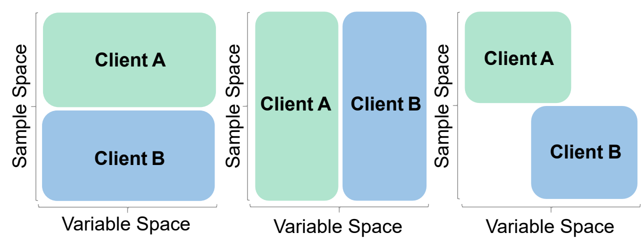
Horizontal and Vertical Federated Causal Structure Learning via Higher-order Cumulants

Wei Chen[1], Wanyang Gu[1], Linjun Peng[1], Ting Yan[1], Ruichu Cai[1,2*], Zhifeng Hao[3], Kun Zhang[4,5]
[1]School of Computer Science, Guangdong University of Technology, Guangzhou, China
[2]Peng Cheng Laboratory, Shenzhen, China
[3]College of Science, Shantou University, Shantou, China
[4]Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA, United States
[5]Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates
chenweidelight@gmail.com, guwanyang@mails.gdut.edu.cn, yting030303@163.com, pljnb666@gmail.com, cairuichu@gmail.com, haozhifeng@stu.edu.cn, kunz1@cmu.edu

Abstract

Federated causal discovery aims to uncover causal relationships while protecting data privacy, with significant real-world applications. Existing methods focus on horizontal federated settings where clients share the same variables but have different samples. However, in practice, clients may have different variables, leading to spurious causal relationships. To address this issue, we comprehensively consider causal structure learning methods under both horizontal and vertical federated settings. Interestingly, we find that, higher-order cumulants rely solely on the joint distribution of the relevant variables and are useful to solve the above problem in the linear non-Gaussian case. This motivates us to provide the identification theories for determining the causal order over observed variables, leveraging the difference in the product of the (cross) cumulants of the specific variables. Based on these theories, we develop a method for learning causal order in the horizontal and vertical federated scenarios. Specifically, we first obtain local (cross) cumulant matrices of observed variables from all participating clients to construct a global cumulant matrix. This global cumulant matrix is then used for recursive source variable identification, ultimately yielding a causal strength matrix of the union of variables from all clients. Our algorithm demonstrates superior performance in experiments on both synthetic and real-world data.



(a) Horizontal Setting (b) Vertical Setting (c) Hybrid Setting

Figure 1: Three types of federated setting.

Introduction

Causal structure learning focus on revealing the causal relationships between variables from observational data (Spirtes, Glymour, and Scheines 2000; Wang and Drton 2020; Chen et al. 2021), which is widely applied in various fields, including Bioinformatics analysis(Zhang et al. 2022), Neuroscience (Cai et al. 2024), recommendation system (Huang et al. 2025), and time series forecasting (Li et al. 2023). Most existing methods for causal structure learning rely on centralized data environments and do not consider privacy issues. With the increasing frequency of data leakage incidents and heightened public awareness regarding data security, a growing number of individuals and organizations are hesitant to share their private datasets (Ng et al. 2021; Yang et al. 2020). This trend presents significant challenges to traditional causal discovery methods. Consequently, federated causal discovery (Guo et al. 2024b; Yu et al. 2025), which aims to uncover causal relationships from decentralized data sources, has garnered considerable attention and is becoming increasingly important in real-world applications. In federated scenarios, there are three main types: 1) horizontal federated setting, where data across different institutions share the same variable space but different samples; 2) vertical federated setting, where data across different institutions share the same samples but different variable spaces; and 3) hybrid federated setting that combined horizontal and vertical federated, as illustrated in Figure 1. Most existing federated causal discovery (FCD) focuses on a horizontal setting, operating at the sample level while addressing the heterogeneity of client data. One typical horizontal federated causal discovery method is based on constraints, which utilizes federated conditional independence tests or scores to search the causal graph (Wang, Ma, and Wang 2023; Huang et al. 2023; Li et al. 2024; Ng et al. 2021). The proposed federated conditional independence tests can address the challenge of independence testing in scenarios where client samples differ. By leveraging the structural information learned from clients and returning it to the server, these methods enable the aggregation of structural information to derive the causal structure (Huang et al. 2023; Guo et al. 2024a,b,c).

However, some causal relationships cannot be uniquely

*Corresponding author.

recovered by utilizing the federated conditional independence tests, since there is a set of directed acyclic graphs (DAGs) that are Markov equivalent to a given DAG and encode the same set of conditional independence relationships. Moreover, those horizontal federated causal discovery methods fail in the vertical federated setting. Clients may be unable to identify complete conditional sets, leading to inaccurate conditional independence (CI) tests. This introduces spurious edges in local causal graphs. Subsequent global aggregation on the server then propagates these errors, resulting in numerous redundant edges in the global causal structure. For each client, the lack of information on all variables prevents them from effectively obtaining the statistics needed for conditional independence testing. Although the CDMiNi method (Huang et al. 2020) is proposed to tackle the non-identical variables set problem, it can access all the original datasets. Recently, Wang et al. (Wang et al. 2025) proposed FedCDnv to learn causal structures from decentralized data with non-identical variable sets, though some causal structures remain unidentifiable due to limited information from conditional independence test results.

To tackle the above issues, we aim to provide a method for federated causal structure learning in both horizontal and vertical federated settings. Interestingly, we find that, higher-order cumulants rely solely on the joint distribution of the relevant variables and are not influenced by the absence of variables (Brillinger 2001; Cai et al. 2023). In the linear non-Gaussian case, higher-order cumulants can capture the information of the absence of variables from observed variables within each client (Robeva and Seby 2021; Dai, Spirtes, and Zhang 2022; Chen et al. 2024; Schkoda and Drton 2025). Inspired by these, we propose identifiability theories for horizontal and vertical federated causal order determination via higher-order cumulants. Based on these theories, we develop a causal structure learning method for both horizontal and vertical federated scenarios, named FedISHC. Specifically, FedISHC first collects all cross-cumulant matrices estimated from all clients and aggregates them into a global cross-cumulant matrix on the server. Second, it iteratively identifies the sources utilizing the elements in the cross-cumulant matrix to remove the influence from the identified ancestors. Finally, it learns the causal structure over all observed variables from the union of observed variables across all clients. Experimental results in synthetic data and real-world data demonstrate the effectiveness of our proposed method.

The contributions of this paper are summarized as: 1) We provide theories on identifying causal order between two observed variables without common ancestors and eliminating the influence of common ancestors on the server side. 2) We propose a source variable recovery method in the multiple observed variables case, and develop a horizontal and vertical federated causal structure learning without data leakage. 3) Experimental results validate that our method can effectively perform causal structure learning without directly accessing raw client data, while requiring only a single communication between the client and the server, ensuring privacy and reducing computational complexity.

Preliminary

Linear Non-Gaussian Acyclic Model

In this paper, we consider the data over d-dimensional vector of observed variables $X = (x_1, x_2, \ldots, x_d)$ are generated by the Linear Non-Gaussian Acyclic Model (LiNGAM) (Shimizu et al. 2006), which is formalized as:

$$\mathbf{X} = \mathbf{B}\mathbf{X} + \mathbf{E}, \tag{1}$$

where $\mathbf{E}$ is a vector of noise terms, each $e_i \in \mathbf{E}$ follows Non-Gaussian distributions and is independent of each other. $\mathbf{B}$ is the causal strength matrix, describing the influences among X. $b_{ij} \in \mathbf{B}$ represents the element in the i-th row and j-th column of $\mathbf{B}$, indicating the causal influence of variable x_j on variable x_i. Since LiNGAM is acyclic, B can be permuted to a strictly lower triangular matrix.

Cumulants

The cumulants are a set of statistical measures that characterizing the shape of a probability distribution, which can capture the information of random variables. The definition of cumulants of a random vector X is given as follows.

Definition 1 (Cumulants (Brillinger 2001)) *Let $X = (X_1, X_2, \ldots, X_d)$ be a random vector of length d. The k-th order cumulant tensor of X is defined as a $d \times \cdots \times d$ (k times) table, $\mathcal{C}^{(k)}$, whose entry at position $(i_1, \ldots, i_k)$ is*

$$\begin{aligned}\mathcal{C}^{(k)}_{i_1,\cdots,i_k} &= Cum(X_{i_1}, \ldots, X_{i_k}) \\ &= \sum_{(D_1,\ldots,D_h)} (-1)^{h-1}(h-1)!\mathbb{E}\left[\prod_{j\in D_i} X_j\right] \cdots \mathbb{E}\left[\prod_{j\in D_h} X_j\right],\end{aligned} \tag{2}$$

where the sum is taken over all partitions $(D_1, \ldots, D_h)$ of the set $\{i_1, \ldots, i_k\}$.

For clarity, we use $C_m(x_i)$ to denote the m-th order cumulant of x_i $Cum(\underbrace{x_i, \ldots}_{m \text{ times}})$, and use $C_{m,n}(x_i, x_j)$ to denote the cross cumulants $Cum(\underbrace{x_i, \ldots}_{m \text{ times}}, \underbrace{x_j, \ldots}_{n \text{ times}})$. For example, $C_3(x_i)$ represents $Cum(x_i, x_i, x_i)$, $C_{1,2}(x_i, x_j)$ represents $Cum(x_i, x_j, x_j)$.

The first-order cumulant of a random variable x_i corresponds to its mean, the second-order cumulant is equivalent to the variance, and the third-order cumulant is the same as the third central moment. But fourth and higher-order cumulants are not equal to central moments. Although the third and higher-order cumulants of a normal distribution are zero, it is non-zero for the non-Gaussian distributions with some specific orders (Feller 1991). For the data generated by LiNGAM, high-order cumulants serve as key tools to capture the high-order characteristics of the data and reveal the essence of the distribution. In addition, cumulants have the additivity properties: cumulants of independent random variables are additive. When two or more random variables are statistically independent, the k-th order cumulant of their sum is equal to the sum of their k-th order cumulants.

Problem Definition and Assumption

Consider that there are K clients and one central server in the federated setting. Each client has a local dataset D^k over observed variables X^k $(k = \{1, \ldots, K\})$ where k is the index of the client. Let d^k denote the dimension of X^k. Different clients may contain different sets of variables, i.e., the variable sets across clients are non-identical. Let X be the union of variables from all K clients, i.e., $X = \bigcup_{k=1}^{K} X^k$, with a total of d variables. Each client has n_k samples. Based on this setting, we aim to recover the causal relationships among all observed variables X within the context of horizontal and vertical federated learning. The necessary assumptions for this paper are presented as Assumption 1.

Assumption 1 *Let X be the union of variables from all K datasets of clients, i.e., $X = \bigcup_{k=1}^{K} X^k = (x_1, x_2, \ldots, x_d)$, with a total of d variables. For any pair of variables x_i and x_j, there exists at least one client containing both.*

Note that this assumption is mild, reasonable, and practical. This assumption ensures that each pair of variables has a certain degree of correlation present in the data of at least one client, thereby helping the model capture the relationships between variables. Compared to each client containing a subset of all variables, this assumption reduces data redundancy and storage overhead. Each client only needs to include a portion of the variable combinations. This assumption provides greater flexibility, allowing different clients to have different combinations of variables, thus better adapting to the actual data distribution and business needs. Considering privacy protection, in some cases, clients may be unwilling to share data for all variables. This assumption allows clients to share only a portion of information of the variables, thereby better protecting privacy.

Horizontal and Vertical Federated Causal Order Identification Criterion

In federated learning scenarios that involve both horizontal and vertical data partitioning, the data samples and variable sets across different clients are not completely identical. This discrepancy renders existing federated conditional independence testing methods ineffective. For instance, suppose there are three variables x_1, x_2, and x_3, and x_1 and x_2 are conditionally independent given x_3. In practice, no single client contains data for all three variables simultaneously. Although it is possible to obtain joint distribution information for any pair of variables among x_1, x_2, and x_3 from different clients, it is insufficient to acquire the joint distribution and conditional probability distribution information for x_1, x_2, and x_3 together. Consequently, it is impossible to compute the statistical measures required for conditional independence testing, making it infeasible to determine whether x_1 and x_2 are conditionally independent given x_3. Although we find that the partial correlation coefficient can be easily extended to the federated case, it is effective when the data are linear and follow Gaussian distributions (the details in given in the Appendix). In this paper, we focus on the linear non-Gaussian case. Fortunately, inspired by the higher-order cumulants that can capture the statistical information from the data, we find that using the (cross) cumulants of different variables can obtain the causal relationships among three observed variables, even when they do not exist in the same client or their samples are different in all considered clients.

Identification of Causal Order without Common Ancestors

Let us first consider the case without common ancestors for two observed variables. Take a simple case as an example. Suppose that there are two observed variables x_i and x_j, where x_i is a direct cause of x_j and x_i and x_j have no common ancestors. This can be simply formalized as $x_i = e_i$, $x_j = b_{ji}x_i + e_j$. Some (cross) cumulants of x_i and x_j can be expressed as follows:

$$\begin{aligned} &C_3(x_i) = C_3(e_i), \\ &C_3(x_j) = b_{ji}^3 C_3(e_i) + C_3(e_j), \\ &C_{1,2}(x_i, x_j) = C_{2,1}(x_j, x_i) = b_{ji}^2 C_3(e_i), \\ &C_{2,1}(x_i, x_j) = C_{1,2}(x_j, x_i) = b_{ji} C_3(e_i). \end{aligned} \tag{3}$$

Then we can observe an interesting conclusion that $C_3(x_i)C_{1,2}(x_i, x_j) - C_{2,1}(x_i, x_j)C_{1,2}(x_j, x_i) = 0$, and $C_3(x_j)C_{1,2}(x_j, x_i) - C_{2,1}(x_j, x_i)C_{1,2}(x_i, x_j) = b_{ji}C_3(e_i)C_3(e_j) \neq 0$, this is because there is an extra term e_j in x_j that cannot be eliminated, which is induced by the result variable. Motivated by these, we derive a causal order identification measure, denoted as τ_{ij}, which is defined as

$$\tau_{ij} = |C_3(x_i)C_{1,2}(x_i, x_j) - C_{2,1}(x_i, x_j)C_{1,2}(x_j, x_i)|. \tag{4}$$

By making use of τ_{ij} and τ_{ji}, we can determine the causal order between x_i and x_j that are not affected by confounders, which is guaranteed by the following theorem.

Theorem 1 *Assumed that two observed variables x_i and x_j are generated by Eq. (1) and x_i and x_j have no common ancestor. Then,*

1. *If and only if both $\tau_{ij} = 0$ and $\tau_{ji} = 0$ hold, then x_i and x_j are mutually independent;*
2. *If and only if both $\tau_{ij} = 0$ and $\tau_{ji} \neq 0$ hold, then x_i is causal earlier than x_j.*

Identification of Causal Order with Common Ancestors

If two observed variables x_i and x_j have at least one common ancestor, Theorem 1 fails to identify the causal relationship between x_i and x_j. But one may think that if the data is generated by LiNGAM, the effect from the common ancestor to its descendant can be eliminated by regression, and the causal relationship between the residuals still follow the LiNGAM (Shimizu et al. 2011). Thus, Theorem 1 still holds for residuals. Suppose that there is a common cause x_k influencing x_i and x_j simultaneously, and x_i is a cause of x_j. Then, the residual of x_i after regressing out x_k is:

$$\tilde{x}_i = x_i - \frac{Cov(x_i, x_k)}{Cov(x_k, x_k)} x_k, \tag{5}$$

where $Cov(x_i, x_k)$ is the covariance of x_i and x_k. However, it should be noted that this way may be infeasible, because x_i, x_j, and x_k might not exist in the same client. For example, the variable sets in three clients A, B, and C are $\{x_i, x_j\}$, $\{x_i, x_k\}$, $\{x_j, x_k\}$, respectively. Since the data from these three clients cannot be shared, it is not possible to eliminate the influence of x_k on x_i and x_j in client A using the method described in Eq. (5). Moreover, it also requires estimating parameter information from the client side, which increases computational complexity and heightens the risk of information leakage. Fortunately, leveraging the calculated (cross) cumulants, we can obtain the total causal effect from x_k to x_i as $a_{ik} = \frac{C_{2,1}(x_i,x_k)}{C_3(x_k)}$. Moreover, we can eliminate the influence of x_k on x_i by only using the cumulants, which is summarized in the following lemma.

Lemma 2 *Assumed that observed variables X are generated by Eq. (1). Then, if x_s is the identified source variable, then for all descendant nodes of x_s with respect to x_i and x_j, the following equations hold:*

$$C_3(\tilde{x}_i) = C_3(x_i) - a_{is}^3 C_3(x_s), \quad (6)$$

$$C_3(\tilde{x}_j) = C_3(x_j) - a_{js}^3 C_3(x_s), \quad (7)$$

$$C_{1,2}(\tilde{x}_i, \tilde{x}_j) = C_{1,2}(x_i, x_j) - a_{is} a_{js}^2 C_3(x_s), \quad (8)$$

$$C_{2,1}(\tilde{x}_i, \tilde{x}_j) = C_{2,1}(x_i, x_j) - a_{is}^2 a_{js} C_3(x_s), \quad (9)$$

where $a_{is} = \frac{C_{2,1}(x_i,x_s)}{C_3(x_s)}$, and $a_{js} = \frac{C_{2,1}(x_j,x_s)}{C_3(x_s)}$.

Lemma 2 guarantees that by combining Equations (6) - (9), the (cross) cumulants used to determine the causal order between variables can also be employed to eliminate the influence of ancestors on the variables. This can be achieved without the need to obtain the variable data from the client side, thereby significantly enhancing privacy and reducing computational complexity.

FedISHC Method

In this section, we propose the FedISHC method for horizontal and vertical federated causal structure learning. The framework of the FedISHC method is illustrated in Figure 2.

The FedISHC method contains three main steps. First, it *estimates the local cross-cumulant matrix* in every client. After obtaining the matrix, all clients send their local cumulant matrix and the sample size to the server. Second, it constructs the global cross-cumulant matrix in the server, by using the local cumulant matrices and the sample size from all clients. Based on the global cross-cumulant matrix, it *recovers the source variables*. Third, it eliminates the influence of the source variables on their dependent variables and recovers the causal order recursively. Based on the identified causal order, it *learns the causal structure in the horizontal and vertical federated scenarios*. The FedISCH method is detailed in Algorithm 2.

Step 1: Cross-Cumulants Matrix Estimation

To obtain the information from different client, in each client, we first estimate the (cross) cumulants of variables

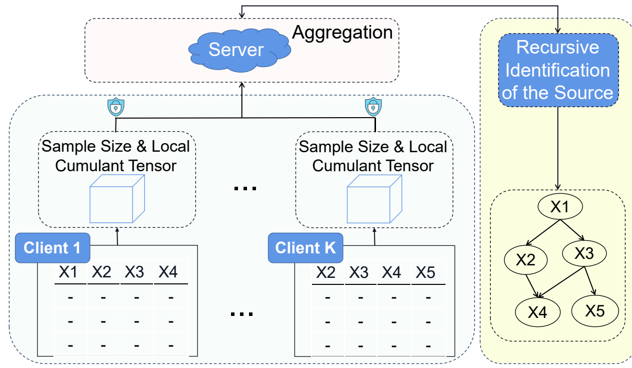


Figure 2: Overall framework of FedISHC.

and construct a cross-cumulants matrix. We denote $\mathbf{M}^i$ as the cross-cumulants matrix for client i, which is defined as:

$$\mathbf{M}^i = \begin{pmatrix} C_3(x_1) & \cdots & C_{1,2}(x_1, x_{d^i}) \\ \vdots & \ddots & \vdots \\ C_{1,2}(x_{d^i}, x_1) & \cdots & C_3(x_{d^i}) \end{pmatrix}. \quad (10)$$

Note that in horizontal and vertical federated scenarios, the sample and the variable set in different clients are different. Like other statistical measures that describe data, the (cross) cumulants in the total sample are equal to the weighted sum of the (cross) cumulants between the same variables after decomposing the total sample into multiple subsamples. That is, the value at position i, j in the global cumulant matrix $\mathbf{M}$ can be calculated by the following equation:

$$\mathbf{M}_{ij} = \frac{n_1}{\sum_{p=1}^{k}} \mathbf{M}_{ij}^1 + \cdots + \frac{n_k}{\sum_{p=1}^{k} n_p} \mathbf{M}_{ij}^k, \quad (11)$$

where n_p is the sample size in client p.

Step 2: Source Variables Recovery

After obtaining the global cross-cumulants matrix, we can use Theorem 1 and Lemma 2 to identify the causal order among observed variables. But there remains a problem: how to recover the source variables? Utilizing the linear additive property of the cumulant, we can recover the source variables by the following theorem.

Theorem 3 *Assume that the variables X are generated by LiNGAM. Then $x_s \in X$ is the source variable, if and only if*

$$\sum_{s \in I,\ j \in I \setminus \{s\}} \tau_{sj} = 0,$$

where $I = \bigcup_{k=1}^{K} I^k$, I^k is the index set of the observed variable set X^k.

Theorem 3 guarantees that if for x_s, $\sum_{s \in I,\ j \in I \setminus \{s\}} \tau_{sj} = 0$ for other variables, then x_s is the source variable. When there are multiple source variables, they are called sibling source variables. The relationship between variables of the same level must be independent. This can be identified by Theorem 1.

Algorithm 1: ISHC algorithm

Input: $\mathbf{M}, I = \{1, 2, \ldots, d\}, \mathbf{A}$
Output: $CO, \mathbf{A}$

1: $CO \leftarrow \{\}$.
2: **while** $\#CO < d - 1$ **do**
3: **for** $i \in I \backslash CO$ **do**
4: **for** $j \in I \backslash CO$ and $j \neq i$ **do**
5: Compute τ_{ij}.
6: **end for**
7: $T_\tau(x_i; I \backslash CO) \leftarrow \sum \tau_{ij}$.
8: **end for**
9: **if** $T_\tau(x_s; I \backslash CO) = 0$ **then**
10: $CO \leftarrow CO \cup \{s\}$.
11: **end if**
12: Update the elements in $\mathbf{M}$ according to Lemma 2.
13: **end while**

Step 3: Federated Causal Structure Learning

First, after identifying the current source variable through Theorem 3, we need to eliminate the influence of the source variable on its descendant nodes. We use equations (6) - (9) to update the global cumulant matrix in order to eliminate the influence of the discovered source variables at the cumulant level. Therefore, we propose Theorem 3 to ensure the correctness of our approach of eliminating the influence of the discovered source variables at the cumulant level.

Algorithm 1 demonstrates Step 2 and Step 3 of our method. Lines 3 - 11 of Algorithm 1 aim to identify the source variables, while lines 12 of Algorithm 1 aims to eliminate the influence of the identified source variables at the cumulant level. Line 12 uses Equations (6) - (9) to eliminate the influence of the identified source variables. At this point, we need to avoid introducing additional noise that could affect two variables without a common ancestor. Therefore, when eliminating the influence of the source variable x_s, we only consider variables that are dependent on x_s and remove the influence of x_s on them.

Throughout the algorithm's execution, we compute the total causal effect $a_{is} = C_{2,1}(x_i, x_s)/C_3(x_s)$ for the source variable x_s and variable x_i. The value of a_{is} is recorded in the (s, i) position of the mixing matrix $\mathbf{A}$. Consequently, our algorithm simultaneously estimates the total mixing matrix $\mathbf{A}$ while determining the causal order. Then, we can obtain the global causal strength matrix $\mathbf{B}$ by $\mathbf{B} = \mathbf{I} - \mathbf{A}^{-1}$.

In practice, τ_{ij} is not exactly zero, and its exact distribution remains unknown. Non-parametric methods like the sign test or permutation test can be utilized to assess whether τ_{ij} can be considered as zero. Thus, instead of transmitting the local cumulant matrix, the client sends the local cumulant tensor to the server.

Privacy and Cost Analysis

Privacy Analysis

Initially, local cross-cumulant matrix $\mathbf{M}^k$ are learned in the respective federated clients, which require the raw data of each client, and then are shared to the server for obtaining the global cumulant matrix $\mathbf{M}$. Next, the server utilizes the global cumulant matrix as the input for Algorithm 1, which is then executed recursively. Upon completion, Algorithm 1 produces the global causal order, the updated global cumulant matrix, and the mixing matrix $\mathbf{A}$. As a result, FedISHC exchanges higher-order statistics information rather than raw data, protecting data privacy to a certain extent. To further avoid data privacy leakage, secure multi-party computation can be implemented using Shamir's secret sharing scheme (Lindell 2020). In this approach, each client securely distributes secret shares of its local cumulant matrix to other participants. The clients then collaboratively reconstruct the aggregated matrix through distributed computation. Alternatively, homomorphic encryption techniques may be employed (Acar et al. 2018). Each client encrypts its local cumulant matrix using homomorphic encryption and transmits the encrypted data to the server. The server then performs aggregation operations directly on the homomorphically encrypted matrices. Subsequent decryption yields the aggregated result in plaintext.

Algorithm 2: FedISHC Algorithm

Input: $\{D^1, D^2, \ldots, D^k\}, I = \{1, 2, \ldots, d\}$
Output: Causal strength matrix $\mathbf{B}$

for each client $k \in \{1, \ldots, K\}$ **do**
$\mathbf{M}^k \leftarrow GetCumMatrix(D^k)$.
$n_k \leftarrow$ sample size of client k.
end for
$\mathbf{M} \leftarrow$ in the server, aggregating all the local Cumulant matrices($\{\mathbf{M}^k\}_{k=1}^K$) into $\mathbf{M}$.
$A \leftarrow$ generate an all-zero matrix of size d.
$CO, \mathbf{A} \leftarrow$ ISHC($\mathbf{M}$, I, $\mathbf{A}$).
$\mathbf{B} = \mathbf{I} - \mathbf{A}^{-1}$.

Communication Cost

Let d^k denote the number of observed variables in client k and d denote the number of global variables in the server. In our method, the client and the server only need to communicate once. Let N denote the number of tests on whether τ_{ij} is zero. Then in practice, for each client k, it just send a $d^k \times d^k \times N \times 1$ tensor, and a value n_k, which requires $O(d^k \times d^k \times N \times 1 + 1)$. The total communication cost for K clients is $O(KN(d^k)^2 + K)$.

Experiments

In this section, we evaluate the proposed FedISHC method by conducting experiments on synthetic datasets and real-world data. All the experiments are run on a machine equipped with an Intel Core i7-10875H CPU. In each experiment, all methods were executed 10 times. In particular, we aim to answer the following research questions (RQs):

- RQ1: How does the proposed method against existing methods for handling small-scale graphs?
- RQ2: How does the proposed method perform on graphs with different dimensional variables?
- RQ3: How does the proposed method perform on graphs of different densities?

- RQ4: How does the proposed method perform in different sample sizes for clients?
- RQ5: How does the proposed method perform in different numbers of clients?

Experiment Setting

Datasets. We use the following two types of datasets.

1) **Synthetic data.** First, we use the ER model to generate real causal graphs with different numbers of variables and different density ratios. In these causal graphs, the range of directed edge density s is $\{0.3, 0.5, 0.7, 0.9, 1\}$, and the range of the total number of observed variables d is $\{5, 10, 20, 30, 40, 50\}$. Then, We generated the data based on Eq. (1). The specific parameters are as follows: i) each causal strength is sampled from a uniform interval $[-0.8, -0.2] \cup [0.2, 0.8]$; ii) the noise term E is sampled from an Exponential Distribution with a shape parameter of 5. iii) the range of the sample size n_k for each client is $\{100, 500, 1000, 2000, 5000\}$. The ratio of the number of observed variables in the dataset of each client to the total number of variables of all clients is λ, i.e., $d^k = d * \lambda$. The range of the number of clients K is $\{3, 5, 6, 10, 15, 30\}$.

2) **Real-world data.** To assess the performance of the methods, we utilize the Sachs dataset (Sachs et al. 2005). The Sachs dataset is a classic dataset for causal discovery in the field of bioinformatics. It focuses on the protein signal transduction network within human immune system cells. Through high-throughput multi-parameter flow cytometry, it measured the expression levels of 11 phosphorylated proteins and phospholipids in individual cells.

Baselines. We have two types of settings.

1) **Horizontal and vertical settings.** We compare FedISHC and FedHC with CDMiNi (Huang et al. 2020). FedHC is a variant of our method FedISHC. It focuses solely on identifying the causal order without removing the influence of source variables. CDMiNi is the only method for identifying causal structures directly from multiple datasets with non-identical variable sets, without considering the federated setting. Under this setting, λ is 0.8.

2) **Horizontal setting.** In the horizontal federated setting, we use horizontal federated learning methods as baseline, including FedC^{2}SL (Wang, Ma, and Wang 2023) and FedCDH (Li et al. 2024). FedC^{2}SL has two methods. The first one is FedPC, and the second one is PC-CIT-Voting, which is a variant of FedPC. Finally, since in CPDAGs, some orientations are not determined, we did the following steps: randomly assign directions to the undirected edges in the CPDAG. Under this setting, λ is defaulted to 1.

Metrics. To evaluate the performance of estimated causal graphs by different methods, we use F1 Score as the evaluation metric, which is calculated as follows:

$$F1\ Score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall},$$

where $Precision$ denotes the percentage of correctly identified edges among all edges in the estimated graph, and $Recall$ denotes the percentage of edges in the ground-truth graph correctly recovered by the estimated graph.

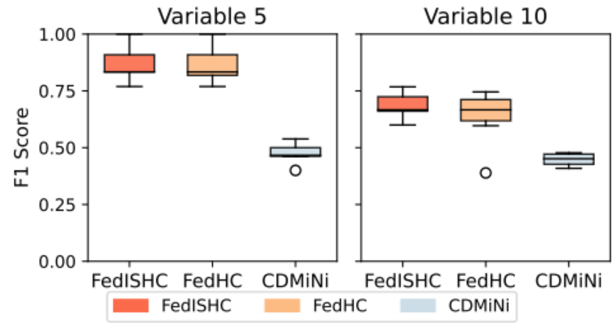


Figure 3: F1 score obtained by our methods and CDMiNi method when the number of variables is 5 or 10.

Results on Synthetic Data

Sensitivity on small-scale graphs. We generated synthetic data according to causal graphs with 5 and 10 variables. The parameters of this experimental setup are as follows: $s = 0.5$, $n_k = 1000$, $K = 5$. The F1 score obtained by our method and baseline methods are shown in Figure 3. It can be observed that in all cases, our method outperforms CDMiNi, and the F1 score of our method is approximately 20% higher than that of CDMiNi. Because CDMiNi is severely affected by potential variables, resulting in poor performance. In contrast, FedISHC and FedHC, which perform collaborative learning by aggregating all local information from the client side, are better suited for accurately identifying causal relationships. More experimental results are presented in Appendix D.

Sensitivity on graphs with different dimensional variables. In this experiment, the setups are: $s = 1$, $n_k = 100$, $K = 10$. The results are given in Figure 4 (a). It is observed that with the increase in the number of variables, FedISHC and FedHC remain stable, while existing federated causal discovery methods tend to decline. Because the increase in variable dimensions leads to a more complex graph structure, which poses a significant challenge to the accuracy of current federated causal discovery methods. Moreover, FedISHC and FedHC utilize higher-order cumulants to capture the asymmetry between variables, which is useful for learning causal orders and causal graph, thereby improving the overall accuracy.

Sensitivity on graphs of different densities. The results of F1 Score are shown in Figure 4 (b), under the parameters of our experimental setup as: $d = 20$, $n_k = 500$, $K = 6$. We can observe that as the density of directed edges increases, the accuracy curves of FedISHC and FedHC rise, while the accuracy curve of the baseline method continues to decline. Because graphs with higher edge density have more complex triangular structures, our method performs better in learning such structures. However, FedCDH, FedPC and PC-CIT-Voting methods determine the causal directions between observed variables by the V-structure and some directional rules. The existing triangular structures in the ground truth make them unable to find the V-structure.

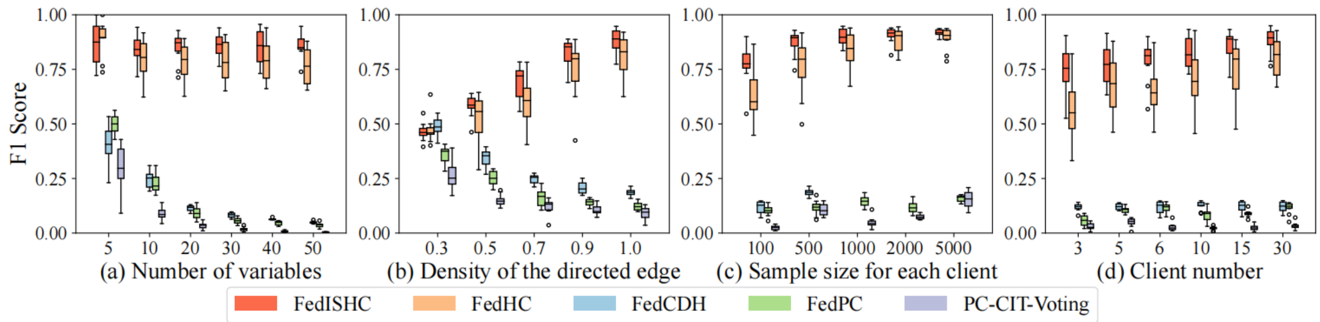


Figure 4: Comparison Results on Exponential Distribution Synthetic Data.

When the density of directed edges is 0.3, the performance of FedISHC and FedHC is slightly lower than that of FedCDH. Due to the estimation bias of cumulant, more redundant edges were not eliminated.

Sensitivity in different sample sizes for clients. The results of F1 Score are given in Figure 4 (c), with the parameters setup as $s = 1$, $d = 20$, $K = 6$. Due to the increase in sample size, the computational complexity of FedCDH has increased exponentially. Therefore, we only provided the results for 100 and 500. It can be observed that as the sample size increases, the performance of FedISHC, FedHC and FedCDH also improves. But the F1 scores of baseline methods are much low than those of our proposed methods. Because for the dense causal graph, utilizing conditional independence tests cannot remove many edges, which leads to difficulties in identifying subsequent v-structures and orienting other causal edges, ultimately reducing the overall effectiveness of FedPC and PC-CIT-Voting method.

Sensitivity in different numbers of clients. The results of F1 Score are shown in Figure 4 (d). The experimental setup is: $s = 1$, $d = 20$, $n_k = 100$. It can be observed that as the number of clients increases, the performance of all methods improves. This is because the volume of aggregated client information expands. Among these methods, FedISHC demonstrates the most significant improvement, followed by FedHC. When the number of clients is relatively small, FedCDH exhibits the greatest stability. Although FedPC shows an upward trend, it becomes unstable when the number of clients falls within the range of 5 to 15. PC-CIT-Voting is a voting-based method. As the number of clients increases, the consistency of the voting strategy may deteriorate and does not show an overall upward trend. Our method achieves collaborative learning by efficiently aggregating client-side information, addressing performance issues in data silo scenarios.

Results on Sachs Data

The results of the horizontal and vertical settings on the Sachs dataset are illustrated in Table 1. Our FedHC method achieved the best performance, while FedISHC achieved suboptimal performance. Compared with CDMiNi, our method improved the F1 score by 7%. This also indicates the advantage of our method in federated causal learning.

Algorithms	F1 Score
FedISHC	0.466 ± 0.019
FedHC	$\mathbf{0.471 \pm 0.013}$
CDMiNi	0.396 ± 0.028

Table 1: The experimental results under horizontal and vertical settings in the Sachs data.

Algorithms	F1-edge	F1-orie
FedISHC	0.601 ± 0.014	0.302 ± 0.026
FedHC	$\mathbf{0.610 \pm 0.013}$	$\mathbf{0.307 \pm 0.026}$
FedCDH	0.500 ± 0.000	0.221 ± 0.101
FedPC	0.451 ± 0.019	0.140 ± 0.061
PC-CIT-Voting	0.370 ± 0.028	0.171 ± 0.086

Table 2: The experimental results under horizontal setting in the Sachs data.

The results of the horizontal setting are shown in Table 2 on the Sachs dataset. FedHC performed the best, while FedISHC's performance was slightly inferior. Specifically, the FedISHC and FedHC methods outperformed FedCDH by 10% on the F1 edge, by 15% compared to FedPC, and by 23% compared to PC-CIT-Voting. Since the comparison methods are based on a constrained framework, comparing in the F1-orie is unfair, so their results were not provided. Overall, both FedISHC and FedHC demonstrated superior performance. These findings indicate that integrating information from all clients is more beneficial for learning causal structures on the Sachs data.

Conclusion

In the paper, we proposed a horizontal and vertical federated causal order identification criterion in a linear non-Gaussian causal model. Based on the criterion, we propose a novel method for horizontal and vertical federated causal structure learning based on high-order cumulants. Experimental results demonstrate that our approach outperforms existing methods on both synthetic and real-world datasets. In particular, our method performs better on complex graph structures (such as triangular structures). Future research will explore more complex scenarios, including cases in which the global causal graph involves latent variables.

Acknowledgments

This research was supported in part by the National Science and Technology Major Project (2021ZD0111502), the Natural Science Foundation of China (U24A20233, 62206064, 62476163), the National Science Fund for Excellent Young Scholars (62122022), the Guangdong Basic and Applied Basic Research Foundation (2025A1515010172, 2023B1515120020), the Guangzhou Basic and Applied Basic Research Foundation (2024A04J4384). KZ would like to acknowledge the support from NSF Award No. 2229881, AI Institute for Societal Decision Making (AI-SDM), the National Institutes of Health (NIH) under Contract R01HL159805, and grants from Quris AI, Florin Court Capital, and MBZUAI-WIS Joint Program, and the Al Deira Causal Education project.

References

Acar, A.; Aksu, H.; Uluagac, A. S.; and Conti, M. 2018. A survey on homomorphic encryption schemes: Theory and implementation. *ACM Computing Surveys (Csur)*, 51(4): 1–35.

Brillinger, D. R. 2001. *Time series: data analysis and theory*. SIAM.

Cai, R.; Huang, Z.; Chen, W.; Hao, Z.; and Zhang, K. 2023. Causal discovery with latent confounders based on higher-order cumulants. In *International conference on machine learning*, 3380–3407. PMLR.

Cai, R.; Wu, Y.; Huang, X.; Chen, W.; Fu, T. Z.; and Hao, Z. 2024. Granger causal representation learning for groups of time series. *Science China Information Sciences*, 67(5): 152103.

Chen, W.; Cai, R.; Zhang, K.; and Hao, Z. 2021. Causal discovery in linear non-gaussian acyclic model with multiple latent confounders. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7): 2816–2827.

Chen, W.; Huang, Z.; Cai, R.; Hao, Z.; and Zhang, K. 2024. Identification of causal structure with latent variables based on higher order cumulants. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 20353–20361.

Dai, H.; Spirtes, P.; and Zhang, K. 2022. Independence testing-based approach to causal discovery under measurement error and linear non-gaussian models. *Advances in Neural Information Processing Systems*, 35: 27524–27536.

Feller, W. 1991. *An introduction to probability theory and its applications, Volume 2*, volume 2. John Wiley & Sons.

Guo, X.; Yi, L.; Wu, X.; Yu, K.; and Wang, G. 2024a. Enhancing causal discovery in federated settings with limited local samples. In *International Workshop on Trustworthy Federated Learning*, 164–179. Springer.

Guo, X.; Yu, K.; Liu, L.; and Li, J. 2024b. FedCSL: a scalable and accurate approach to federated causal structure learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 12235–12243.

Guo, X.; Yu, K.; Wang, H.; Cui, L.; Yu, H.; and Li, X. 2024c. Sample quality heterogeneity-aware federated causal discovery through adaptive variable space selection. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 4071–4079.

Huang, B.; Zhang, K.; Gong, M.; and Glymour, C. 2020. Causal discovery from multiple data sets with non-identical variable sets. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 10153–10161.

Huang, J.; Guo, X.; Yu, K.; Cao, F.; and Liang, J. 2023. Towards privacy-aware causal structure learning in federated setting. *IEEE Transactions on Big Data*, 9(6): 1525–1535.

Huang, Z.; Cheng, D.; Li, J.; Liu, L.; Lu, G.; and Zhang, S. 2025. Interaction-Data-guided Conditional Instrumental Variables for Debiasing Recommender Systems. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*.

Li, L.; Ng, I.; Luo, G.; Huang, B.; Chen, G.; Liu, T.; Gu, B.; and Zhang, K. 2024. Federated causal discovery from heterogeneous data. In *International Conference on Learning Representations*.

Li, Z.; Cai, R.; Fu, T. Z.; Hao, Z.; and Zhang, K. 2023. Transferable time-series forecasting under causal conditional shift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4): 1932–1949.

Lindell, Y. 2020. Secure multiparty computation. *Communications of the ACM*, 64(1): 86–96.

Ng, D.; Lan, X.; Yao, M. M.-S.; Chan, W. P.; and Feng, M. 2021. Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets. *Quantitative Imaging in Medicine and Surgery*, 11(2): 852.

Robeva, E.; and Seby, J.-B. 2021. Multi-trek separation in linear structural equation models. *SIAM Journal on Applied Algebra and Geometry*, 5(2): 278–303.

Sachs, K.; Perez, O.; Pe'er, D.; Lauffenburger, D. A.; and Nolan, G. P. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721): 523–529.

Schkoda, D.; and Drton, M. 2025. Goodness-of-fit tests for linear non-Gaussian structural equation models. *Biometrika*.

Shimizu, S.; Hoyer, P. O.; Hyvärinen, A.; Kerminen, A.; and Jordan, M. 2006. A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(10).

Shimizu, S.; Inazumi, T.; Sogawa, Y.; Hyvarinen, A.; Kawahara, Y.; Washio, T.; Hoyer, P. O.; Bollen, K.; and Hoyer, P. 2011. DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model. *Journal of Machine Learning Research-JMLR*, 12(Apr): 1225–1248.

Spirtes, P.; Glymour, C. N.; and Scheines, R. 2000. *Causation, prediction, and search*. MIT press.

Wang, Y.; Fuyuan, C.; Yu, K.; and Liang, J. 2025. Federated Causal Structure Learning with Non-identical Variable Sets. In *Forty-second International Conference on Machine Learning*.

Wang, Y. S.; and Drton, M. 2020. High-dimensional causal discovery under non-Gaussianity. *Biometrika*, 107(1): 41–59.

Wang, Z.; Ma, P.; and Wang, S. 2023. Towards practical federated causal structure learning. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 351–367. Springer.

Yang, L.; Tan, B.; Zheng, V. W.; Chen, K.; and Yang, Q. 2020. Federated recommendation systems. In *Federated Learning: Privacy and Incentive*, 225–239. Springer.

Yu, K.; Rong, C.; Wang, H.; Cao, F.; and Liang, J. 2025. Federated local causal structure learning. *Science China Information Sciences*, 68(3): 132105.

Zhang, H.; Yan, C.; Xia, Y.; Guan, J.; and Zhou, S. 2022. Causal gene identification using non-linear regression-based independence tests. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 20(1): 185–195.