: $\hat{\mathcal{C}}^{(l)}_X = \{\hat{C}_k\}_{k=1}^{\hat{K}^{(l)}_X} \leftarrow$ maximal-cliques$(\mathcal{G}^{(l)}_X)$.
6: For each agent m, set $\mathcal{K}^{(l)}_{X_m} = \hat{C}_k$, if $m \in \hat{C}_k$.

3.3 COLLABORATIVE LEARNING PHASE

Upon identifying clusters of agents at the level of component subset, our system proceeds to the collaborative learning phase. In this phase, agents optimistically select nodes to intervene, and collaborate with other similar agents to enhance their local learning performance. Since cluster structures are different for distinct nodes and subsets of their parents, we utilize an Alternating Minimization (AM) method presented in Li and Wang [2022a] to separately update different component subsets of parameter vectors. This method can achieve the effective collaboration among agents in the same cluster. We introduce the procedure of AM method below.

Alternating Minimization. At each round $t \in [T_0 + 1, T]$, after each agent observes new feedback $(\mathbf{X}_m(t), Y_m(t))$, it alternates among the following $[L_X]$ steps for each node $X \notin \mathbf{S}_m(t)$ which is not intervened in this round:

$$\hat{\theta}^{(l)}_{X_m}(t) = \Pi_{\mathbb{B}^{|\boldsymbol{Pa}^{(l)}(X)|}}\Big(\big(M^{(l)}_{X_m}(t) + V^{(l)}_{X_m}(t)(V^{(l)}_{X_m}(t))^\top\big)^- \big(b^{(l)}_{X_m}(t) + V^{(l)}_{X_m}(t)\hat{X}^{(l)}_m(t)\big)\Big), \forall l \in [L_X] \tag{2}$$

where Π denotes the projection operation and $\mathbb{B}^{|\boldsymbol{Pa}^{(l)}(X)|} := [0,1]^{|\boldsymbol{Pa}^{(l)}(X)|}$. In this phase, we estimate the unobservable partial response $X^{(l)}_m(t)$ by the following way: $\hat{X}^{(l)}_m(t) = X_m(t) - \sum_{j\in[L_X]\setminus\{l\}} V^{(j)}_{X_m}(t) \cdot \hat{\theta}^{(j)}_{X_m}(t)$, where $\{\hat{\theta}^{(l)}_{X_m}(t)\}$ are initialized by the MLE obtained in the clustering stage.

Algorithm 2 Clustering-based Linear Regression for FedCCB with Linear Models

1: **Input:** Graph $\mathcal{G}$, intervention budget K.
2: Initialize $M_{X_m}^{(l)}(T_0) \leftarrow \lambda \mathbf{I} \in \mathbb{R}^{|\boldsymbol{Pa}^{(l)}(X)|\times|\boldsymbol{Pa}^{(l)}(X)|}$, $b_{X_m}^{(l)}(T_0) \leftarrow 0^{|\boldsymbol{Pa}^{(l)}(X)|}$, and $\rho_{X_m}^{(l)}(t) \leftarrow (\sigma_X + 2|\boldsymbol{Pa}^{(l)}(X)|)\sqrt{|\boldsymbol{Pa}^{(l)}(X)|\log(1+t|\boldsymbol{Pa}^{(l)}(X)|)+2\log\frac{1}{\delta}}$, for $X \in \mathcal{V}$ and $l \in [L_X]$.
3: **for** $t = T_0+1, ..., T$ **do**
4: **for** agent $m \in [M]$ **do**
5: Compute the confidence ellipsoid for any node X and any subset $l \in [L_X]$ of its parents.
6: $\mathbf{S}_m(t), \tilde{\theta}_m(t) = \arg\max_{\mathbf{S}\subset\mathbf{X},|\mathbf{S}|=K,\{\theta'^{(l)}_{X_m}\in \mathrm{CB}^{(l)}_{X_m}(t)\}_{X\in\mathcal{V},l\in[L_X]}} \mathbb{E}[Y_m|do(\mathbf{S})]$.
7: Intervene all the nodes in $\mathbf{S}_m(t)$ to 1 and observe the feedback.
8: Run AM by Eq. (2) to estimate partial observations.
9: **for** $X \in \mathcal{V}$ and $l \in [L_X]$ **do**
10: **if** $X \notin \mathbf{S}_m(t)$ **then**
11: Update: $M_{X_m}^{(l)}(t) += V_{X_m}^{(l)}(t)(V_{X_m}^{(l)}(t))^\top$, $b_{X_m}^{(l)}(t) += V_{X_m}^{(l)}(t)\hat{X}_m^{(l)}(t)$, $\Delta M_{X_m}^{(l)}(t) += V_{X_m}^{(l)}(t)(V_{X_m}^{(l)}(t))^\top$, $\Delta b_{X_m}^{(l)}(t) += V_{X_m}^{(l)}(t)\hat{X}_m^{(l)}(t)$, $\Delta t_{X_m}^{(l)} += 1$.
12: **if** $\Delta t_{X_m}^{(l)} \log(\det(M_{X_m}^{(l)}(t))/\det(M_{X_m}^{(l)}(t) - \Delta M_{X_m}^{(l)}(t))) > D(\mathcal{K}_{X_m}^{(l)})$ **then**
13: Send a synchronization signal to the server
14: **if** Synchronization is started **then**
15: Each agent $m \in \mathcal{K}_{X_m}^{(l)}$ sends $\Delta M_{X_m}^{(l)}(t)$ and $\Delta b_{X_m}^{(l)}(t)$ to server.
16: Each agent $m \in \mathcal{K}_{X_m}^{(l)}$ receives $M_X^{(l)}(t) = \sum_{n\in\mathcal{K}_{X_m}^{(l)}} \Delta M_{X_n}^{(l)}(t)$ and $b_X^{(l)}(t) = \sum_{n\in\mathcal{K}_{X_m}^{(l)}} \Delta b_{X_n}^{(l)}(t)$ from server.
17: Local updates: $M_{X_m}^{(l)}(t) += M_X^{(l)}(t) - \Delta M_{X_m}^{(l)}(t)$, $b_{X_m}^{(l)}(t) += b_X^{(l)}(t) - \Delta b_{X_m}^{(l)}(t)$, $\Delta M_{X_m}^{(l)}(t) = 0$, $\Delta b_{X_m}^{(l)}(t) = 0$, $\Delta t_{X_m}^{(l)} = 0$.

This estimation is utilized to implement feasible collaboration. Then $\{\hat{X}_m^{(l)}(t)\}_{l\in[L_X]}$ are used to update the sufficient statistics (line 11 in Algorithm 2).

Arm Selection. Following the spirit of optimism in the face of uncertainty, our algorithm requires agents to select the best action in confidence ellipsoids (line 6 in Algorithm 2). Then we formally provide the rule of arm selection: $\mathbf{S}_m(t) = \arg\max_{|\mathbf{S}|=K,\{\theta'^{(l)}_{X_m}\in \mathrm{CB}^{(l)}_{X_m}(t)\}_{X,l}} \mathbb{E}[Y_m|do(\mathbf{S})]$, where the confidence ellipsoids $\mathrm{CB}_{X_m}^{(l)}(t)$ are defined in the following Lemma 3.1.

Lemma 3.1 (Confidence Ellipsoids)**.** *For any node $X \in \mathcal{V}$ and any subset $l \in [L_X]$ of its parents, suppose agent m is a member of cluster $\hat{C}_k \in \hat{\mathcal{C}}_X^{(l)}$, and is thus collaborating with agents $n \in \hat{C}_k$. With probability at least $1-\delta$, for $t > 0$, $\theta_{X_m}^{(l)}$ lies in the set:*

$$
\begin{aligned}
CB_{X_m}^{(l)}(t) = \Big\{ & \theta \in [0,1]^{|\boldsymbol{Pa}^{(l)}(X)|} : \|\theta - \hat{\theta}_{X_m}^{(l)}(t)\|_{M_{X_m}^{(l)}(t)} \\
& \le \big(\sigma_X + 2|\boldsymbol{Pa}(X)|\big)\sqrt{2\log\big(\frac{\det(M_{X_m}^{(l)}(t))^{1/2}}{\det(\lambda I)^{1/2}\delta}\big)} \\
& + \Big\| \sum_{n\in\hat{C}_k\setminus\{m\}} (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(\theta_{X_n}^{(l)} - \theta_{X_m}^{(l)}\big) \Big\|_{(M_{X_m}^{(l)}(t))^{-1}} \\
& + \sqrt{\lambda|\boldsymbol{Pa}^{(l)}(X)|}\Big\}.
\end{aligned} \tag{3}
$$

The detailed proof of Lemma 3.1 is provided in Appendix E.

Communication Protocol. Our devised algorithm combines the event-triggered communication protocol proposed in Wang et al. [2020] to balance the trade-off between communication cost and regret minimization within clusters. This protocol requires agents to store observations in two local buffers denoted by $\Delta M_{X_m}^{(l)}(t))$ and $\Delta b_{X_m}^{(l)}(t)$. Clients request server collaboration when the informativeness of the stored updates surpass a certain threshold. In particular, if $\Delta t_{X_m}^{(l)} \log(\det(M_{X_m}^{(l)}(t))/\det(M_{X_m}^{(l)}(t) - \Delta M_{X_m}^{(l)}(t))) > D(\mathcal{K}_{X_m}^{(l)})$, the agent sends a synchronization signal to the central server, where $D(\mathcal{K}_{X_m}^{(l)})$ denotes the threshold for the estimated cluster $\mathcal{K}_{X_m}^{(l)}$ including agent m with respect to the associated node $X \in \mathcal{X}$ and the subset $l \in [L_X]$ of its parents $\boldsymbol{Pa}(X)$.

We outline the overall procedure of the collaborative learning phase in Algorithm 2.

4 THEORETICAL RESULTS

In this section, we first prove that with our homogeneity test, the clustering algorithm correctly identifies the underlying clusters. Due to the space limit, the details of proof are deferred to Appendix F.

Theorem 4.1 (Clustering Correctness)**.** *Under the condition that we set the homogeneity test threshold* $\nu_X^{(l)} \geq F^{-1}(1-\frac{\delta}{M^2ND}; df_X^{(l)}; \frac{1}{\tilde{\sigma}_X^2})$*, with probability at least* $1-\delta$*, we have* $\hat{\mathcal{C}}_X^{(l)} = \mathcal{C}_X^{(l)}$*, for any node* $X \in \mathcal{V}$ *and any subset* $l \in [L_X]$ *of components.*

$F^{-1}(\cdot)$ *is the inverse of the CDF of the non-central* χ^2 *distribution,* $df_X^{(l)} = rank(\mathbf{V}_{X_m}^{(l)}) + rank(\mathbf{V}_{X_n}^{(l)}) - rank(\begin{bmatrix}\mathbf{V}_{X_m}^{(l)} \\ \mathbf{V}_{X_n}^{(l)}\end{bmatrix})$ *denotes the degree of freedom.*

We provide the detailed proof of Theorem 4.1 in Appendix D. Then we provide the cumulative regret $R(T)$ and communication cost $C(T)$ of our algorithm below.

Theorem 4.2 (Regret and Communication cost)**.** *With an exploration phase length of* $T_0 = \sum_{X\in\mathbf{X}}\sum_{l\in[L_X]}\frac{16M\tilde{\sigma}_X^2\psi_X^{(l)}}{\lambda_c\gamma^2}$*, with probability* $1-\delta$ *our protocol achieves a cumulative regret of*

$$R(T) = \mathcal{O}\Big(\sum_{X\in\mathbf{X}}\sum_{l\in[L_X]}\frac{M\tilde{\sigma}_X^2\psi_X^{(l)}}{\lambda_c\gamma^2} + \sum_{X\in\mathcal{V}}|\boldsymbol{Pa}(X)|^2 \sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]}\big(\sqrt{T|C_k||\boldsymbol{Pa}^{(l)}(X)|}\log(|C_k|T) + \sqrt{T|C_k|}\log^2(|C_k|T)\big)\Big), \quad (4)$$

where λ_c *denotes the minimum eigenvalue of* $\mathbb{E}[V_X^{'}(t)(V_X^{'}(t))^\top]$ *for non-intervened subset* $\boldsymbol{Pa}'(X)$ *of any* $X \in \mathcal{V}$*.* $\psi_X^{(l)} := F^{-1}(\frac{\delta}{M^2(K_X^{(l)}-1)ND}; df_X^{(l)}; \nu_X^{(l)})$*, with communication cost*

$$C(T) = \mathcal{O}\Big(M\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}|\boldsymbol{Pa}^{(l)}(X)|^2 + \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}|\boldsymbol{Pa}^{(l)}(X)|^3\sum_{k\in[K_X^{(l)}]}|C_k|^{1.5}\Big). \quad (5)$$

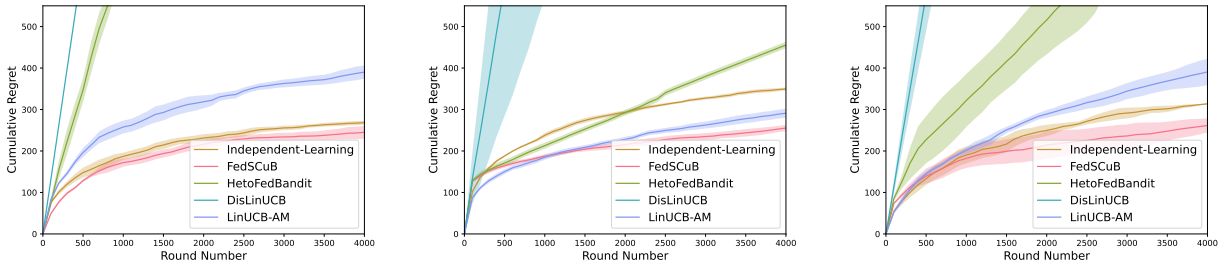
Remark 4.3. The regret upper bound consists of two parts. The first term, arising from the clustering stage, reflects the inherent difficulty of clustering in our problem setting and is independent of T. Such a term is also common in existing clustered bandit studies [Li et al., 2021, Blaser et al., 2024]. The second term represents the standard regret upper bound of FedCCB, incurred during the collaborative learning phase within the ground-truth clusters. Similarly, the communication cost of our protocol comprises two components: the first part corresponds to the cost of uploading sufficient statistics required for clustering, while the second part, associated with collaborative learning, aligns with the communication protocol outlined in Wang et al. [2020] within the ground-truth clusters.

Trade-off between communication cost and regret. Notice that when the cluster structure is defined at the level of individual components (i.e., $L_X = |\boldsymbol{Pa}(X)|$), the regret becomes $\tilde{\mathcal{O}}(\sum_{X\in\mathcal{V}}|\boldsymbol{Pa}(X)|^2\sum_{l\in[L_X]}\sum_{k\in K_X^{(l)}}\sqrt{T|C_k|})$, while the communication cost is $\mathcal{O}(M\sum_{X\in\mathcal{V}}|\boldsymbol{Pa}(X)| + \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in K_X^{(l)}}|C_k|^{1.5})$. Compared to performing clustering on the overall parameter vectors, component-wise clustering reduces communication cost but increases regret: in the case of homogeneous causal influences, the regret shifts from $\tilde{\mathcal{O}}(ND^2\sqrt{TMD})$ to $\tilde{\mathcal{O}}(ND^3\sqrt{TM})$, and the communication cost changes from $\mathcal{O}(M^{1.5}ND^3)$ to $\mathcal{O}(M^{1.5}ND)$. This reduction in communication cost is achieved because, under component-wise clustering, the uploaded covariance matrices are reduced to at most D scalars.

Then we compare the regret bound of FedSCuB with available baselines. To the best of our knowledge, this work is the first to study federated learning for CCB. We mainly extend four baseline algorithms in federated linear bandits and compute their regret orders in the CCB setting: a) independent optimistic learning [Feng and Chen, 2023]; b) DisLinUCB [Wang et al., 2020]; c) LinUCB-AM [Li and Wang, 2022b]; d) HetoFedBandit [Blaser et al., 2024].

Comparison to Prior Methods. For the FedCCB problem with heterogeneous causal influences, the regret bound of independent learning is $\tilde{\mathcal{O}}\Big(\sum_X|\boldsymbol{Pa}(X)|M\sqrt{T|\boldsymbol{Pa}(X)|}\Big)$. Besides, the regret bound of DisLinUCB [Wang et al., 2020] is $\tilde{\mathcal{O}}\Big(\sum_X|\boldsymbol{Pa}(X)|\big(\sqrt{TM|\boldsymbol{Pa}(X)|} + \epsilon' MT\sqrt{|\boldsymbol{Pa}(X)|}\big)\Big)$, where $\epsilon' \geq \max\{\|\theta_{X_m} - \theta_{X_n}\| : m \neq n \in [M]\}$ denotes the dissimilarity between local causal models. LinUCB-AM's regret order is $\tilde{\mathcal{O}}\Big(\sum_X|\boldsymbol{Pa}(X)|\big(\sqrt{MT|\boldsymbol{Pa}^{(g)}(X)|} + \sum_m\sqrt{T|\boldsymbol{Pa}^{(m)}(X)|}\big)\Big)$, where $\boldsymbol{Pa}^{(g)}(X)$ denotes the globally shared covariates and $\boldsymbol{Pa}^{(m)}(X)$ denotes the local personalized covariates. The regret bound of HetoFedBandit is $\tilde{\mathcal{O}}\Big(\sum_X|\boldsymbol{Pa}(X)|\sum_k\big(\sum_l\sqrt{T|C_k||\boldsymbol{Pa}^{(l)}(X)|} + \sum_j\epsilon'|C_k|T\sqrt{|\boldsymbol{Pa}^{(j)}(X)|}\big)\Big)$, where the second term is caused by the dissimilarity from certain components across agents within the same cluster.

The regret bound of independent learning scales with the order of $M\sqrt{T}$. In contrast, FedSCuB achieves a tighter regret order of $\sqrt{MT}$ when we consider homogeneous causal influences and perform clustering on parameter vectors. The inferior regret bound of independent learning stems from its failure to leverage collaboration among agents with similar distributions. It implies that enabling collaboration among similar agents is crucial to enhance the performance of FedCCB. Regarding DisLinUCB, this method simply aggregates statistics from all agents. This introduces an additional heterogeneity term of the order $\epsilon' MT$ in the regret bound. If the dissimilarity ϵ' exceeds the order of $\sqrt{T}$, the regret bound of DisLinUCB becomes worse than



(a) Parallel graph with six nodes. (b) Parallel graph with eight nodes. (c) Parallel graph with ten nodes.

Figure 2: Baseline methods for FedCCB using parallel graphs.

that of independent learning, indicating that collaboration among dissimilar agents will worsen the performance of FedCCB. In contrast, FedSCuB identifies similar agents through clustering and restricts collaboration to agents within the same cluster, effectively eliminating the additional heterogeneity term. LinUCB-AM only facilitates collaboration on specific components, introducing an additional term $\sum_{m\in[M]}\sqrt{|\boldsymbol{Pa}^{(m)}(X)|T} = \mathcal{O}(M\sqrt{T})$. The order of $M\sqrt{T}$ in LinUCB-AM's regret results in a worse performance compared to FedSCuB. Besides, LinUCB-AM depends on prior knowledge of which features are globally shared covariates, a requirement that significantly limits its practicality in reality. HetoFedBandit conducts clustering based on the overall parameter vector. However, the parameter vectors of agents may differ significantly in certain components. Consequently, collaboration across all components within a cluster introduces a heterogeneity term $\epsilon'|C_k|T$ for some components in the regret bound.

4.1 LOWER BOUND FOR ONE-LAYER PARALLEL GRAPH

In this section, we present a lower bound analysis for FedCCB to demonstrate the optimality of our proposed FedSCuB method. Here we focus on a special causal graph consisting of a one-layer parallel structure and consider the known cluster structure for any parent subset of the target node [Li and Zhang, 2018, Liu et al., 2022]. We defer the proof of lower bound to Appendix G.

Theorem 4.4 (Lower bound for FedCCB under the One-layer Parallel Graph)**.** *Let $\mathcal{I}^{L_Y}(\epsilon)$ denote the class of ϵ-FedCCB under one-layer parallel graph instances, we have the following,*

$$\inf_{\mathcal{A}}\sup_{\mathcal{I}\in\mathcal{I}^{L_Y}(\epsilon)}\mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T) = \Omega\left(\sum_{l\in[L_Y]}\sum_{k\in[K_Y^{(l)}]}|\boldsymbol{Pa}^{(l)}(Y)|\sqrt{|C_k|T}\right). \tag{6}$$

Remark 4.5. We note that our algorithm is near-optimal under the setting of FedCCB under one-layer parallel graph, since its regret upper bound $\tilde{\mathcal{O}}(|\boldsymbol{Pa}(Y)|^2\sum_l\sum_k\sqrt{|C_k|T})$ differs from the lower bound by only a logarithmic factor, which is common in general linear bandits. Although the upper bound of the proposed algorithm contains a factor of $|\boldsymbol{Pa}(Y)|^2$, this deterioration stems from the fact that the algorithm uses the AM method to estimate the partial response value, having little impact on the regret, as causal graphs are sparse in most applications [Cai et al., 2023].

5 EXPERIMENTS

In this section, we evaluate the empirical performance of our proposed FedSCuB algorithm, by comparing it with baselines on simulated datasets. All experiments are conducted on the A100 GPU. We first describe how we generate the synthetic dataset. Following Feng and Chen [2023], we adopt parallel graphs with 6, 8, and 10 nodes. We set the number of communication round as $T = 4000$, the number of agents as $M = 16$. In this paper, we only consider performing clustering and collaborative learning for the parameter vector of the target variable Y. Specifically, we assume $\boldsymbol{Pa}(Y)$ can be partitioned into two groups $\{\boldsymbol{Pa}^{(1)}(Y), \boldsymbol{Pa}^{(2)}(Y)\}$. The corresponding parameter vector is $\theta_Y = (\theta_Y^{(1)}, \theta_Y^{(2)})$. For each group, we follow the manner of data generation proposed in Blaser et al. [2024] to sample cluster centers via standard normal distribution that are away from each other. Then for each group, we randomly assign each agent to one of the clusters. In our evaluation, we employ baseline methods introduced in the previous section. To conserve space, the details of experimental settings are provided in Appendix H.

Results. In Figure 2, we compare FedSCuB with baselines on three parallel graphs with different numbers of nodes. We simulate 20 times of each experiment and draw the 95% confidence intervals of average regrets. We notice that DisLinUCB experiences linear regret under heterogeneous environment. Although independent learning achieves sublinear regret, its cumulative regret is higher than that of our

method, since it overlooks accelerating the exploration process via collaboration. In addition, LinUCB-AM underperforms compared to FedSCuB, as it restricts collaboration to only certain components. HetoFedBandit also experiences linear regret since it performs clustering on the overall vector, resulting in significant heterogeneity under our simulated environment.

More empirical results including the ablation study on the clustering level, and additional comparative experiments on the two-layer graph are presented in Appendix H.

6 RELATED WORK

Causal Bandits. The causal bandit problem is first defined in Lattimore et al. [2016] under the assumption of a fully known causal graph and known interventional distributions. Subsequent research has sought to relax these assumptions to address more scenarios. For example, Lu et al. [2021], Malek et al. [2023] remove the requirement of a known causal graph structure, while Yabe et al. [2018b], Maiti et al. [2022], Sawarni et al. [2023] assume that interventional distributions are unavailable. Additionally, Yan and Tajer [2024], Varici et al. [2023b] investigate the use of soft interventions on the causal graph, and Yan et al. [2024] extend the problem to non-stationary environments. To address the challenge of an exponentially large action space in causal bandits, Feng and Chen [2023], Xiong and Chen [2023], Feng et al. [2023] introduce a combinatorial causal bandit framework. Despite these advancements, the majority of existing causal bandit algorithms are developed for centralized settings, where a central server has full access to all observational data. In contrast, our work explores a federated combinatorial causal bandit problem, where decentralized agents collaborate to identify their respective optimal interventions without sharing raw local data.

Federated/Distributed Linear Contextual Bandits. Several studies have explored multi-agent bandit learning with decentralized data. Our work is more related to the line of federated/distributed linear contextual bandits, where multiple agents collaborate to solve a linear contextual bandit problem under the coordination of a central server [Wang et al., 2020, Li and Wang, 2022a, He et al., 2022, Huang et al., 2021, Blaser et al., 2024, Do et al., 2023a, Yang et al., 2024]. The key distinction between our problem setting and previous federated linear bandits lies in two aspects: (a) we incorporate a Bayesian network characterized by multiple linear models, whereas prior works typically consider only a single linear model; (b) our framework involves a combinatorial action space.

Online Clustering of Bandits. The online clustering of bandits is first presented by Gentile et al. [2014] and has been followed by several studies to further tackle with the challenge of user heterogeneity in the recommendation application of contextual bandits [Gentile et al., 2017, Li and Zhang, 2018, Yang et al., 2020]. Some recent works aim at utilizing clustered bandit algorithms to group similar agents for achieving feasible collaboration in solving federated contextual bandit problem [Blaser et al., 2024, Liu et al., 2022, Yang et al., 2024]. These works commonly assume that the cluster structure among agents is defined with respect to the overall parameter vector. Our study introduces a more practical and finer-grained perspective by defining the cluster structure at the level of component subset of parameters.

7 CONCLUSION AND LIMITATIONS

In this paper, we study the FedCCB problem and develop novel methods to tackle the challenge of heterogeneous causal influences in FedCCB. We propose a more fine-grained clustered bandits algorithm called FedSCuB to facilitate feasible collaboration among heterogeneous agents. Through theoretical analysis and empirical studies, we demonstrate that our method can achieve significant regret reduction compared to baselines. As the first attempt to integrate FL into causal bandits, we focus on the most common setting of linear SEM with hard interventions. An important direction for future research involves extending our approach to more general settings, including non-linear causal models with soft interventions and latent variables.

Acknowledgements

The corresponding author Fang Kong is supported by Guangdong Basic and Applied Basic Research Foundation (2025A1515011412) and National Natural Science Foundation of China (62506150).

References

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

Ethan Blaser, Chuanhao Li, and Hongning Wang. Federated linear contextual bandits with heterogeneous clients. In *International Conference on Artificial Intelligence and Statistics*, pages 631–639. PMLR, 2024.

Léon Bottou, Jonas Peters, Joaquin Quiñonero Candela, Denis Xavier Charles, Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Y. Simard, and Ed Snelson. Counterfactual reasoning and learning systems: the example of computational advertising. *J. Mach. Learn. Res.*, 14 (1):3207–3260, 2013.

Hengrui Cai, Yixin Wang, Michael Jordan, and Rui Song. On learning necessary and sufficient causal graphs. *Advances in Neural Information Processing Systems*, 36: 42148–42160, 2023.

R Stephen Cantrell, Peter M Burrows, and Quang H Vuong. Interpretation and use of generalized chow tests. *International Economic Review*, pages 725–741, 1991.

Gregory C Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605, 1960.

Anh Do, Thanh Nguyen-Tang, and Raman Arora. Multi-agent learning with heterogeneous linear contextual bandits. *Advances in Neural Information Processing Systems*, 36:78768–78790, 2023a.

Anh Do, Thanh Nguyen-Tang, and Raman Arora. Multi-agent learning with heterogeneous linear contextual bandits. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023b. URL https://openreview.net/forum?id=7f6vH3mmhr.

Shi Feng and Wei Chen. Combinatorial causal bandits. In *AAAI*, pages 7550–7558. AAAI Press, 2023.

Shi Feng, Nuoya Xiong, and Wei Chen. Combinatorial causal bandits without graph skeleton. *CoRR*, abs/2301.13392, 2023.

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

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

Jiafan He, Tianhao Wang, Yifei Min, and Quanquan Gu. A simple and provably efficient algorithm for asynchronous federated contextual linear bandits. *Advances in neural information processing systems*, 35: 4762–4775, 2022.

Ruiquan Huang, Weiqiang Wu, Jing Yang, and Cong Shen. Federated linear contextual bandits. *Advances in neural information processing systems*, 34:27057–27068, 2021.

Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2):1–210, 2021.

Sameer Kanase, Yan Zhao, Shenghe Xu, Mitchell Goodman, Manohar Mandalapu, Benjamyn Ward, Chan Jeon, Shreya Kamath, Ben Cohen, Yujia Liu, Hengjia Zhang, Yannick Kimmel, Saad Khan, Brent Payne, and Patricia Grao. An application of causal bandit to content optimization. 2022.

Gi-Soo Kim, Young Suh Hong, Tae Hoon Lee, Myunghee Cho Paik, and Hongsoo Kim. Bandit-supported care planning for older people with complex health and care needs. In *2023 IEEE 5th International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, pages 1–5. IEEE, 2023.

Finnian Lattimore, Tor Lattimore, and Mark D. Reid. Causal bandits: Learning good interventions via causal inference. In *Advances in neural information processing systems*, pages 1181–1189, 2016.

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

Sanghack Lee and Elias Bareinboim. Structural causal bandits: Where to intervene? In *Advances in neural information processing systems*, pages 2573–2583, 2018.

Chuanhao Li and Hongning Wang. Asynchronous upper confidence bound algorithms for federated linear bandits. In *AISTATS*, volume 151 of *Proceedings of Machine Learning Research*, pages 6529–6553. PMLR, 2022a.

Chuanhao Li and Hongning Wang. Asynchronous upper confidence bound algorithms for federated linear bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 6529–6553. PMLR, 2022b.

Chuanhao Li, Qingyun Wu, and Hongning Wang. Unifying clustered and non-stationary bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 1063–1071. PMLR, 2021.

Junfan Li, Zheshun Wu, Zenglin Xu, and Irwin King. On the necessity of collaboration for online model selection with decentralized data. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Shuai Li and Shengyu Zhang. Online clustering of contextual cascading bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Shuai Li, Fang Kong, Kejie Tang, Qizhi Li, and Wei Chen. Online influence maximization under linear threshold model. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1192–1204. Curran Associates, Inc., 2020a. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/

0d352b4d3a317e3eae221199fdb49651-Paper.pdf.

Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, 37(3): 50–60, 2020b.

Siqi Liu, Kay Choong See, Kee Yuan Ngiam, Leo Anthony Celi, Xingzhi Sun, and Mengling Feng. Reinforcement learning for clinical decision support in critical care: comprehensive review. *Journal of medical Internet research*, 22(7):e18477, 2020.

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

Yangyi Lu, Amirhossein Meisami, Ambuj Tewari, and William Yan. Regret analysis of bandit problems with causal background knowledge. In *Conference on Uncertainty in Artificial Intelligence*, pages 141–150. PMLR, 2020.

Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Causal bandits with unknown graph structure. *Advances in Neural Information Processing Systems*, 34:24817–24828, 2021.

Aurghya Maiti, Vineet Nair, and Gaurav Sinha. A causal bandit approach to learning good atomic interventions in presence of unobserved confounders. In *Uncertainty in Artificial Intelligence*, pages 1328–1338. PMLR, 2022.

Alan Malek, Virginia Aglietti, and Silvia Chiappa. Additive causal bandits with unknown graph. In *International Conference on Machine Learning*, pages 23574–23589. PMLR, 2023.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

Vineet Nair, Vishakha Patil, and Gaurav Sinha. Budgeted and non-budgeted causal bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 2017–2025. PMLR, 2021.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Ayush Sawarni, Rahul Madhavan, Gaurav Sinha, and Siddharth Barman. Learning good interventions in causal graphs via covering. In *Uncertainty in Artificial Intelligence*, pages 1827–1836. PMLR, 2023.

Burak Varici, Karthikeyan Shanmugam, Prasanna Sattigeri, and Ali Tajer. Causal bandits for linear structural equation models. *J. Mach. Learn. Res.*, 24:297:1–297:59, 2023a.

Burak Varici, Karthikeyan Shanmugam, Prasanna Sattigeri, and Ali Tajer. Causal bandits for linear structural equation models. *Journal of Machine Learning Research*, 24 (297):1–59, 2023b.

Yuanhao Wang, Jiachen Hu, Xiaoyu Chen, and Liwei Wang. Distributed bandit learning: Near-optimal regret with efficient communication. In *ICLR*. OpenReview.net, 2020.

Zheshun Wu, Zenglin Xu, Dun Zeng, Qifan Wang, and Jie Liu. Advocating for the silent: Enhancing federated generalization for nonparticipating clients. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15, 2024. doi: 10.1109/TNNLS.2024.3516692.

Nuoya Xiong and Wei Chen. Combinatorial pure exploration of causal bandits. In *ICLR*. OpenReview.net, 2023.

Akihiro Yabe, Daisuke Hatano, Hanna Sumita, Shinji Ito, Naonori Kakimura, Takuro Fukunaga, and Ken-ichi Kawarabayashi. Causal bandits with propagating inference. In *ICML*, volume 80 of *Proceedings of Machine Learning Research*, pages 5508–5516. PMLR, 2018a.

Akihiro Yabe, Daisuke Hatano, Hanna Sumita, Shinji Ito, Naonori Kakimura, Takuro Fukunaga, and Ken-ichi Kawarabayashi. Causal bandits with propagating inference. In *International Conference on Machine Learning*, pages 5512–5520. PMLR, 2018b.

Zirui Yan and Ali Tajer. Linear causal bandits: Unknown graph and soft interventions. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Zirui Yan, Arpan Mukherjee, Burak Varıcı, and Ali Tajer. Robust causal bandits for linear models. *IEEE Journal on Selected Areas in Information Theory*, 2024.

Hantao Yang, Xutong Liu, Zhiyong Wang, Hong Xie, John CS Lui, Defu Lian, and Enhong Chen. Federated contextual cascading bandits with asynchronous communication and heterogeneous users. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20596–20603, 2024.

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

APPENDIX

In the appendices, we provide more details and results omitted in the main paper. The appendices are structured as follows:

- Appendix A provides the technical Lemmas used in this paper.
- Appendix B provides the notations used in this paper.
- Appendix C provides the missing algorithm details in the main paper.
- Appendix D provides the proof of Theorem 4.1.
- Appendix E provides the proof of Lemma 3.1
- Appendix F provides the proof of Theorem 4.2.
- Appendix G presents a lower-bound analysis for FedCCB under the one-layer parallel graph.
- Appendix H elaborates on the experiment details and missing experiment results.

A TECHNICAL LEMMAS

In this section, we introduce the technical lemmas utilized in the proofs.

Lemma A.1 (Lemma 11 in Abbasi-Yadkori et al. [2011]). *Let $\{X_t\}_{t=1}^{\infty}$ be a sequence in $\mathbb{R}^d$, V is a $d \times d$ positive definite matrix and define $\bar{V}_t = V + \sum_{s=1}^{t} X_s X_s^{\top}$, where $V = \lambda I$. Additionally we have that $\lambda_{\min}(V) \geq \max(1, L^2)$ and $\|X_t\|_2 \leq L$ for all t, then*

$$\log\left(\frac{\det(\bar{V}_n)}{\det(V)}\right) \leq \sum_{t=1}^{T} \|X_t\|^2_{\bar{V}_{t-1}^{-1}} \leq 2\log\left(\frac{\det(\bar{V}_n)}{\det(V)}\right). \tag{7}$$

Lemma A.2 (Theorem 1 of Abbasi-Yadkori et al. [2011]). *Let $\{\mathcal{F}_t\}_{t=0}^{\infty}$ be a filtration. Let $\{\eta_t\}_{t=1}^{\infty}$ be a real-valued stochastic process such that η_t is $\mathcal{F}_t$-measurable, and η_t follows conditionally zero mean R-sub-Gaussian for some $R \geq 0$. Let $\{X_t\}_{t=1}^{\infty}$ be an $\mathbb{R}^d$-valued stochastic process such that X_t is $\mathcal{F}_{t-1}$-measurable. Assume that V is a $d \times d$ positive definite matrix. For any $t > 0$, define*

$$V_t = V + \sum_{\tau=1}^{t} X_\tau X_\tau^{\top} \quad \mathcal{S}_t = \sum_{\tau=1}^{t} \eta_\tau X_\tau.$$

Then for any $\delta > 0$, with probability at least $1 - \delta$,

$$||\mathcal{S}_t||_{V_t^{-1}} \leq R\sqrt{2\log\frac{\det(V_t)^{1/2}}{\det(V)^{1/2}\delta}}, \quad \forall t \geq 0.$$

Lemma A.3 (Determinant-Trace Inequality). *Suppose $X_1, X_2, ..., X_t \in \mathbb{R}^d$ and for any $1 \leq \tau \leq t$, $\|X_\tau\|_2 \leq L$, Let $\overline{V_t} = \lambda I + \sum_{\tau=1}^{t} X_\tau X_\tau^{\top}$ for some $\lambda > 0$. Then,*

$$\det(\overline{V_t}) \leq (\lambda + tL^2/d)^d.$$

Lemma A.4 (Lemma 8 of Blaser et al. [2024]). *When the underlying bandit parameters θ_i^* and θ_j^* of two observation sequence $\mathcal{H}_{i,t-1}$ and $\mathcal{H}_{j,t-1}$ are not the same, the probability that cluster identification module clusters them together can be upper bounded by:*

$$P\Big(S(\mathcal{H}_{i,t-1}, \mathcal{H}_{j,t-1}) \leq \upsilon^c | \|\theta_i^* - \theta_j^*\| > \epsilon\Big) \leq F(\upsilon^c; d, \psi^c)$$

under the condition that both $\lambda_{\min}(\sum_{(\mathbf{x}_k, y_k) \in \mathcal{H}_{i,t-1}} \mathbf{x}_k \mathbf{x}_k^{\top})$ and $\lambda_{\min}(\sum_{(\mathbf{x}_k, y_k) \in \mathcal{H}_{j,t-1}} \mathbf{x}_k \mathbf{x}_k^{\top})$ are at least $\frac{2\psi^c \sigma^2}{\gamma^2}$.

Lemma A.5 (Lemma 9 of Blaser et al. [2024]). *Denote the number of samples that have been used to update $\{V_{i,t}, b_{i,t}\}$ as τ_i. At each round t, for each agent $m \in [M]$ and any node $X \in \mathcal{V}$, if any non-intervened subset $\boldsymbol{Pa}'(X)$ of its parents $\boldsymbol{Pa}(X)$, i.e., $\boldsymbol{Pa}'(X) \nsubseteq \mathbf{S}_m(t)$, the corresponding covariance matrix $\mathbb{E}[V'_{X_m}(t)(V'_{X_m}(t))^{\top}]$ is full-rank with minimum eigenvalue $\lambda_c > 0$, then with probability at least $1 - \delta$, we have*

$$\lambda_{\min}(V_{i,t}) \geq \lambda + \frac{\lambda_c \tau_i}{8}, \tag{8}$$

$\forall \tau_i \in \{\tau_i, \tau_i + 1, ..., T\}$, where $\tau_{\min} = \lceil \frac{64}{3\lambda_c} \log(\frac{2Td}{\delta}) \rceil$.

Remark A.6. According to our considered CCB model, the non-intervened parents are sub-Gaussian with non-zero variance. As a result, this lemma holds naturally in our problem since any combination of parents constitutes a sub-Gaussian random vector.

B NOTATIONS

In this section, we introduce the main notations used in this paper. We list the main notations below

Notations	Meanings
$\mathcal{G}$	a causal graph $\mathcal{G} = \{\mathbf{X} \cup \{Y\}, \mathcal{E}\}$
$\mathcal{V}$	$\mathcal{V} = \mathbf{X} \cup \{Y\}$ the set of nodes in $\mathcal{G}$
Y	the reward node in $\mathcal{G}$
N	number of nodes in $\mathcal{G}$
D	the maximum in-degree of nodes in $\mathcal{G}$
K	the maximal number of nodes that can be chosen to intervene in a round
P	used to define the propagating rule by $P(X\|\boldsymbol{Pa}(X))$
$\boldsymbol{Pa}(X)$	set of parents of node X
$\boldsymbol{Pa}^{(l)}(X)$	the l-th subset of parents $\boldsymbol{Pa}(X)$
L_X	the number of subsets of $\boldsymbol{Pa}(X)$
ϵ	denotes the upper bound of dissimilarity between parameters within a cluster
γ	denotes the lower bound of dissimilarity between parameters in two different clusters
$R(T)$	the cumulative regret of federated combinatorial causal bandits system
$\mathbf{S}_m(t)$	the set of nodes that agent m perform $do(\mathbf{S}_m(t) = \mathbf{s}_m(t))$ in round t
T	total number of rounds in the federated CCB problem
df	the degree of freedom in a χ^2 distribution
$(\mathbf{X}_m(t), Y_m(t))$	the observed propagating result of nodes in agent m's causal model in round t
$X_m(t)$	denotes the observation of node X for agent m in round t
$V_{X_m}(t)$	denotes the observation of parents $\boldsymbol{Pa}(X)$ of node X for agent m in round t
$X_m^{(l)}(t)$	the propagating result for node X contributed by subset l of its parents for agent m in round t
$V_{X_m}^{(l)}(t)$	denotes the observation of subset l of parents $\boldsymbol{Pa}(X)$ for agent m in round t
$M_{X_m}^{(l)}(t)$	the covariance matrix for subset l of $\boldsymbol{Pa}(X)$ in agent m's causal model
$\mathbf{V}_{X_m}^{(l)}(t)$	the design matrix, where each row of $\mathbf{V}_{X_m}^{(l)}$ is the feature vector corresponding to subset $\boldsymbol{Pa}^{(l)}(X)$.
$\mathcal{H}_{X_m}^{(l)}(t)$	$\mathcal{H}_{X_m}^{(l)}(t) = \{V_{X_m}^{(l)}(\tau), X_m^{(l)}(\tau)\}_{\tau \in [t]}$ denotes the set of collected covariate-response pairs
λ_c	denotes the minimum eigenvalue of matrix $\mathbb{E}[V_{X_m}^{(l)}(t)(V_{X_m}^{(l)}(t))^\top]$
δ	a constant in the online algorithms
θ_{X_m}	the real parameters of node $X \in \mathcal{V}$ in agent m's causal model
$\theta_{X_m}^{(l)}$	the real parameters of subset l of components within θ_{X_m}
$\hat{\theta}_{X_m}$	estimated parameters
$\tilde{\theta}_{X_m}$	parameter with the upper confidence bound or from the confidence ellipsoid
$\mathcal{C}_X^{(l)}$	denotes the set of ground-truth clusters with respect to the component subset $\theta_{X_m}^{(l)}$
$K_X^{(l)}$	$K_X^{(l)} = \|\mathcal{C}_X^{(l)}\|$ denotes the size of $\mathcal{C}_X^{(l)}$
$\lambda_{\min}(M)$	minimum eigenvalue of matrix M
$\{\rho_{X_m}^{(l)}\}$	constants in the input of our online algorithms
$\sigma(\mathbf{S}, \theta)$	the expected reward under parameter θ and intervention $do(\mathbf{S})$

C ALGORITHM DETAILS

In this section, we provide the details of our proposed algorithm which are missing in the main paper.

INDEPENDENT EXPLORATION PHASE

In this phase, agents independently execute pure exploration of length $T_0 = \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]} T_X^{(l)}$, where $T_X^{(l)}$ denotes the rounds needed for guaranteeing the accuracy of the homogeneity test for the node $X \in \mathcal{V}$ and the subset $l \in [L_X]$ of its parents. Different with Blaser et al. [2024] requiring agents to select actions by uniformly sampling, we provide an intervention-based exploration method to obtain partial observations needed for clustering agents at the level of component subset.

During this exploration stage, each agent $m \in [M]$ should construct datasets $\mathcal{H}_{X_m}^{(l)}(t) = \{V_{X_m}^{(l)}(\tau), X_m^{(l)}(\tau)\}_{\tau\in[t]}$. For accessing such unobservable $X_m^{(l)}(t)$, agent m first observes the non-intervened realization of $X_m(t)$. Then the agent selects the intervention set $\mathbf{S}_m(t) = \boldsymbol{Pa}^{(l)}(X_m)$, and intervenes by setting all nodes in $\mathbf{S}_m(t)$ to zero. The difference between two realizations of X_m, which is defined as $X_m^{(l)}(t) := \theta_{X_m}^{(l)}\cdot\boldsymbol{Pa}^{(l)}(X_m)+\sum_{l'\neq l}\theta_{X_m}^{(l')}\cdot(\boldsymbol{Pa}^{(l')}(X_m)-\tilde{\boldsymbol{Pa}}^{(l')}(X_m))+\eta_{X_m}-\tilde{\eta}_{X_m}$, has mean $\mathbb{E}\big[\theta_{X_m}^{(l)}\cdot\boldsymbol{Pa}^{(l)}(X_m)\big]$ and noise variance $\mathcal{O}(\sigma_{X_m}^2 + L_X^2)$. Below we define $\tilde{\sigma}_X^2 := \mathcal{O}(\sigma_{X_m}^2 + L_X^2)$. Based on this operation, the causal influences from the subsets $\boldsymbol{Pa}^{(k)}(X), k \in [L_X]\setminus\{l\}$ to X can be regarded to be removed indirectly, so we can regard the practical observation $X_m(t)$ as the partial observation $X_m^{(l)}(t)$ in this scenario.

After receiving the $(V_{X_m}^{(l)}(t), X_m^{(l)}(t))$ pair, each agent updates its sufficient statistics as: $M_{X_m}^{(l)}(t) += V_{X_m}^{(l)}(t)(V_{X_m}^{(l)}(t))^\top$, $b_{X_m}^{(l)}(t) += V_{X_m}^{(l)}(t)X_m^{(l)}(t)$. Upon completion of $T_X^{(l)}$ rounds of pure exploration for each node $X \in \mathcal{V}$ and each subset $l \in [L_X]$ of its parents, each agent $m \in [M]$ then uploads its sufficient statistics $\{V_{X_m}^{(l)}(T_X^{(l)}), b_{X_m}^{(l)}(T_X^{(l)})\}_{X\in\mathcal{V},l\in[L_X]}$ to the central server.

CLUSTER ESTIMATION

In this paper, we follow the idea proposed in Blaser et al. [2024] of clustering similar parameters (ϵ defined in Assumption 2.1). The key distinction is that we further conduct the more fine-grained component-subset-wise clustering to broaden the scope of beneficial collaboration. We implement this by testing whether $\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \le \epsilon$ via the homogeneity test provided in Blaser et al. [2024], Li et al. [2021]. The key idea of this homogeneity test is to construct the test statistic $s(\mathcal{H}_{X_m}^{(l)}(t), \mathcal{H}_{X_n}^{(l)}(t))$ following the non-central χ^2-distribution. This homogeneity test determines whether the parameters of linear regression models associated with two datasets are similar under the equal variance. Since $\{\theta_{X_m}^{(l)}\}_{m\in[M]}$ are not observable, this test has to be performed on their estimates, for which maximum likelihood estimator (MLE) is a prevalent choice.

Based on the sufficient statistics uploaded by agents, the server can obtain the MLE of $\theta_{X_m}^{(l)}$ defined by $v_{X_m}^{(l)} = (M_{X_m}^{(l)}(t))^{-}b_{X_m}^{(l)}(t)$, for any node and any subset of its parents. We apply the following test statistics that has been widely used in this type of problems [Chow, 1960, Cantrell et al., 1991].

$$s(\mathcal{H}_{X_m}^{(l)}(t), \mathcal{H}_{X_n}^{(l)}(t)) := \frac{\|\mathbf{V}_{X_m}^{(l)}(v_{X_m}^{(l)} - v_{X_m,X_n}^{(l)})\|^2 + \|\mathbf{V}_{X_n}^{(l)}(v_{X_n}^{(l)} - v_{X_m,X_n}^{(l)})\|^2}{\tilde{\sigma}_X^2}, \tag{9}$$

where $v_{X_m,X_n}^{(l)}$ denotes the estimator using data from $\mathcal{H}_{X_m}^{(l)}(t)$ and $\mathcal{H}_{X_n}^{(l)}(t)$. $\mathbf{V}_{X_m}^{(l)}$ denotes the design matrix, where each row of $\mathbf{V}_{X_m}^{(l)}$ is the covariate vector corresponding to subset $\boldsymbol{Pa}^{(l)}(X)$.

When $s(\mathcal{H}_{X_m}^{(l)}(t), \mathcal{H}_{X_n}^{(l)}(t))$ is above a selected threshold $\nu_X^{(l)}$, it shows that the aggregated estimator deviates heavily from the individual estimators on two datasets. Hence, for any node and any subset of its parents, to determine the corresponding sets of agents that can collaborate, the central server performs this pairwise homogeneity test among each pair of agents. If the homogeneity test satisfies $s(\mathcal{H}_{X_m}^{(l)}(t), \mathcal{H}_{X_n}^{(l)}(t)) \le \nu_X^{(l)}$, then we add an undirected edge between them in a graph $\mathcal{G}_X^{(l)}$ demonstrating they benefit from collaborating with each other. To ensure each agent within a cluster is sufficiently similar to one another, we follow a similar idea in Blaser et al. [2024] to require each of our estimated clusters be a maximal clique of $\mathcal{G}_X^{(l)}$ (line 6 in Algorithm 1). Therefore, for every $X \in \mathcal{V}$ and $l \in [L_X]$, we denote the set of estimated clusters as $\hat{\mathcal{C}}_X^{(l)} = \{\hat{C}_k\}_{k=1}^{\hat{K}_X^{(l)}}$, and allocate each agent $m \in [M]$ to one estimated cluster $\mathcal{K}_{X_m}^{(l)} = \hat{C}_k$, where $m \in \hat{C}_k$.

D PROOF OF THEOREM 4.1

In this section, we provide the full proof of Theorem 4.1. For any node $X \in \mathcal{V}$ and any subset $l \in [L_X]$ of its parents, Theorem 4.1 states that utilizing the homogeneity test with threshold $\nu_X^{(l)} \geq F^{-1}(1 - \frac{\delta}{NDM^2}, df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2})$, after the exploration phase of length $T_0 = \sum_{X \in \mathcal{V}} \sum_{l \in [L_X]} \frac{16 \psi_X^{(l)} \tilde{\sigma}_X^2}{\lambda_c \gamma^2}$, the estimated clusters $\hat{\mathcal{C}}_X^{(l)}$ match the ground-truth clusters $\mathcal{C}_X^{(l)}$, for any node $X \in \mathcal{V}$ and any subset $l \in [L_X]$ of its parents.

For a given node and one specific subset of its parents, we first assume that the probability of $\hat{\mathcal{C}}_X^{(l)} \neq \mathcal{C}_X^{(l)}$ can be bounded by $\tilde{\delta}$. Based on it, we focus on the probability of the event $\{\exists X \in \mathcal{V}, l \in [L_X] : \hat{\mathcal{C}}_X^{(l)} \neq \mathcal{C}_X^{(l)}\}$:

$$\begin{aligned}
\mathbb{P}\{\exists X \in \mathcal{V}, l \in [L_X] : \hat{\mathcal{C}}_X^{(l)} \neq \mathcal{C}_X^{(l)}\} &= \mathbb{P}\{ \bigcup_{X \in \mathcal{V}, l \in [L_X]} \hat{\mathcal{C}}_X^{(l)} \neq \mathcal{C}_X^{(l)}\} \\
&\leq \sum_{X \in \mathcal{V}} \sum_{l \in [L_X]} \mathbb{P}\{\hat{\mathcal{C}}_X^{(l)} \neq \mathcal{C}_X^{(l)}\} \\
&\leq \sum_{X \in \mathcal{V}} \sum_{l \in [L_X]} \tilde{\delta} \\
&\leq ND\tilde{\delta}.
\end{aligned} \tag{10}$$

We denote the derived upper bound as δ and then we can know that $\tilde{\delta} = \frac{\delta}{ND}$. Hence, we then focus on deriving the probability $\tilde{\delta}$.

The homogeneity test statistic follows a non-central χ^2 distribution, i.e., $s(\mathcal{H}_{X_m}^{(l)}(T_{X_m}^{(l)}), \mathcal{H}_{X_n}^{(l)}(T_{X_n}^{(l)})) \sim \chi^2(df_X^{(l)}, \psi_X^{(l)})$, where the degree of freedom is

$$df_X^{(l)} = rank(\mathbf{V}_{X_m}^{(l)}) + rank(\mathbf{V}_{X_n}^{(l)}) - rank([\mathbf{V}_{X_m}^{(l)}, \mathbf{V}_{X_n}^{(l)}]), \tag{11}$$

and the non-centrality parameter is

$$\begin{aligned}
&\psi_X^{(l)} \\
&= \frac{\begin{bmatrix} \mathbf{V}_{X_m}^{(l)} \theta_{X_m}^{(l)} \\ \mathbf{V}_{X_n}^{(l)} \theta_{X_n}^{(l)} \end{bmatrix}^\top \left[I_{T_{X_m}^{(l)} + T_{X_n}^{(l)}} - \begin{bmatrix} \mathbf{V}_{X_m}^{(l)} \\ \mathbf{V}_{X_n}^{(l)} \end{bmatrix} \left((\mathbf{V}_{X_m}^{(l)})^\top \mathbf{V}_{X_m}^{(l)} + (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \right)^- \begin{bmatrix} (\mathbf{V}_{X_m}^{(l)})^\top & (\mathbf{V}_{X_n}^{(l)})^\top \end{bmatrix} \right] \begin{bmatrix} \mathbf{V}_{X_m}^{(l)} \theta_{X_m}^{(l)} \\ \mathbf{V}_{X_n}^{(l)} \theta_{X_n}^{(l)} \end{bmatrix}}{\tilde{\sigma}_X^2}.
\end{aligned} \tag{12}$$

Motivated by the proof of Blaser et al. [2024], we then also prove two corollaries. The first step is to prove that with high probability $\hat{\mathcal{C}}_{X_m}^{(l)} \subset \mathcal{C}_{X_m}^{(l)}$. Next we move to prove that $\mathcal{C}_{X_m}^{(l)} \subset \hat{\mathcal{C}}_{X_m}^{(l)}$. Consequently, the conjunction of these two events holding simultaneously shows that $\mathcal{C}_{X_m}^{(l)} = \hat{\mathcal{C}}_{X_m}^{(l)}$.

LOWER BOUNDING $\mathcal{C}_{X_m}^{(l)} \subset \hat{\mathcal{C}}_{X_m}^{(l)}$

In order to prove that $\mathcal{C}_{X_m}^{(l)} \subset \hat{\mathcal{C}}_{X_m}^{(l)}$, we need to show that the set of edges in the ground-truth agent graph $\mathcal{G}_X^{(l)}$ is a subset of the edges in the estimated agent graph $\hat{\mathcal{G}}_X^{(l)}$. Hence, we should upper-bound the type-I error probability of the homogeneity test, which corresponds the probability that our algorithm fails to cluster two agents together when the underlying bandit parameters $\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \leq \epsilon$

Lemma D.1. *The type-I error probability can be upper bounded by:*

$$\mathbb{P}\Big(s\big(\mathcal{H}_{X_m}^{(l)}(T_{X_m}^{(l)}), \mathcal{H}_{X_n}^{(l)}(T_{X_n}^{(l)})\big) > v \Big| \|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \leq \epsilon \Big) \leq 1 - F(v; df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2}),$$

where $F(v; df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2})$ denotes the cumulative density function of distribution $\chi^2(df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2})$ evaluated at v, and $\frac{1}{\tilde{\sigma}_X^2}$ denotes the non-centrality parameter.

Proof. Denote $\zeta = \theta^{(l)}_{X_n} - \theta^{(l)}_{X_m}$ and thus $\theta^{(l)}_{X_n} = \zeta + \theta^{(l)}_{X_m}$. Then we know that the non-centrality parameter $\psi^{(l)}_X$ defined in Eq. (12) becomes

$$\begin{aligned}
\tilde{\sigma}^2_X \psi^{(l)}_X = &\begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \theta^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \theta^{(l)}_{X_m} \end{bmatrix}^\top \left[I_{T^{(l)}_{X_m}+T^{(l)}_{X_n}} - \begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \end{bmatrix} \left((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} + (\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \right)^- \begin{bmatrix} (\mathbf{V}^{(l)}_{X_m})^\top & (\mathbf{V}^{(l)}_{X_n})^\top \end{bmatrix} \right] \begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \theta^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \theta^{(l)}_{X_m} \end{bmatrix} \\
&+ \begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \theta^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \theta^{(l)}_{X_m} \end{bmatrix}^\top \left[I_{T^{(l)}_{X_m}+T^{(l)}_{X_n}} - \begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \end{bmatrix} \left((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} + (\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \right)^- \begin{bmatrix} (\mathbf{V}^{(l)}_{X_m})^\top & (\mathbf{V}^{(l)}_{X_n})^\top \end{bmatrix} \right] \begin{bmatrix} 0 \\ \mathbf{V}^{(l)}_{X_n} \zeta \end{bmatrix} \\
&+ \begin{bmatrix} 0 \\ \mathbf{V}^{(l)}_{X_n} \zeta \end{bmatrix}^\top \left[I_{T^{(l)}_{X_m}+T^{(l)}_{X_n}} - \begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \end{bmatrix} \left((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} + (\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \right)^- \begin{bmatrix} (\mathbf{V}^{(l)}_{X_m})^\top & (\mathbf{V}^{(l)}_{X_n})^\top \end{bmatrix} \right] \begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \theta^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \theta^{(l)}_{X_m} \end{bmatrix} \\
&+ \begin{bmatrix} 0 \\ \mathbf{V}^{(l)}_{X_n} \zeta \end{bmatrix}^\top \left[I_{T^{(l)}_{X_m}+T^{(l)}_{X_n}} - \begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \end{bmatrix} \left((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} + (\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \right)^- \begin{bmatrix} (\mathbf{V}^{(l)}_{X_m})^\top & (\mathbf{V}^{(l)}_{X_n})^\top \end{bmatrix} \right] \begin{bmatrix} 0 \\ \mathbf{V}^{(l)}_{X_n} \zeta \end{bmatrix}.
\end{aligned} \tag{13}$$

Since $\begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \theta^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \theta^{(l)}_{X_m} \end{bmatrix}$ is in the column space of $\begin{bmatrix} \mathbf{V}^{(l)}_{X_m} \\ \mathbf{V}^{(l)}_{X_n} \end{bmatrix}$, the first term in the second equality of Eq. (13) is zero. The second and third terms can be shown equal to zero as well using the property that matrix product is distributive with respect to matrix addition. Hence, we have

$$\begin{aligned}
\psi^{(l)}_X &= \frac{1}{\tilde{\sigma}^2_X} (\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n})^\top (\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \left((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} + (\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \right) (\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} (\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}) \\
&\leq \frac{1}{\tilde{\sigma}^2_X} \|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\|^2 \lambda_{\max} \Big((\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \left((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} + (\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \right) (\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} \Big) \\
&\leq \frac{\epsilon^2}{\tilde{\sigma}^2_X} \lambda_{\max} \Big((\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \left((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} + (\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \right) (\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} \Big).
\end{aligned} \tag{14}$$

Moreover, we can utilize the fact that $\mathbf{Y}(\mathbf{Y}+\mathbf{Y})^{-1}\mathbf{X} = (\mathbf{Y}^{-1}+\mathbf{X}^{-1})^{-1}$, where $\mathbf{Y}$ and $\mathbf{X}$ are both invertible matrices to further have

$$\begin{aligned}
\psi^{(l)}_X &\leq \frac{\epsilon^2}{\tilde{\sigma}^2_X} \lambda_{\max} \Big(\Big(\big((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} \big)^{-1} + \big((\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \big)^{-1} \Big)^{-1} \Big) \\
&\leq \frac{\epsilon^2}{\tilde{\sigma}^2_X} \frac{1}{\lambda_{\min} \Big(\big((\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \big)^{-1} \Big) + \lambda_{\min} \Big(\big((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} \big)^{-1} \Big)} \\
&= \frac{\epsilon^2}{\tilde{\sigma}^2_X} \frac{1}{\frac{1}{\lambda_{\max} \Big((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} \Big)} + \frac{1}{\lambda_{\max} \Big((\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \Big)}} \\
&\leq \frac{\epsilon^2}{\tilde{\sigma}^2_X} \max \Big\{ \lambda_{\max} \Big((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} \Big), \lambda_{\max} \Big((\mathbf{V}^{(l)}_{X_n})^\top \mathbf{V}^{(l)}_{X_n} \Big) \Big\}.
\end{aligned} \tag{15}$$

Denote the number of samples in $\mathbf{V}^{(l)}_{X_m}$ as $T^{(l)}_{X_m}$. Given that $\|V^{(l)}_{X_m}(t)\| \leq |\boldsymbol{Pa}^{(l)}(X)|$, we know that $\lambda_{\max} \Big((\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m} \Big) \leq \|(\mathbf{V}^{(l)}_{X_m})^\top \mathbf{V}^{(l)}_{X_m}\|_F \leq T^{(l)}_{X_m} |\boldsymbol{Pa}^{(l)}(X)|$. Thus we further have

$$\begin{aligned}
\psi^{(l)}_X &\leq \frac{\epsilon^2 |\boldsymbol{Pa}^{(l)}(X)|}{\tilde{\sigma}^2_X} \max\{T^{(l)}_{X_m}, T^{(l)}_{X_n}\} \\
&\leq \frac{\epsilon^2 |\boldsymbol{Pa}^{(l)}(X)| T}{\tilde{\sigma}^2_X} \\
&\leq \frac{|\boldsymbol{Pa}^{(l)}(X)| T}{DT\tilde{\sigma}^2_X} \\
&\leq \frac{1}{\tilde{\sigma}^2_X},
\end{aligned} \tag{16}$$

where the third inequality holds since $\epsilon \leq \frac{1}{M\sqrt{DT}}$.

Therefore, when $\|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| \leq \epsilon$, the test statistics $s \sim \chi^2(df^{(l)}_X, \frac{1}{\tilde{\sigma}^2_X})$. The type-I error probability can be thus upper bounded by $1 - F(v; df^{(l)}_X, \frac{1}{\tilde{\sigma}^2_X})$. □

Since our algorithm leverages the pairwise homogeneity test to evaluate whether each pair of agents is within ϵ of one another, to achieve a type-I error probability of $\tilde{\delta}/M^2$ between two individual agents, we can solve for the needed threshold $\nu^{(l)}_X$. Given that we have

$$\begin{aligned} \frac{\tilde{\delta}}{M^2} &\leq 1 - F(v; df^{(l)}_X, \frac{1}{\tilde{\sigma}^2_X}) \\ \Rightarrow \frac{\delta}{NDM^2} &\leq 1 - F(v; df^{(l)}_X, \frac{1}{\tilde{\sigma}^2_X}) \end{aligned} \tag{17}$$

Based on it, we know the threshold $\nu^{(l)}_X$ should satisfies

$$\nu^{(l)}_X \geq F^{-1}(1 - \frac{\delta}{NDM^2}, df^{(l)}_X, \frac{1}{\tilde{\sigma}^2_X}). \tag{18}$$

Taking the union bound over all M^2 pairwise tests proves the that the set of edges in the ground-truth agent graph is a subset of the edges estimated agent graph. Hence we have proved that $\mathcal{C}^{(l)}_{X_m} \subset \hat{\mathcal{C}}^{(l)}_{X_m}$

LOWER BOUNDING $\hat{\mathcal{C}}^{(l)}_{X_m} \subset \mathcal{C}^{(l)}_{X_m}$

In this section, we prove that with high probability $\hat{\mathcal{C}}^{(l)}_{X_m} \subset \mathcal{C}^{(l)}_{X_m}$ for a specific subset of one node. Similarly, we are supposed to indicate that the set of edges in the estimated agent graph is a subset of the ground-truth edges in $\mathcal{G}^{(l)}_X$. To achieve this, we utilize the type-II error probability upper-bound to ensure that with high probability agents with different underlying parameters are not clustered together. Using this type-II error probability, we follow similar steps in Lemma 13 of Blaser et al. [2024] to prove the following lemma.

Lemma D.2. *If the cluster identification module clusters observation history $(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m})$ and $(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m})$ together, the probability that they actually have the same underlying bandit parameters satisfies*

$$\mathbb{P}\Big(\|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| \leq \epsilon \Big| s\big(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m}), \mathcal{H}^{(l)}_{X_n}(T^{(l)}_{X_n})\big) \leq \nu^{(l)}_X\Big) \geq F(\nu^{(l)}_X; df^{(l)}_X, \frac{1}{\tilde{\sigma}^2_X}), \tag{19}$$

when $\lambda_{\min}\Big(\sum_{(V^{(l)}_{X_m}(\tau), X^{(l)}_m(\tau)) \in \mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m})} V^{(l)}_{X_m}(\tau)(V^{(l)}_{X_m}(\tau))^\top\Big)$ and $\lambda_{\min}\Big(\sum_{(V^{(l)}_{X_n}(\tau), X^{(l)}_n(\tau)) \in \mathcal{H}^{(l)}_{X_n}(T^{(l)}_{X_m})} V^{(l)}_{X_n}(\tau)(V^{(l)}_{X_n}(\tau))^\top\Big)$ are at least $\frac{2\psi^{(l)}_X \tilde{\sigma}^2_X}{\gamma^2}$, where $\psi^{(l)}_X := F^{-1}\big(\frac{\delta}{M^2(K^{(l)}_X - 1)ND}; df^{(l)}_X; \nu^{(l)}_X\big)$.

Proof. Denote the events $\big\{\|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| > \epsilon\big\} \cap \big\{s\big(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m}), \mathcal{H}^{(l)}_{X_n}(T^{(l)}_{X_n})\big) > \nu^{(l)}_X\big\}$ as True Positive (TP), $\big\{\|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| \leq \epsilon\big\} \cap \big\{s\big(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m}), \mathcal{H}^{(l)}_{X_n}(T^{(l)}_{X_n})\big) \leq \nu^{(l)}_X\big\}$ as True Negative (TN), $\big\{\|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| \leq \epsilon\big\} \cap \big\{s\big(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m}), \mathcal{H}^{(l)}_{X_n}(T^{(l)}_{X_n})\big) > \nu^{(l)}_X\big\}$ as False Positive (FP), and $\big\{\|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| > \epsilon\big\} \cap \big\{s\big(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m}), \mathcal{H}^{(l)}_{X_n}(T^{(l)}_{X_n})\big) \leq \nu^{(l)}_X\big\}$ as False Negative (FN) of cluster identification, respectively. We can rewrite the probabilities in Lemma A.4, D.1 and D.2 as:

$$\begin{aligned} P\big(s\big(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m}), \mathcal{H}^{(l)}_{X_n}(T^{(l)}_{X_n})\big) > \nu^{(l)}_X \Big| \|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| \leq \epsilon\big) &= \frac{P(FP)}{P(TN + FP)} \leq 1 - F(\nu^{(l)}_X; df^{(l)}_X, \frac{1}{\tilde{\sigma}^2_X}) \\ P\big(s\big(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m}), \mathcal{H}^{(l)}_{X_n}(T^{(l)}_{X_n})\big) \leq \nu^{(l)}_X \Big| \|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| > \epsilon\big) &= \frac{P(FN)}{P(FN + TP)} \leq F(\nu^{(l)}_X; df^{(l)}_X, \psi^{(l)}_X) \\ P\big(\|\theta^{(l)}_{X_m} - \theta^{(l)}_{X_n}\| \leq \epsilon \Big| s\big(\mathcal{H}^{(l)}_{X_m}(T^{(l)}_{X_m}), \mathcal{H}^{(l)}_{X_n}(T^{(l)}_{X_n})\big) \leq \nu^{(l)}_X\big) &= \frac{P(TN)}{P(TN + FN)} = \frac{1}{1 + \frac{P(FN)}{P(TN)}} \end{aligned}$$

We can upper bound $\frac{P(FN)}{P(TN)}$ by:

$$\frac{P(FN)}{P(TN)} \leq \frac{P(TP+FN)}{P(TN+FP)} \cdot \frac{F(\nu_X^{(l)}; df_X^{(l)}, \psi_X^{(l)})}{F(\nu_X^{(l)}; df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2})}$$

where $\frac{P(TP+FN)}{P(TN+FP)}$ denotes the ratio between the number of positive instances and negative instances in the population. We can upper bound this ratio for any pair (m, n) uniformly sampled from $[M]$, since we need to run the test on allM^2 pairs. First we find that $\frac{P(TP+FN)}{P(TN+FP)} = \frac{P(\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| > \epsilon)}{P(\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \leq \epsilon)}$. We upper-bound this ratio by giving a lower bound on the probability of two randomly sampled clients belonging to the same cluster as $P(\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \leq \epsilon)$ with

$$\begin{aligned} P(\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \leq \epsilon) &= \sum_{k=1}^{K_X^{(l)}} \frac{|C_k|}{M} \times \frac{|C_k|-1}{M-1} \\ &> \sum_{k=1}^{K_X^{(l)}} (\frac{|C_k|-1}{M-1})^2 \\ &> \sum_{k=1}^{K_X^{(l)}} \frac{1}{(K_X^{(l)})^2} \\ &= \frac{1}{K_X^{(l)}}. \end{aligned} \tag{20}$$

The second inequality is true because the probability that two uniformly sampled clients belonging to the same cluster is minimized when the clusters are all of equal sizes. Therefore we have

$$\frac{P(\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| > \epsilon)}{P(\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \leq \epsilon)} \leq \frac{1 - \frac{1}{K_X^{(l)}}}{\frac{1}{K_X^{(l)}}} = K_X^{(l)} - 1 \tag{21}$$

With this upper bound of $\frac{P(FN)}{P(TN)}$, we can now have

$$P\Big(\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \leq \epsilon \Big| s\big(\mathcal{H}_{X_m}^{(l)}(T_{X_m}^{(l)}), \mathcal{H}_{X_n}^{(l)}(T_{X_n}^{(l)})\big) \leq \nu_X^{(l)}\Big) \geq 1/\Big(1 + (K_X^{(l)} - 1) \cdot \frac{F(\nu_X^{(l)}; df_X^{(l)}, \psi_X^{(l)})}{F(\nu_X^{(l)}; df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2})}\Big) \tag{22}$$

Then by setting $\psi_X^{(l)} := F^{-1}\Big(\frac{1-F\left(\nu_X^{(l)}; df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2}\right)}{K_X^{(l)}-1}; df_X^{(l)}; \nu_X^{(l)}\Big)$, we have:

$$P\Big(\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \leq \epsilon \Big| s\big(\mathcal{H}_{X_m}^{(l)}(T_{X_m}^{(l)}), \mathcal{H}_{X_n}^{(l)}(T_{X_n}^{(l)})\big) \leq \nu_X^{(l)}\Big) \geq F(\nu_X^{(l)}; df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2}). \tag{23}$$

□

Based on Lemma D.2, we can further prove that $\hat{\mathcal{C}}_{X_m}^{(l)} \subset \mathcal{C}_{X_m}^{(l)}$. Under the condition that we set the threshold $\nu_X^{(l)} \geq F^{-1}(1 - \frac{\tilde{\delta}}{M^2}, df_X^{(l)}), \frac{1}{\tilde{\sigma}_X^2}$, with an exploration phase length of $T_X^{(l)} = \min\{\frac{64}{3\lambda_c} \log \frac{2T|\boldsymbol{Pa}^{(l)}(X)|}{\tilde{\delta}}, \frac{16\psi_X^{(l)}\tilde{\sigma}_X^2}{\lambda_c \gamma^2}\}$, the application of Lemma A.5 gives with probability $1 - \tilde{\delta}$ that

$$\lambda_{\min}\Big((\mathbf{V}_{X_m}^{(l)})^\top \mathbf{V}_{X_m}^{(l)}\Big) \geq \frac{\lambda_c T_X^{(l)}}{8} = \frac{2\psi_X^{(l)}\tilde{\sigma}_X^2}{\gamma^2}. \tag{24}$$

Consequently, we can apply Lemma D.2 which states

$$P\big(\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \leq \epsilon \big| s\big(\mathcal{H}_{X_m}^{(l)}(T_{X_m}^{(l)}), \mathcal{H}_{X_n}^{(l)}(T_{X_n}^{(l)})\big) \leq \nu_X^{(l)}\big) \geq F(\nu_X^{(l)}; df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2}). \tag{25}$$

Using the similar steps, we can see our choice of test statistics $\nu_X^{(l)} \geq F^{-1}(1-\frac{\tilde{\delta}}{M^2}, df_X^{(l)}, \frac{1}{\tilde{\sigma}_X^2})$ leads to this event happening with probability $1-\frac{\tilde{\delta}}{M^2}$. Because our algorithm conducts this pairwise homogeneity test across all pairs of clients, a union bound over all M^2 pairwise tests proves the corollary.

The combination of the results given in section D and section D proves that based on our choice of $\nu_X^{(l)}$ and $T_0 = \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]} T_X^{(l)}$, $\mathcal{C}_X^{(l)} = \hat{\mathcal{C}}_X^{(l)}$ with probability $1-\delta$.

E PROOF OF LEMMA 3.1

Lemma E.1. *For any node $X \in \mathcal{V}$ and any subset $l \in [L_X]$ of its parents, if each agent $m \in [M]$ is provided the unbiased estimation $\{\hat{\theta}_{X_m}^{(l)}\}$, the term $e_{X_m}^{(l)}(\tau) := \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(\tau)\cdot(\theta_{X_m}^{(j)} - \hat{\theta}_{X_m}^{(j)}(\tau))$ is zero-mean $2|\boldsymbol{Pa}(X)|$-sub-Gaussian, condition on $\mathcal{F}_{t-1}, \forall t$.*

Proof. Recall that the unbiased estimators $\{\hat{\theta}_{X_m}^{(l)}\}$ are MLE estimators obtained before in the independent exploration phase. When conditioning on $\mathcal{F}_{t-1} = \{(\mathbf{X}_m(\tau), Y_m(\tau))\}_{\tau=1}^t$, $\sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(t)\cdot(\theta_{X_m}^{(j)} - \hat{\theta}_{X_m}^{(j)})$ is a random variable. We define $\mathbf{1}_X^{(l)} \in 1^{|\boldsymbol{Pa}^{(1)}(X)|}$ as a all-ones vector whose dimension is $|\boldsymbol{Pa}^{(1)}(X)|$, then we have

$$
\begin{aligned}
\Big|\sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(t)\cdot(\theta_{X_m}^{(j)} - \hat{\theta}_{X_m}^{(j)})\Big| &\leq \Big|\sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(t)\cdot\theta_{X_m}^{(j)}\Big| + \Big|\sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(t)\cdot\hat{\theta}_{X_m}^{(j)}\Big| \\
&\overset{(a)}{=} 2\sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(t)\cdot\theta_{X_m}^{(j)} \\
&\leq 2\sum_{j\in[L_X]\setminus\{l\}} \mathbf{1}_X^{(j)}\cdot\mathbf{1}_X^{(j)} \\
&= 2\sum_{j\in[L_X]\setminus\{l\}} |\boldsymbol{Pa}^{(j)}(X)| \\
&\leq 2\sum_{j\in[L_X]} |\boldsymbol{Pa}^{(j)}(X)| \\
&= 2|\boldsymbol{Pa}(X)|,
\end{aligned}
\tag{26}
$$

where (a) holds from the fact that $\hat{\theta}_X^{(l)}, \theta_X^{(l)} \in [0,1]^{|\boldsymbol{Pa}^{(l)}(X)|}$ and $X \in [0,1]$, for any node X and any subset $l \in [L_X]$.

As mentioned above, $\{\hat{\theta}_{X_m}^{(l)}\}$ are unbiased estimators obtained before the independent exploration phase using historical data. Hence we have $\mathbb{E}[\sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(t)\cdot(\theta_{X_m}^{(j)} - \hat{\theta}_{X_m}^{(j)})] = \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(t)\cdot\mathbb{E}[\theta_{X_m}^{(j)} - \hat{\theta}_{X_m}^{(j)}] = 0$. Based on the above analysis, we know that $\sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(t)\cdot(\theta_{X_m}^{(j)} - \hat{\theta}_{X_m}^{(j)})$ is zero-mean $2|\boldsymbol{Pa}(X)|$-sub-Gaussian.

□

Now we are equipped to prove Lemma 3.1.

Proof. In the following analysis, we solely focus on one specific subset $l \in [L_X]$ of $\boldsymbol{Pa}(X)$ for node $X \in \mathcal{V}$. Besides, we only consider the case where agent is collaborating with members of its ground-truth cluster, since we have already proved with high probability $\mathcal{C}_X^{(l)} = \hat{\mathcal{C}}_X^{(l)}$. We can demonstrate that the regret caused by $\mathcal{C}_X^{(l)} \neq \hat{\mathcal{C}}_X^{(l)}$ is upper bounded by a constant. In this proof, we assume without loss of generality, that agent m belongs to ground-truth cluster $C_k \in \mathcal{C}_X^{(l)}$. Let $t_{X_m}^{(l)}$ be the last round that synchronization occurs for subset l and node X, where $t_{X_m}^{(l)} < t$. Then each agent's statistics can be divided

into two parts, its own data and the pooled data from other agents in its cluster since round $t_{X_m}^{(l)}$. In particular, we have

$$
\begin{aligned}
M_{X_m}^{(l)}(t) &= \lambda I + \sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)(V_{X_m}^{(l)}(\tau))^\top + \sum_{n\in C_k\setminus\{m\}} \sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)(V_{X_n}^{(l)}(\tau))^\top, \\
b_{X_m}^{(l)}(t) &= \sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)\hat{X}_m^{(l)}(\tau) + \sum_{n\in C_k\setminus\{m\}} \sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)\hat{X}_n^{(l)}(\tau),
\end{aligned}
\tag{27}
$$

where $\hat{X}_m^{(l)}(\tau) = X_m(\tau) - \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(\tau)\cdot\hat{\theta}_{X_m}^{(j)}(\tau)$. □

Since $\hat{\theta}_{X_m}^{(l)}(t) = (M_{X_m}^{(l)}(t))^{-1}b_{X_m}^{(l)}(t)$ and $\hat{X}_m^{(l)}(t) = X_m^{(l)}(t) - \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(t)\cdot\hat{\theta}_{X_m}^{(j)}(t)$, we then have

$$
\begin{aligned}
\hat{\theta}_{X_m}^{(l)}(t) &= (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)\hat{X}_m^{(l)}(\tau) + \sum_{n\in C_k\setminus\{m\}} \sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)\hat{X}_n^{(l)}(\tau)\Big] \\
&= (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)\big(X_m(\tau) - \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(\tau)\cdot\hat{\theta}_{X_m}^{(j)}(\tau)\big) \\
&\quad + \sum_{n\in C_k\setminus\{m\}} \sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)\big(X_n(\tau) - \sum_{j\in[L_X]\setminus\{l\}} V_{X_n}^{(j)}(\tau)\cdot\hat{\theta}_{X_n}^{(j)}(\tau)\big)\Big] \\
&= (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)\big(\theta_{X_m}\cdot V_{X_m}(\tau) + \eta_{X_m}(\tau) - \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(\tau)\cdot\hat{\theta}_{X_m}^{(j)}(\tau)\big) \\
&\quad + \sum_{n\in C_k\setminus\{m\}} \sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)\big(\theta_{X_n}\cdot V_{X_n}(\tau) + \epsilon_{X_n}(\tau) - \sum_{j\in[L_X]\setminus\{l\}} V_{X_n}^{(j)}(\tau)\cdot\hat{\theta}_{X_n}^{(j)}(\tau)\big)\Big].
\end{aligned}
\tag{28}
$$

Based on the considered partition of the variable subset, we have

$$
\begin{aligned}
\hat{\theta}_{X_m}^{(l)}(t) &= (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)\big(\sum_{l\in[L_X]} \theta_{X_m}^{(l)}\cdot V_{X_m}^{(l)}(\tau) + \eta_{X_m}(\tau) - \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(\tau)\cdot\hat{\theta}_{X_m}^{(j)}(\tau)\big) \\
&\quad + \sum_{n\in C_k\setminus\{m\}} \sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)\big(\sum_{l\in[L_X]} \theta_{X_n}^{(l)}\cdot V_{X_n}^{(l)}(\tau) + \epsilon_{X_n}(\tau) - \sum_{j\in[L_X]\setminus\{l\}} V_{X_n}^{(j)}(\tau)\cdot\hat{\theta}_{X_n}^{(j)}(\tau)\big)\Big] \\
&= (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)\big(\theta_{X_m}^{(l)}\cdot V_{X_m}^{(l)}(\tau) + \eta_{X_m}(\tau) + \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(\tau)\cdot(\theta_{X_m}^{(j)} - \hat{\theta}_{X_m}^{(j)}(\tau))\big) \\
&\quad + \sum_{n\in C_k\setminus\{m\}} \sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)\big(\theta_{X_n}^{(l)}\cdot V_{X_n}^{(l)}(\tau) + \epsilon_{X_n}(\tau) + \sum_{j\in[L_X]\setminus\{l\}} V_{X_n}^{(j)}(\tau)\cdot(\theta_{X_n}^{(j)} - \hat{\theta}_{X_n}^{(j)}(\tau))\big)\Big]. \\
&= (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)\big(\theta_{X_m}^{(l)}\cdot V_{X_m}^{(l)}(\tau) + \eta_{X_m}(\tau) + \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(\tau)\cdot(\theta_{X_m}^{(j)} - \hat{\theta}_{X_m}^{(j)}(\tau))\big) \\
&\quad + \sum_{n\in C_k\setminus\{m\}} \sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)\big((\theta_{X_n}^{(l)} + \theta_{X_m}^{(l)} - \theta_{X_m}^{(l)})\cdot V_{X_n}^{(l)}(\tau) + \epsilon_{X_n}(\tau) + \sum_{j\in[L_X]\setminus\{l\}} V_{X_n}^{(j)}(\tau)\cdot(\theta_{X_n}^{(j)} - \hat{\theta}_{X_n}^{(j)}(\tau))\big)\Big].
\end{aligned}
\tag{29}
$$

Reorganize it and we have

$$\begin{aligned}\hat{\theta}_{X_m}^{(l)}(t) = (M_{X_m}^{(l)}(t))^{-1}\Big[\Big(\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)(V_{X_m}^{(l)}(\tau))^\top + \sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)(V_{X_n}^{(l)}(\tau))^\top\Big)\theta_{X_m}^{(l)}\Big] \\ + (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)(V_{X_n}^{(l)}(\tau))^\top\big(\theta_{X_n}^{(l)} - \theta_{X_m}^{(l)}\big)\Big] \\ + (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)\eta_{X_m}(\tau) + \sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)\epsilon_{X_n}(\tau)\Big] \\ + (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)e_{X_m}^{(l)}(\tau) + \sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)e_{X_n}^{(l)}(\tau)\Big],\end{aligned} \tag{30}$$

where $e_{X_m}^{(l)}(\tau) := \sum_{j\in[L_X]\setminus\{l\}} V_{X_m}^{(j)}(\tau)\cdot(\theta_{X_m}^{(j)} - \hat{\theta}_{X_m}^{(j)}(\tau))$.
For the first term, we have

$$\begin{aligned}&(M_{X_m}^{(l)}(t))^{-1}\Big[\Big(\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)(V_{X_m}^{(l)}(\tau))^\top + \sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)(V_{X_n}^{(l)}(\tau))^\top\Big)\theta_{X_m}^{(l)}\Big] \\ &= (M_{X_m}^{(l)}(t))^{-1}\Big[\Big(\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)(V_{X_m}^{(l)}(\tau))^\top + \sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)(V_{X_n}^{(l)}(\tau))^\top + \lambda I - \lambda I\Big)\theta_{X_m}^{(l)}\Big] \\ &= (M_{X_m}^{(l)}(t))^{-1}\Big[\Big(\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)(V_{X_m}^{(l)}(\tau))^\top + \sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)(V_{X_n}^{(l)}(\tau))^\top + \lambda I\Big)\theta_{X_m}^{(l)} - \lambda\theta_{X_m}^{(l)}\Big] \\ &= \theta_{X_m}^{(l)} - \lambda(M_{X_m}^{(l)}(t))^{-1}\theta_{X_m}^{(l)}\end{aligned} \tag{31}$$

As a result, we have

$$\begin{aligned}\hat{\theta}_{X_m}^{(l)}(t) - \theta_{X_m}^{(l)} = -\lambda(M_{X_m}^{(l)}(t))^{-1}\theta_{X_m}^{(l)} + \underbrace{(M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)(V_{X_n}^{(l)}(\tau))^\top\big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)\Big]}_{\text{Dissimilarity term}} \\ \underbrace{+ (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)\eta_{X_m}(\tau) + \sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)\epsilon_{X_n}(\tau)\Big]}_{\text{Noise term}} \\ \underbrace{+ (M_{X_m}^{(l)}(t))^{-1}\Big[\sum_{\tau=1}^{t} V_{X_m}^{(l)}(\tau)e_{X_m}^{(l)}(\tau) + \sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t_{X_m}^{(l)}} V_{X_n}^{(l)}(\tau)e_{X_n}^{(l)}(\tau)\Big]}_{\text{Estimation error of partial observation}}.\end{aligned} \tag{32}$$

Applying triangle inequality, we have the following for $\|\hat{\theta}_{X_m}^{(l)}(t) - \theta_{X_m}^{(l)}\|_{\left(M_{X_m}^{(l)}(t)\right)^{-1}}$.

$$\begin{aligned}\|\hat{\theta}^{(l)}_{X_m}(t)-\theta^{(l)}_{X_m}\|_{M^{(l)}_{X_m}(t)} \le \lambda\|\theta^{(l)}_{X_m}\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}} &+ \Big\|\sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t^{(l)}_{X_m}} V^{(l)}_{X_n}(\tau)(V^{(l)}_{X_n}(\tau))^\top\big(\theta^{(l)}_{X_m}-\theta^{(l)}_{X_n}\big)\Big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}\\ &+\Big\|\sum_{\tau=1}^{t} V^{(l)}_{X_m}(\tau)\eta_{X_m}(\tau)+\sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t^{(l)}_{X_m}} V^{(l)}_{X_n}(\tau)\epsilon_{X_n}(\tau)\Big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}\\ &+\Big\|\sum_{\tau=1}^{t} V^{(l)}_{X_m}(\tau)e^{(l)}_{X_m}(\tau)+\sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t^{(l)}_{X_m}} V^{(l)}_{X_n}(\tau)e^{(l)}_{X_n}(\tau)\Big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}\end{aligned} \tag{33}$$

The first term $\lambda\|\theta^{(l)}_{X_m}\|_{(M^{(l)}_{X_m}(t))^{-1}} \le \frac{\lambda}{\lambda_{\min}\left(M^{(l)}_{X_m}(t)\right)^{1/2}}\|\theta^{(l)}_{X_m}\|_2 \le \sqrt{\lambda}\|\theta^{(l)}_{X_m}\|_2 \le \sqrt{\lambda|\boldsymbol{Pa}^{(l)}(X)|}$.

For the third term, we know that the exogenous variable $\eta_{X_m}(t)$ is zero-mean σ_X-sub-Gaussian, then we have

$$\Big\|\sum_{\tau=1}^{t} V^{(l)}_{X_m}(\tau)\eta_{X_m}(\tau)+\sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t^{(l)}_{X_m}} V^{(l)}_{X_n}(\tau)\epsilon_{X_n}(\tau)\Big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}} \le \sigma_X\sqrt{2\log\Big(\frac{\det(M^{(l)}_{X_m}(t))^{1/2}\det(\lambda I)^{-1/2}}{\delta}\Big)}. \tag{34}$$

Similarly, based on Lemma E.1, we know that $e^{(l)}_{X_m}(\tau)$ is zero-mean $2|\boldsymbol{Pa}(X)|$-sub-Gaussian, then the fourth term becomes

$$\begin{aligned}&\Big\|\sum_{\tau=1}^{t} V^{(l)}_{X_m}(\tau)e^{(l)}_{X_m}(\tau)+\sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t^{(l)}_{X_m}} V^{(l)}_{X_n}(\tau)e^{(l)}_{X_n}(\tau)\Big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}\\ &\le 2|\boldsymbol{Pa}(X)|\sqrt{2\log\Big(\frac{\det(M^{(l)}_{X_m}(t))^{1/2}\det(\lambda I)^{-1/2}}{\delta}\Big)}.\end{aligned} \tag{35}$$

F PROOF OF THEOREM 4.2

We use $\sigma(\mathbf{S},\theta)$ to represent the reward function $\mathbb{E}[Y|do(\mathbf{S})]$ under parameter θ, to make the parameters of the reward explicit.

Lemma F.1 (Heterogeneity Term Bound)**.** *Under the condition that we set the homogeneity test threshold* $\nu^{(l)}_X \ge F^{-1}(1-\frac{\delta}{M^2ND};df^{(l)}_X;\psi_X)$*, and with an exploration phase length of* $T_0=\min\{\frac{64}{3\lambda_c}\log(\frac{2TD}{\delta}),\frac{16\psi(\sigma_X^2+4D^2)}{\lambda_c\gamma^2}\}$*, we have with probability* $1-\delta$*:*

$$\Big\|\sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t^{(l)}_{X_m}} V^{(l)}_{X_n}(\tau)(V^{(l)}_{X_n}(\tau))^\top\big(\theta^{(l)}_{X_m}-\theta^{(l)}_{X_n}\big)\Big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}} \le 1 \tag{36}$$

Proof. First we notice that $\sum_{\tau=1}^{t^{(l)}_{X_m}} V^{(l)}_{X_n}(\tau)(V^{(l)}_{X_n}(\tau))^\top = (\mathbf{V}^{(l)}_{X_n})^\top\mathbf{V}^{(l)}_{X_n}$, where $\mathbf{V}^{(l)}_{X_n}\in[0,1]^{t^{(l)}_{X_m}\times|\boldsymbol{Pa}^{(l)}(X)|}$ denotes the design matrix. Each row of $\mathbf{V}^{(l)}_{X_n}$ is a feature vector sampled at the corresponding time step. Hence, we can rewrite the term $\Big\|\sum_{n\in C_k\setminus\{m\}}\sum_{\tau=1}^{t^{(l)}_{X_m}} V^{(l)}_{X_n}(\tau)(V^{(l)}_{X_n}(\tau))^\top\big(\theta^{(l)}_{X_m}-\theta^{(l)}_{X_n}\big)\Big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}$ as $\Big\|\sum_{n\in C_k\setminus\{m\}}(\mathbf{V}^{(l)}_{X_n})^\top\mathbf{V}^{(l)}_{X_n}\big(\theta^{(l)}_{X_m}-$

$\theta_{X_n}^{(l)}\big)\Big\|_{\left(M_{X_m}^{(l)}(t)\right)^{-1}}$ below. Then we have

$$
\begin{aligned}
&\Big\| \sum_{n\in C_k\setminus\{m\}} (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)\Big\|_{\left(M_{X_m}^{(l)}(t)\right)^{-1}} \\
&\le \sum_{n\in C_k\setminus\{m\}} \Big\|(\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)\Big\|_{\left(M_{X_m}^{(l)}(t)\right)^{-1}} \\
&= \sum_{n\in C_k\setminus\{m\}} \sqrt{\big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)^\top (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(M_{X_m}^{(l)}(t)\big)^{-1} (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)} \\
&\overset{(a)}{\le} \sum_{n\in C_k\setminus\{m\}} \sqrt{\big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)^\top (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big((\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)}\big)^{-1} (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)} \\
&= \sum_{n\in C_k\setminus\{m\}} \sqrt{\big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)^\top (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)},
\end{aligned}
\tag{37}
$$

where (a) holds since the sum over all agents $M_{X_m}^{(l)}(t)$ necessarily includes $(\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)}$, hence $M_{X_m}^{(l)}(t) \succeq (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)}$. Additionally, $(\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)}$ is positive semi-definite so that $(M_{X_m}^{(l)}(t))^{-1} \preceq \big((\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)}\big)^{-1}$.

We further have

$$
\begin{aligned}
\Big\| \sum_{n\in C_k\setminus\{m\}} (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)\Big\|_{\left(M_{X_m}^{(l)}(t)\right)^{-1}} &\overset{(a)}{\le} \sum_{n\in C_k\setminus\{m\}} \sqrt{\big\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big\| \big\|(\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)\big\|} \\
&\overset{(b)}{\le} \sum_{n\in C_k\setminus\{m\}} \big\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big\| \sqrt{\big\|(\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)}\big\|} \\
&\overset{(c)}{\le} \sum_{n\in C_k\setminus\{m\}} \big\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big\| \sqrt{\big\|\mathbf{V}_{X_n}^{(l)}\big\|_F^2} \\
&\overset{(d)}{\le} \sum_{n\in C_k\setminus\{m\}} \big\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big\| \sqrt{|\boldsymbol{Pa}^{(l)}(X)| t_{X_m}^{(l)}} \\
&\le \sum_{n\in C_k\setminus\{m\}} \big\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big\| \sqrt{DT},
\end{aligned}
\tag{38}
$$

where (a) holds because of the CauchySchwarz inequality. (b) is given by the definition of the spectral norm of the matrix. (c) follows from the fact that $\|(\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)}\|_2 \le \|(\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)}\|_F$ and $\|(\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)}\|_F \le \|\mathbf{V}_{X_n}^{(l)}\|_F^2$. (d) holds since we know that $X \in [0,1], \forall X \in \mathcal{V}$.

Theorem 4.1 indicates theta by setting $\nu_X^{(l)}$ and T_0 properly, with probability $1-\delta$, we have $\hat{\mathcal{C}}_{X_m}^{(l)} = \mathcal{C}_{X_m}^{(l)}$. Thus, since m, n belong to the same ground-truth cluster C_k, we have by Assumption 2.1, $\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\| \le \epsilon = \frac{1}{M\sqrt{DT}}$. Consequently, we can further upper bound the heterogeneity term by

$$
\begin{aligned}
\Big\| \sum_{n\in C_k\setminus\{m\}} (\mathbf{V}_{X_n}^{(l)})^\top \mathbf{V}_{X_n}^{(l)} \big(\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big)\Big\|_{\left(M_{X_m}^{(l)}(t)\right)^{-1}} &\le \sum_{n\in C_k\setminus\{m\}} \big\|\theta_{X_m}^{(l)} - \theta_{X_n}^{(l)}\big\| \sqrt{DT} \\
&\le \sum_{n\in C_k\setminus\{m\}} \frac{\sqrt{DT}}{M\sqrt{DT}} \\
&\le 1.
\end{aligned}
\tag{39}
$$

□

Lemma F.2. *With probability* $1 - MND\delta$*, for* $\theta \in [0,1]^{|\boldsymbol{Pa}^{(l)}(X)|}$*,* $\rho^{(l)}_{X_m}(t) = \|\theta - \hat{\theta}^{(l)}_{X_m}(t)\|_{M^{(l)}_{X_m}(t)}$ *is bounded by*

$$\begin{aligned}\rho^{(l)}_{X_m}(t) &\leq (\sigma_X + 2|\boldsymbol{Pa}(X)|)\sqrt{2\log\Big(\frac{\det(M^{(l)}_{X_m}(t))^{1/2}\det(\lambda I)^{-1/2}}{\delta}\Big)} + \sqrt{\lambda D} + \mathcal{O}(1)\\ &= \mathcal{O}\Big((\sigma_X + 2|\boldsymbol{Pa}(X)|)\sqrt{|\boldsymbol{Pa}^{(l)}(X)|\log\frac{T}{\delta}}\Big)\end{aligned} \tag{40}$$

Lemma F.3 (Group observation modulated Bounded Smoothness of LM [Li et al., 2020a, Feng and Chen, 2023])**.** *For any two weight vectors* $\theta^1, \theta^2 \in \Theta$ *for a linear model* $\mathcal{G}$*, the difference of their expected reward for any intervened set* $\mathbf{S}$ *can be bounded as*

$$|\sigma(\mathbf{S}, \theta^1) - \sigma(\mathbf{S}, \theta^2)| \leq \mathbb{E}_\epsilon\Big[\sum_{X\in\mathbf{X}_{\mathbf{S},Y}} |V_X^\top(\theta^1_X - \theta^2_X)|\Big], \tag{41}$$

where $\mathbf{X}_{\mathbf{S},Y}$ *is the set of paths from* $\mathbf{S}$ *to* Y *excluding* $\mathbf{S}$*, and* V_X *is the propagation result of the parents of* X *under parameter* θ^2*. The expectation is taken over the randomness of the noise* ϵ*.*

Remark F.4. The authors in [Li et al., 2020a, Feng and Chen, 2023] only provide the proof of this lemma in binary linear model (BLM) setting, but we point out that this lemma also holds under the continuous variables setting. Here we provide the formal proof of this lemma for the linear model with continuous variables setting considered in this paper.

Proof. In the following, we uses $\boldsymbol{\gamma}$ to denote a random vector with independent entries uniformly distributed on $[0,1]$, and denotes $\boldsymbol{\eta}$ to be a sub-Gaussian random vector. For BLM models, we know that $\mathbb{E}_{\boldsymbol{\gamma}}[X] = \mathbb{E}_{\boldsymbol{\gamma}}[\mathbb{I}\{\theta_X \cdot V_X \geq \gamma\}] = \mathbb{P}[\theta_X \cdot V_X \geq \gamma]$. For linear models with continuous variables, we have $\mathbb{E}_{\boldsymbol{\eta}}[X] = \mathbb{E}_{\boldsymbol{\eta}}[\theta_X \cdot V_X + \eta_X] = \theta_X \cdot \mathbb{E}_{\boldsymbol{\eta}}[V_X]$.

Now we prove that, under the same causal graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, for any node $X \in \mathcal{V}$, $\mathbb{E}_{\boldsymbol{\eta}}[X] = \mathbb{E}_{\boldsymbol{\gamma}}[X]$, i.e., we prove that the expected propagation results of node X under linear models and BLM models are the same. We know that for any intervened node $X \in \mathbf{S}$, $\mathbb{E}_{\boldsymbol{\eta}}[X] = \mathbb{E}_{\boldsymbol{\gamma}}[X] = 1$. Besides, for any non-intervened node $X \notin \mathbf{S}$, we first assume that $\mathbb{E}_{\boldsymbol{\eta}}[V_X] = \mathbb{E}_{\boldsymbol{\gamma}}[V_X]$. Then, for the node X itself, we have $\mathbb{E}_{\boldsymbol{\gamma}}[X] = \mathbb{P}[\theta_X \cdot V_X \geq \gamma] = \theta_X \cdot \mathbb{E}_{\boldsymbol{\gamma}}[V_X]$, and $\mathbb{E}_{\boldsymbol{\eta}}[X]$.

To sum up, we prove that for any node $X \in \mathcal{V}$, $\mathbb{E}_{\boldsymbol{\eta}}[X] = \mathbb{E}_{\boldsymbol{\gamma}}[X]$.

Given the below analysis, we conclude that the GOM lemma for BLM models also holds for linear models with continuous variables considered in this paper. □

Lemma F.5. *We define the single step pseudo regret* $\mathbb{E}[r_{m,t}] = \mathbb{E}[\sigma(\mathbf{S}^{opt}_m, \theta_m) - \sigma(\mathbf{S}_m(t), \theta_m)]$*. With probability* $1-\delta$*,* $r_{m,t}$ *is bounded by*

$$r_{m,t} \leq \mathbb{E}\Big[\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]} \mathcal{O}\Big((\sigma_X + 2|\boldsymbol{Pa}(X)|)\sqrt{|\boldsymbol{Pa}^{(l)}(X)|\log\frac{T}{\delta}}\Big)\|W^{(l)}_{X_m}(t)\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}\Big], \tag{42}$$

where $W_{X_m}(t)$ *denotes a vector satisfying that* $X \in \mathbf{S}_m(t)$*,* $W_{X_m}(t) = \mathbf{0}^{|\boldsymbol{Pa}(X)|}$*; if* $X \notin \mathbf{S}_m(t)$*,* $W_{X_m}(t) = V_{X_m}(t)$*.*

Proof. With probability at most $\delta = NDM\tilde{\delta}$, the event $\{\exists t, X \in \mathcal{V}, l \in [L_X] : \|\hat{\theta}^{(l)}_X(t) - \theta^{(l)}_X\| > \rho^{(l)}_X(t)\}$ occurs. Next we bound the regret conditioned on the absence of this event. In detail, based on Lemma F.3, we can deduce that

$$\begin{aligned}\mathbb{E}[r_{m,t}] &= \mathbb{E}[\sigma(\mathbf{S}^{opt}_m, \theta_m) - \sigma(\mathbf{S}_m(t), \theta_m)]\\ &\leq \mathbb{E}[\sigma(\mathbf{S}_m(t), \tilde{\theta}_m(t)) - \sigma(\mathbf{S}_m(t), \theta_m)]\\ &\leq \mathbb{E}\Big[\sum_{X\in\mathbf{X}_{\mathbf{S}_m(t),Y}} \big|V_{X_m}(t)\cdot(\tilde{\theta}_{X_m}(t) - \theta_{X_m}(t))\big|\Big].\end{aligned} \tag{43}$$

For convenience, we define $W_{X_m}(t)$ as a vector such that if $X \in \mathbf{S}_m(t)$, $W_{X_m}(t) = \mathbf{0}^{|\boldsymbol{Pa}(X)|}$; if $X \notin \mathbf{S}_m(t)$, $W_{X_m}(t) = V_{X_m}(t)$. Then we have

$$
\begin{aligned}
\mathbb{E}[r_{m,t}] &\leq \mathbb{E}\Big[\sum_{X\in\mathcal{V}} \big|W_{X_m}(t)\cdot(\tilde{\theta}_{X_m}(t)-\theta_{X_m}(t))\big|\Big] \\
&= \mathbb{E}\Big[\sum_{X\in\mathcal{V}} \big|\sum_{l\in[L_X]} W^{(l)}_{X_m}(t)\cdot(\tilde{\theta}^{(l)}_{X_m}(t)-\theta^{(l)}_{X_m}(t))\big|\Big] \\
&\leq \mathbb{E}\Big[\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]} \big|W^{(l)}_{X_m}(t)\cdot(\tilde{\theta}^{(l)}_{X_m}(t)-\theta^{(l)}_{X_m}(t))\big|\Big] \\
&\leq \mathbb{E}\Big[\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]} \big\|W^{(l)}_{X_m}(t)\big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}\big\|\tilde{\theta}^{(l)}_{X_m}(t)-\theta^{(l)}_{X_m}(t)\big\|_{M^{(l)}_{X_m}(t)}\Big] \\
&\leq \mathbb{E}\Big[\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]} 2\rho^{(l)}_{X_m}(t)\big\|W^{(l)}_{X_m}(t)\big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}\Big] \\
&\leq \mathbb{E}\Big[\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]} \mathcal{O}\Big(\big(\sigma_X+2|\boldsymbol{Pa}(X)|\big)\sqrt{|\boldsymbol{Pa}^{(l)}(X)|\log\frac{T}{\delta}}\Big)\big\|W^{(l)}_{X_m}(t)\big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}\Big].
\end{aligned}
\tag{44}
$$

□

Now we are ready to prove Theorem 4.2.

Proof. We notice that the cumulative regret of our considered federated combinatorial causal bandits system can be decomposed into three parts. The first part is the regret accumulated under the independent exploration stage. During these rounds we can trivially upper bound the instantaneous regret by two (since $Y_m \in [0,1], \forall m \in [M]$). The second part considers the regret during rounds where our estimated clusters are correct. The third part considers the regret accumulated during the rounds in which our estimated clusters are incorrect, which we can also upper bound the instantaneous regret by two, yielding

$$
\begin{aligned}
R(T) \leq &\sum_{t\in[T_0]}\sum_{m\in[M]} 2 + \sum_{t=T_0+1}^{T}\sum_{m\in[M]} r_{m,t}\cdot\mathbf{1}\{\hat{C}^{(l)}_X = C^{(l)}_X\} \\
&+ \sum_{t=T_0+1}^{T}\sum_{m\in[M]} 2\cdot\mathbf{1}\{\hat{C}^{(l)}_X \neq C^{(l)}_X\}, \forall X\in\mathcal{V}, \forall l\in[L_X].
\end{aligned}
\tag{45}
$$

Based on Theorem 4.1, if we select $\nu^{(l)}_X$ and $T_0 = \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]} T^{(l)}_X$, the probability that $\hat{C}^{(l)}_X \neq C^{(l)}_X, \forall X\in\mathcal{V}, \forall l \in [L_X]$ is $1-\delta$. Therefore, by setting $\delta = \frac{1}{M^2T}$, the regret contributed by the rightmost term is of the order $\mathcal{O}(1)$. Consequently, our high-probability regret bound is given by:

$$
R(T) \leq \mathcal{O}\Big(\frac{M\psi^{(l)}_X\tilde{\sigma}^2_X}{\lambda_c\gamma^2}\Big) + \sum_{t=T_0+1}^{T}\sum_{m\in[M]} r_{m,t}\cdot\mathbf{1}\{\hat{C}^{(l)}_X = C^{(l)}_X\} + \mathcal{O}(1). \tag{46}
$$

In the subsequent steps, we will focus on the regret accumulated when $\hat{C}^{(l)}_X = C^{(l)}_X$. This means we only need to examine the cases that $\forall X\in\mathcal{V}, \forall l\in[L_X]$, when the estimated number of clusters and their compositions exactly match the actual clusters. Consequently, in subsequent discussions $\hat{K}^{(l)}_X = K^{(l)}_X$ and $\hat{C}_k = C_k$ for all $k \in [K^{(l)}_X]$. Hence, in the following discussion, we focus on the below term

$$
R'(T) = \sum_{t=T_0+1}^{T}\sum_{m\in[M]} r_{m,t} \leq \sum_{t=T_0+1}^{T}\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K^l_X]}\sum_{m\in C_k} 2\rho^{(l)}_{X_m}(T)\mathbb{E}\Big[\big\|W^{(l)}_{X_m}(t)\big\|_{\left(M^{(l)}_{X_m}(t)\right)^{-1}}\Big] \tag{47}
$$

In the following, we only focus on the set $\mathcal{C}^{(l)}_X = \{C_k\}_{k\in K^{(l)}_X}$ of clusters corresponding to one specific component subset $l\in[L_X]$ of one specific node $X\in\mathcal{V}$. For each cluster C_k, there are some epochs separated by communication rounds. If there are P_k epochs within cluster C_k, then M_{P_k} will be the matrix with all samples from C_k included. We denote the last globally shared M to the agents in C_k in epoch p as M_p.

Observe that $\det(M_0) = \lambda^{|\boldsymbol{Pa}^{(l)}(X)|}$, and $\det(M_{P,k}) \leq \left(\frac{tr(M_p)}{|\boldsymbol{Pa}^{(l)}(X)|}\right)^{|\boldsymbol{Pa}^{(l)}(X)|} \leq \left(\lambda + \frac{|C_k|\,T}{|\boldsymbol{Pa}^{(l)}(X)|}\right)^{|\boldsymbol{Pa}^{(l)}(X)|}$. Therefore,

$$\log \frac{\det(M_p)}{\det(M_0)} \leq |\boldsymbol{Pa}^{(l)}(X)| \log\left(1 + \frac{|C_k|\,T}{\lambda|\boldsymbol{Pa}^{(l)}(X)|}\right).$$

It follows that for all but $R_k := |\boldsymbol{Pa}^{(l)}(X)| \log\left(1 + \frac{|C_k|T}{\lambda|\boldsymbol{Pa}^{(l)}(X)|}\right)$ epochs,

$$1 \leq \frac{\det(M_j)}{\det(M_{j-1})} \leq 2. \tag{48}$$

In these "good epochs" where Eq (48) is satisfied, we can treat all of the $|C_k|T$ observations from cluster k as observations from an imaginary single agent in a round-robin manner. We similarly use $\tilde{M}_{m,t} = \lambda I + \sum_{\{(p,q):(q<t)\vee(p<m\wedge q=t)\}} V_{p,q}V_{p,q}^\top$ to denote the $\overline{M}_{m,t}$ this agent calculates before seeing $M_{m,t}$. If $M_{m,t}$ is in a good epoch, then:

$$1 \leq \frac{\det(\tilde{M}_{m,t})}{\det(\overline{M}_{m,t})} \leq \frac{\det(M_j)}{\det(M_{j-1})} \leq 2.$$

We similarly denote $\mathcal{B}_{p,k}$ as the set of (m,t) pairs that belong to epoch p and $P_{good,k}$ as the set of good epochs in cluster k. In that event, we can use the regret bound for a single agent which gives

$$\begin{aligned}
R_{good}(T) &= \sum_{t=T_0+1}^{T} \sum_{X\in\mathcal{V}} \sum_{l\in[L_X]} \sum_{k\in[K_X^{(l)}]} \sum_{m\in C_k} 2\rho_{X_m}^{(l)}(T)\left\|W_{X_m}^{(l)}(t)\right\|_{\left(M_{X_m}^{(l)}(t)\right)^{-1}} \\
&\leq \mathcal{O}\Big(\sum_{X\in\mathcal{V}} \sum_{l\in[L_X]} \sum_{k\in[K_X^{(l)}]} 2\rho_{X_m}^{(l)}(T) \sqrt{|C_k|T \sum_{p\in P_{good,k}} \sum_{(m,t)\in\mathcal{B}_{p,k}} \left\|W_{X_m}^{(l)}(t)\right\|^2_{\left(M_{X_m}^{(l)}(t)\right)^{-1}}}\Big) \\
&\overset{(a)}{\leq} \mathcal{O}\Big(\sum_{X\in\mathcal{V}} \sum_{l\in[L_X]} \sum_{k\in[K_X^{(l)}]} 2\rho_{X_m}^{(l)}(T) \\
&\qquad \sqrt{|C_k|T \sum_{p\in P_{good,k}} \sum_{(m,t)\in\mathcal{B}_{p,k}} \frac{|\boldsymbol{Pa}^{(l)}(X)|}{\log(|\boldsymbol{Pa}^{(l)}(X)|+1)} \log\left(1 + \left\|W_{X_m}^{(l)}(t)\right\|^2_{\left(M_{X_m}^{(l)}(t)\right)^{-1}}\right)}\Big),
\end{aligned}$$

where (a) is given by the inequality that $u \leq \frac{a}{\log(1+a)}\log(1+u)$ for $u \in [0,a]$ and $\|W_{X_m}^{(l)}(t)\|^2 \leq |\boldsymbol{Pa}^{(l)}(X)|$.

Then we have

$$
\begin{aligned}
R_{good}(T) &\le \mathcal{O}\Big(\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]} 2\rho_{X_m}^{(l)}(T)\sqrt{|C_k|T\frac{|\boldsymbol{Pa}^{(l)}(X)|}{\log(|\boldsymbol{Pa}^{(l)}(X)|+1)}\sum_{p\in P_{good,k}}\log\Big(\frac{\det(M_p)}{\det(M_{p-1})}\Big)}\Big)\\
&\le \mathcal{O}\Big(\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]} 2\rho_{X_m}^{(l)}(T)\sqrt{|C_k|T\frac{|\boldsymbol{Pa}^{(l)}(X)|}{\log(|\boldsymbol{Pa}^{(l)}(X)|+1)}\log\Big(\frac{\det(M_p)}{\det(M_0)}\Big)}\Big)\\
&\le \mathcal{O}\Big(\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]} 2\rho_{X_m}^{(l)}(T)\sqrt{|C_k|T\frac{|\boldsymbol{Pa}^{(l)}(X)|^2}{\log(|\boldsymbol{Pa}^{(l)}(X)|+1)}\log\Big(\frac{\mathrm{tr}(M_p)}{|\boldsymbol{Pa}^{(l)}(X)|}\Big)}\Big)\\
&\le \mathcal{O}\Big(\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]} 2\rho_{X_m}^{(l)}(T)\sqrt{|C_k|T\frac{|\boldsymbol{Pa}^{(l)}(X)|^2}{\log(|\boldsymbol{Pa}^{(l)}(X)|+1)}\log(|C_k|T)}\Big)\\
&\le \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]}\mathcal{O}\Big(\big(\sigma_X+2|\boldsymbol{Pa}(X)|\big)\sqrt{|\boldsymbol{Pa}^{(l)}(X)|\log\frac{T}{\delta}}\sqrt{|C_k|T\frac{|\boldsymbol{Pa}^{(l)}(X)|^2}{\log(|\boldsymbol{Pa}^{(l)}(X)|+1)}\log(|C_k|T)}\Big)\\
&\le \mathcal{O}\Big(\sum_{X\in\mathcal{V}}|\boldsymbol{Pa}(X)|^2\sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]}\sqrt{|\boldsymbol{Pa}^{(l)}(X)||C_k|T}\log(|C_k|T)\Big).
\end{aligned}
\tag{49}
$$

Now we focus on the regret incurred by the bad epochs, where $R_k = \mathcal{O}(|\boldsymbol{Pa}^{(l)}(X)|\log(|C_k|T))$ within each cluster $C_k \in \mathcal{C}_X^{(l)}$. Consider the regret for a particular cluster $C_k \in \mathcal{C}_X^{(l)}$. Assume that the epoch starts from round t_0, and the length of the epoch is t_k. For any node and any subset of components, we have $M_{X_m}^{(l)}(t) - \Delta M_{X_m}^{(l)}(t)$ at the start of the bad epoch. The regret in this epoch satisfies

$$
\begin{aligned}
R_{bad}(T) &= \sum_{t=T_0+1}^{T}\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]}\sum_{m\in C_k} 2\rho_{X_m}^{(l)}(T)\big\|W_{X_m}^{(l)}(t)\big\|_{\left(M_{X_m}^{(l)}(t)\right)^{-1}}\\
&\le \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K_X^{l}]}\sum_{m\in C_k}\sum_{t=t_0}^{t_0+t_k} 2\rho_{X_m}^{(l)}(T)\big\|W_{X_m}^{(l)}(t)\big\|_{\left(M_{X_m}^{(l)}(t)\right)^{-1}}\\
&\le \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K_X^{l}]}\sum_{m\in C_k} 2\rho_{X_m}^{(l)}(T)\sqrt{t_k\log\frac{\det(M_{X_m}^{(l)}(t_0+t_k))}{\det(M_{X_m}^{(l)}(t_0+t_k)-\Delta M_{X_m}^{(l)}(t_0+t_k))}}
\end{aligned}
\tag{50}
$$

For all but one agent, $t_k\log\frac{\det(M_{X_m}^{(l)}(t_0+t_k))}{\det(M_{X_m}^{(l)}(t_0+t_k)-\Delta M_{X_m}^{(l)}(t_0+t_k))} < D_k$. Hence we can show that

$$
R_{bad}(T) \le \mathcal{O}\Big(\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{k\in[K_X^{l}]}\big(\sigma_X+2|\boldsymbol{Pa}(X)|\big)\sqrt{|\boldsymbol{Pa}^{(l)}(X)|\log\frac{T}{\delta}}|C_k|\sqrt{D_k}\Big).
\tag{51}
$$

Since we know that at most $R_k = \mathcal{O}(|\boldsymbol{Pa}^{(l)}(X)|\log(|C_k|T))$ bad epochs, we can upper bound it by

$$
R_{bad}(T) \le \mathcal{O}\Big(\sum_{X\in\mathcal{V}}|\boldsymbol{Pa}(X)|^2\sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]}\sqrt{|\boldsymbol{Pa}^{(l)}(X)|}|C_k|\sqrt{D_k}\log^{1.5}\big(|C_k|T\big)\Big).
\tag{52}
$$

Recall that we select $D_k = \frac{T\log(|C_k|T)}{|\boldsymbol{Pa}^{(l)}(X)||C_k|}$, the regret caused by bad epochs becomes:

$$
R_{bad}(T) \le \mathcal{O}\Big(\sum_{X\in\mathcal{V}}|\boldsymbol{Pa}(X)|^2\sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]}\sqrt{|C_k|T}\log^2(|C_k|T)\Big).
\tag{53}
$$

Combining the regret from the independent exploration phase, good epochs, and bad epochs gives a final cumulative regret upper bound of:

$$R(T) \leq \mathcal{O}\Big(\frac{M\psi_X(\sigma_X^2+4D^2)}{\lambda_c\gamma^2} + \sum_{X\in\mathcal{V}}|\boldsymbol{Pa}(X)|^2 \sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]}\sqrt{|\boldsymbol{Pa}^{(l)}(X)||C_k|T}\log(|C_k|T) \\ + \sum_{X\in\mathcal{V}}|\boldsymbol{Pa}(X)|^2 \sum_{l\in[L_X]}\sum_{k\in[K_X^{(l)}]}\sqrt{|C_k|T}\log^2(|C_k|T\Big) \tag{54}$$

COMMUNICATION COST

The cumulative communication $C(T)$ of our algorithm consists of two parts. The first part is the communication cost associated with the independent exploration phase and cluster estimation phase. During the independent exploration phase, no agents communicate with the server, so that the communication cost corresponding to this phase is just zero. At the end of the independent exploration phase, all the agents $m \in [M]$ upload sufficient statistics $\{M_{X_m}^{(l)}(T_X^{(l)})\}_{X\in\mathcal{V},l\in[L_X]}$ and $\{\hat{b}_{X_m}^{(l)}(T_X^{(l)})\}_{X\in\mathcal{V},l\in[L_X]}$ to the central matrix. We notice that $M_{X_m}^{(l)}(T_X^{(l)}) \in [0,1]^{|\boldsymbol{Pa}^{(l)}(X)|\times|\boldsymbol{Pa}^{(l)}(X)|}$ and $\hat{b}_{X_m}^{(l)}(T_X^{(l)}) \in [0,1]^{|\boldsymbol{Pa}^{(l)}(X)|\times 1}$ respectively. Therefore, the communication cost of the cluster estimation is phase is $C_{clustering} = (\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}\sum_{m\in[M]}(|\boldsymbol{Pa}^{(l)}(X)|^2+|\boldsymbol{Pa}^{(l)}(X)|) = \mathcal{O}(M(\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}|\boldsymbol{Pa}^{(l)}(X)|^2)$.

Then we turn to analyze the communication cost of the second phase, the collaborative learning phase. For the communication protocol considered in this paper, for each cluster $C_k \in [K_X^{(l)}]$ associated with one specific node $X \in \mathcal{V}$ and one specific subset $l \in [L_X$ of components, there are a number of epochs separated by communication rounds. Denote the length of an epoch as α_k, so that there can be at most $\lceil\frac{T}{\alpha_k}\rceil$ epochs with length longer than α_k. For an epoch with less than α_k time steps, similarly, we denote the first time step of this epoch as $t_k(s)$ and the last as $t_k(e)$, i.e., $t_k(e)-t_k(s)<\alpha_k$. Therefore, $\log\frac{\det(M_{t_k(e)})}{\det(M_{t_k(s)})} > \frac{D_k}{\alpha}$. Following the same argument as in the regret proof, the number of epochs with less than α_k time steps is at most $\lceil\frac{R_k\alpha_k}{D_k}\rceil$. Then $C_{collaboration}(k) = |C_k|\cdot(\lceil\frac{T}{\alpha_k}\rceil+\lceil\frac{R_k\alpha}{D_k}\rceil)$, because at the end of each epoch, the synchronization round incurs $2|C_k|$ communication cost. We minimize $C_{collaboration}(k)$ by setting $\alpha_k = \sqrt{\frac{D_kT}{R_k}}$, so that $C_{collaboration}(k) = O(|C_k|\sqrt{\frac{TR_k}{D_k}})$. With our choice of $D_k = \frac{T\log|C_k|T}{|\boldsymbol{Pa}^{(l)}(X)||C_k|}$, we have

$$\begin{aligned} C_{collaboration}(k) &= O(|C_k|\sqrt{\frac{TR_k}{\frac{T\log|C_k|T}{|\boldsymbol{Pa}^{(l)}(X)||C_k|}}}), \\ &= O(|C_k|\cdot|\boldsymbol{Pa}^{(l)}(X)|\sqrt{|C_k|}). \end{aligned}$$

Combining our communication cost from our two phases together gives:

$$C(T) = \mathcal{O}\Big(M\sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}|\boldsymbol{Pa}^{(l)}(X)|^2 + \sum_{X\in\mathcal{V}}\sum_{l\in[L_X]}|\boldsymbol{Pa}^{(l)}(X)|^3\sum_{k\in[K_X^{(l)}]}|C_k|^{1.5}\Big). \tag{55}$$

□

G LOWER BOUND FOR FEDCCB UNDER THE ONE-LAYER PARALLEL GRAPH

In this section, we present a lower bound analysis for FedCCB with heterogeneous causal influences to demonstrate the optimality of our proposed FedSCuB method. Since deriving lower bounds for general linear threshold models remains an open problem, here we focus on a special causal graph consisting of a one-layer parallel structure where all parent nodes directly influence a single target node. Following the convention of online clustering of bandits Li and Zhang [2018], Liu et al. [2022], we consider deriving the lower bound under the known cluster structure for any parent subset of the target node.

To establish this lower bound, we first revisit Assumption 2.1 as stated in the main text, and then define the problem instances considered in this section.

Assumption G.1 (Proximity within clusters)**.** For any node $X \in \mathcal{V}$ and any subset of parents $\boldsymbol{Pa}^{(l)}(X), l \in [L_X]$, if any two agents $m, m' \in [M]$ belong to a specific cluster in $\mathcal{C}_X^{(l)}$, we have $\|\theta_{X_m}^{(l)} - \theta_{X_{m'}}^{(l)}\| \leq \epsilon \leq \frac{1}{M\sqrt{DT}}$.

Definition G.2 (p-norm ϵ-FedCCB under one-layer parallel graph instance)**.** We define the class of p-norm ϵ-FedCCB under one-layer parallel graph with the target node Y as $\mathcal{I}_p^{L_Y}(\epsilon)$, if for any $l \in [L_Y]$ and any two agents $m, m' \in [M]$ belong to a specific cluster in $C_Y^{(l)}$, $\|\theta_{Y_m}^{(l)} - \theta_{Y_{m'}}^{(l)}\|_p \leq \epsilon$.

Now we begin to introduce the lower bound for FedCCB under the one-layer parallel graph formally.

Theorem G.3 (Lower bound for FedCCB under the One-layer Parallel Graph)**.** *Let $\mathcal{I}^{L_Y}(\epsilon)$ denote the class of ϵ-FedCCB under one-layer parallel graph instances, we have the following,*

$$\inf_{\mathcal{A}} \sup_{\mathcal{I} \in \mathcal{I}^{L_Y}(\epsilon)} \mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T) = \Omega\left(\sum_{l \in [L_Y]} \sum_{k \in [K_Y^{(l)}]} |\boldsymbol{Pa}^{(l)}(Y)| \sqrt{|C_k|T}\right). \tag{56}$$

Remark G.4. We note that our algorithm is near-optimal under the setting of FedCCB under one-layer parallel graph, since its regret upper bound $\tilde{\mathcal{O}}(|\boldsymbol{Pa}(Y)|^2 \sum_l \sum_k \sqrt{|C_k|T})$ differs from the lower bound by only a logarithmic factor, which is common in general linear bandits. Furthermore, although the upper bound of the proposed algorithm contains a factor of $|\boldsymbol{Pa}(Y)|^2$, this deterioration stems from the fact that the algorithm uses the AM method to estimate the partial response value, which has little impact on the regret, as causal graphs are sparse in most applications.

Proof. Based on Assumption 2.1, we can prove that solving any ϵ-FedCCB under one-layer parallel graph instance is at least as hard as solving a clustered linear bandits for $|C_k|T$ rounds within each cluster $C_k, k \in [K_Y^{(l)}]$ when $\epsilon = 0$ and true partial responses can be attained. We prove it by contradiction, which is based on the linear bandits lower bound [Lattimore and Szepesvári, 2020]. For $\epsilon \geq 0$, $\inf_{\mathcal{A}} \sup_{\mathcal{I} \in \mathcal{I}_\infty^{L_Y}(\epsilon)} \mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T) \geq \inf_{\mathcal{A}} \sup_{\mathcal{I} \in \mathcal{I}_\infty^{L_Y}(0)} \mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T)$.

Below we assume there exists an algorithm $\mathcal{A}$ achieving

$$\sup_{\mathcal{I} \in \mathcal{I}_\infty^{L_Y}(0)} \mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T) < c \sum_{l \in [L_Y]} \sum_{k \in [K_Y^{(l)}]} |\boldsymbol{Pa}^{(l)}(Y)| \sqrt{|C_k|T}, \tag{57}$$

where c denotes some constant.

We observe that $\mathcal{I}_\infty^{L_Y}(0)$ is the class multi-agent solving a clustered linear bandits problem with exactly the same parameters within each cluster. Then for any $l \in [L_Y]$ and $k \in [K_Y^{(l)}]$, we simulate the algorithm $\mathcal{B}$ on a single agent representing a specific cluster. This virtual agent solves linear bandits for $|C_k|T$ rounds by the protocol of $|C_k|$ agents solving an identical linear bandits for T rounds. Hence, if we have $\mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T) \leq c \sum_{l \in [L_Y]} \sum_{k \in [K_Y^{(l)}]} |\boldsymbol{Pa}^{(l)}(Y)| \sqrt{|C_k|T}$, we also have $\mathcal{R}_{\mathcal{B},\mathcal{I}}(\{|C_k|T\}_{k \in [K_Y^{(l)}]}) \leq c \sum_{l \in [L_Y]} \sum_{k \in [K_Y^{(l)}]} |\boldsymbol{Pa}^{(l)}(Y)| \sqrt{|C_k|T}$, which contradicts the lower bound of single-agent linear bandits [Lattimore and Szepesvári, 2020]. Thus we have $\sup_{\mathcal{I} \in \mathcal{I}_\infty^{L_Y}(0)} \mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T) \geq c \sum_{l \in [L_Y]} \sum_{k \in [K_Y^{(l)}]} |\boldsymbol{Pa}^{(l)}(Y)| \sqrt{|C_k|T}$.

Since $\|x\|_\infty \leq \epsilon$ implies $\|x\|_2 \leq \epsilon\sqrt{d}$, we have $\mathcal{I}_\infty^{L_Y}(0) \subset \mathcal{I}_2^{L_Y}(0)$. In words, we have

$$\sup_{\mathcal{I} \in \mathcal{I}_2^{L_Y}(0)} \mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T) = \Omega\left(\sum_{l \in [L_Y]} \sum_{k \in [K_Y^{(l)}]} |\boldsymbol{Pa}^{(l)}(Y)| \sqrt{|C_k|T}\right). \tag{58}$$

Besides, following Do et al. [2023b], we know that for the partition $[L_Y]$ and cluster structure $[K_Y^{(l)}]$, the regret can be also lower bounded as $\inf_{\mathcal{A}} \sup_{\mathcal{I} \in \mathcal{I}_2^{L_Y}(\epsilon)} \mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T) = \Omega\left(\epsilon \sum_{l \in [L_Y]} \sum_{k \in [K_Y^{(l)}]} |C_k|T\right)$. By Assumption 2.1, we have $\epsilon \leq \frac{1}{M\sqrt{DT}}$, which implies: $\inf_{\mathcal{A}} \sup_{\mathcal{I} \in \mathcal{I}_2^{L_Y}(\epsilon)} \mathcal{R}_{\mathcal{A},\mathcal{I}}(M,T) = \Omega\left(\sum_{l \in [L_Y]} \sum_{k \in [K_Y^{(l)}]} \sqrt{T}\right)$. This result is dominated by Equation (58).

To sum up, we have finished the proof. □

H EXPERIMENT DETAILS

In this section, we provide the details of running simulations on linear causal model with our proposed FedSCuB method and other baselines. We show that our algorithms have much smaller best-in-hindsight regrets than the baseline algorithms. This is consistent with our discussions in Section 3. Also, we further compare the performance of conducting clustering on different levels of component subset.

EXPERIMENTAL SETTINGS

In this subsection, we detail our simulation settings. We primarily follow the data generation approach proposed in Feng and Chen [2023] and Blaser et al. [2024]. We set the number of communication rounds $T = 4000$ and the number of participating agents $M = 16$. Following Feng and Chen [2023], we adopt the parallel graph and two-layer graph as our simulated model. In specific, we consider three distinct parallel graphs and one specific two-layer graph in the following. For parallel graphs, we set the number of nodes as 6, 8 and 10 respectively.

Taking the parallel graph with eight nodes as the example. The simulated graph is a parallel linear model where X_1 points to $X_2, X_3, \ldots, X_7$, and $X_2, X_3, \ldots, X_7$ all point to Y. Here, X_1 is always set to 1. Its corresponding topological structure is illustrated in Figure 3. Next, we describe how we generate the weights of the incoming edges for all other nodes, i.e., how we construct the parameter vectors for these nodes. In this experiment, we focus on clustering and collaborative learning for the parameter vector of the target variable Y. Specifically, we assume that $\boldsymbol{Pa}(Y) = \{X_2, X_3, \ldots, X_7\}$ can be partitioned into two groups, $\{\boldsymbol{Pa}^{(1)}(Y), \boldsymbol{Pa}^{(2)}(Y)\}$, where $\boldsymbol{Pa}^{(1)}(Y) = \{X_2, X_3, X_4\}$ and $\boldsymbol{Pa}^{(2)}(Y) = \{X_5, X_6, X_7\}$. The corresponding parameter vector is $\theta_Y = (\theta_Y^{(1)}, \theta_Y^{(2)}) \in [0, 1]^6$. Here, $\theta_Y^{(1)}$ represents the vector whose components are the weights of the directed edges from $\{X_2, X_3, X_4\}$ to Y, respectively. Similarly, $\theta_Y^{(2)}$ denotes the vector whose components are the weights of the directed edges from $\{X_5, X_6, X_7\}$ to Y, respectively.

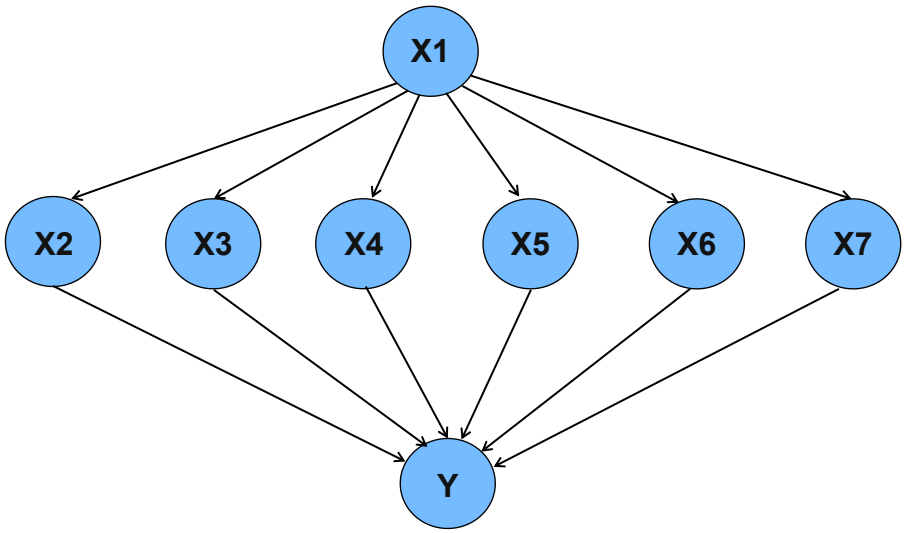


Figure 3: Topology structure of parallel graph with eight nodes.

Besides, we utilize the two-layer causal graph introduced in Feng and Chen [2023] to validate the method's generalizability. The experimental setup employs an eight-node, two-layer causal graph. To maintain consistency with previous experimental settings for parallel graphs, we specifically introduced grouped heterogeneity in the second layer: the four nodes pointing to the target node Y were divided into two groups (X_4,X_5 and X_6, X_7), with distinct edge weight clustering structures among 16 agents for each group's connections to Y. The topology structure of this two-layer graph is shown in Figure 4.

For each group, we follow the data generation method proposed in Blaser et al. [2024] to sample cluster centers from a standard normal distribution, ensuring that they are at least $\gamma + 2\epsilon$ apart from each other. We then randomly assign each agent to one of these clusters. We set the number of clusters to 3 for each group. We set the hyperparamters γ to 0.85, and

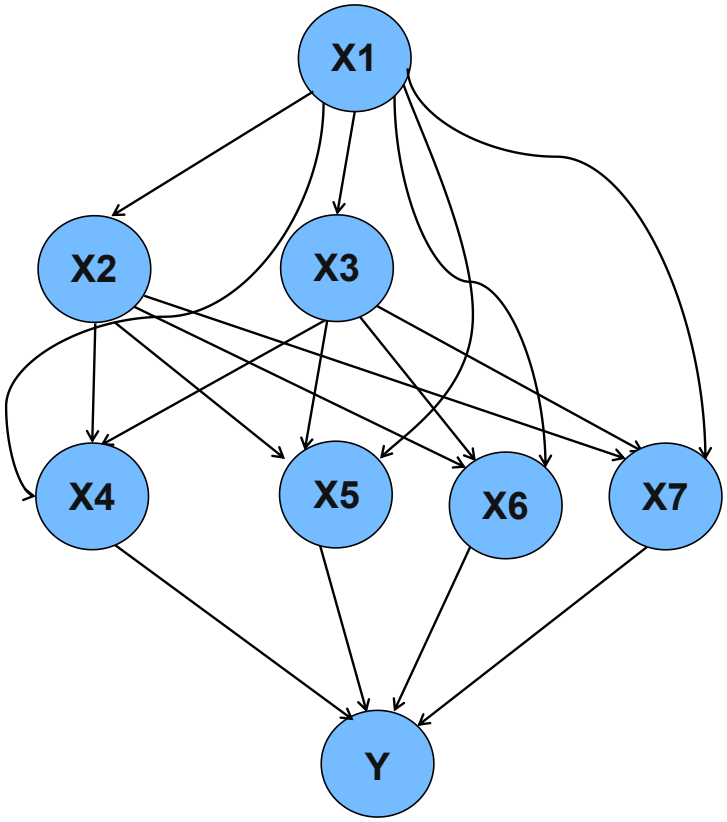


Figure 4: Topology structure of two-layer graph.

$\epsilon = \frac{1}{M\sqrt{T}}$ respectively. About the independent exploration phase, we set $T_0 = \sum_X \sum_l T_X^{(l)}$ with known γ and ϵ.

BASELINE METHODS

In our evaluation, we compared our proposed FedSCuB method with below baseline methods: 1) independent optimistic learning for FedCCB proposed in Feng and Chen [2023]; 2) distributed optimistic learning for FedCCB via the communication protocol proposed in DisLinUCB [Wang et al., 2020], which simply aggregate sufficient statistics from all the other agents; c) LinUCB-AM proposed in Li and Wang [2022b] partition the context vector into two parts including the global features and local features. This method utilizes the alternating minimization method to update parameter vector via partial collaboration on the globally shared component, while conducts independent learning for the local personalized component ; d) HetoFedBandit [Blaser et al., 2024] conducts clustering on the level of the overall parameter vector, and aggregates sufficient statistics in terms of the overall parameter vector.

EXPERIMENTAL RESULTS

The effect of clustering level on collaboration for FedSCuB

In this section, we conduct experiments to evaluate the impact of different clustering levels on the regret of FedSCuB. We use the parallel graph with eight nodes for this experiment, whose topology is shown in Figure 3.For this experiment, we simulate 30 times of each setting, and draw the 95% confidence intervals of average regrets. Recall that we partition the parameter vector of the target node Y into two groups and generate distinct clusters for each group. In the following experiment, we compare performance across four clustering and collaboration levels: a) clustering based on the overall parameter vector; b) clustering based on two groups of components, aligned with the real data generation process; c) clustering based on three groups of components, where each group contains two components; d) individual component-wise clustering.

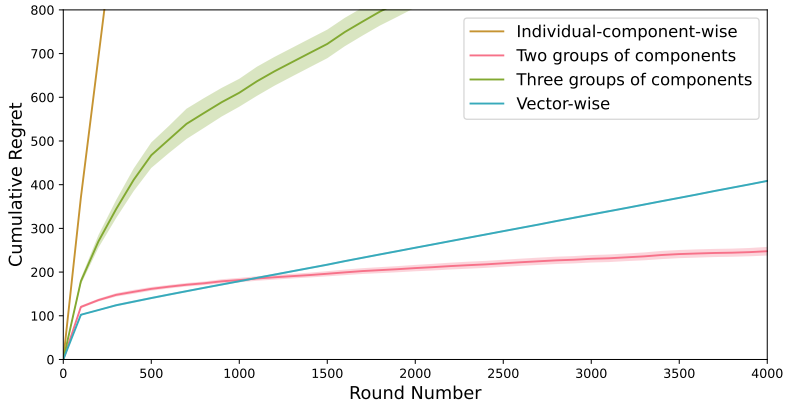


Figure 5: Different clustering levels for FedCCB.

From Figure 5, we observe that clustering and collaboration based on the ground-truth grouping (i.e., two groups, each with three components) outperforms the other scenarios. This is because this approach aligns with the actual data generation process. Furthermore, we find that adopting a more fine-grained clustering level results in worse regret. Specifically, component-wise clustering performs worse than clustering based on three groups of components. Additionally, both component-wise clustering and clustering on three groups of components yield significantly higher regret compared to vector-wise clustering. These empirical results are consistent with our theoretical findings.

Additional experiment on the two-layer causal graph

In Figure 6, we compare FedSCuB with baselines introduced above on the two-layer graph whose topology structure is illustrated in Figure 4. The experimental results demonstrate that even in more complex two-layer graph structures, our proposed method continues to exhibit significantly superior performance compared to other baseline algorithms, which demonstrates that our proposed method can be generalized to more complex setting.

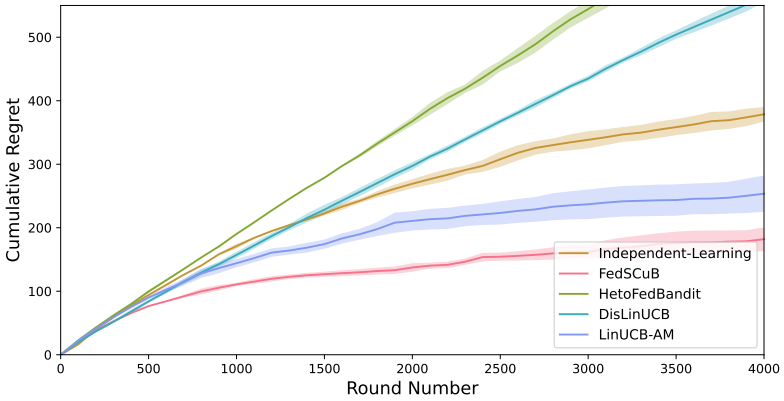


Figure 6: Different baseline methods for FedCCB on the two-layer graph with eight nodes.