Hybrid Local Causal Discovery

Zhaolong Ling [1], **Honghui Peng** [1], **Yiwen Zhang** [1 *], **Debo Cheng** [2 *], **Xingyu Wu** [3], **Peng Zhou** [1] and **Kui Yu** [4]
[1]Anhui University
[2]University of South Australia
[3]Hong Kong Polytechnic University
[4]Hefei University of Technology
zlling@ahu.edu.cn, penghonghui@stu.ahu.edu.cn, zhangyiwen@ahu.edu.cn, debo.cheng@unisa.edu.au, xingy.wu@polyu.edu.hk, doodzhou@ahu.edu.cn, yukui@hfut.edu.cn

Abstract

Local causal discovery aims to identify and distinguish the direct causes and effects of a target variable from observational data. Due to the inherent incompleteness of local information, popular methods from global causal discovery often face new challenges in local causal discovery tasks, such as 1) erroneous symmetry constraint tests and the resulting cascading errors in constraint-based methods, and 2) confusion within score-based approaches caused by local spurious equivalence classes. To address the above issues, we propose a Hybrid Local Causal Discovery algorithm, called HLCD. Specifically, HLCD initially utilizes a constraint-based approach with the OR rule to obtain a candidate skeleton, which is subsequently refined using a score-based method to eliminate redundant structures. Furthermore, during the local causal orientation phase, HLCD distinguishes between V-structures and equivalence classes by comparing local structure scores between the two, thereby avoiding orientation interference caused by local equivalence class ambiguities. Comprehensive experiments on 14 benchmark Bayesian networks and two real datasets validate that the proposed algorithm outperforms the existing local causal discovery methods.

1 Introduction

Causal discovery has always been an important goal in many areas of scientific research [Huang *et al.*, 2019; Prosperi *et al.*, 2020]. It reveals the underlying causal mechanisms of data generation and contributes to solving decision-making problems in machine learning [Yu *et al.*, 2020; Chen *et al.*, 2023]. Learning a Bayesian network (BN) from observational data is a popular method for causal discovery [Cui *et al.*, 2020; Zhang *et al.*, 2021]. The structure of a BN takes the form of a directed acyclic graph (DAG), where nodes signify variables, and edges represent cause-effect relationships [Xie *et al.*, 2020; Spirtes *et al.*, 2000]. In recent years, many global causal discovery algorithms have been proposed, which aim to learn the entire causal network [Chickering *et al.*, 2004]. In general, learning a global causal network over a large number of variables is computationally intractable [Zeng *et al.*, 2021; Scutari *et al.*, 2019]. To reduce computational complexity, the local-to-global approach was introduced to limit the search space for causal networks [Tsamardinos *et al.*, 2006; Gao *et al.*, 2017]. Rather than exploring the entire network across all variables at once, these methods first identify the Markov blanket (MB) or the parent-child (PC) set of a target variable and gradually construct the DAG skeleton from these subsets [Yu *et al.*, 2023]. In many practical cases, however, focusing on the causal relationships around a specific variable can eliminate the need to build a global causal network [Wu *et al.*, 2023], increasing the importance of local causal discovery algorithms.

Local causal discovery aims to uncover the causal structure surrounding a specific variable. However, due to the unavailability of complete global information, many edge directions determined by relationships with distant variables[1] remain unidentified. As a result, most existing methods adopt a progressive learning approach to gradually acquire outer layer information, until the causal directions around the target variable are identified. Consequently, local causal discovery commonly employs the faster constraint-based methods [You *et al.*, 2023; Ling *et al.*, 2025c], as score-based methods exhibit higher time complexity and are not well-suited for this gradual information acquisition process [Ling *et al.*, 2022a].

Similar to global causal discovery methods, local causal discovery is susceptible to common issues associated with conditional independence (CI) testing, which can impact accuracy [Wu *et al.*, 2024]. One prominent concern is that CI testing cannot accurately determine the causal skeleton. As a result, many approaches employ symmetry tests to address this limitation [Wu *et al.*, 2022; Ling *et al.*, 2025b]. However, the prevailing AND rule[2] and OR rule[3] used in these symme-

*Corresponding author

[1]"Distant variables" refer to nodes that are located further away from the target variable along the causal paths.

[2]If X belongs to the PC of Y, but Y does not belong to the PC of X, then X should be removed from the PC of Y.

[3]If X belongs to the PC of Y, but Y does not belong to the PC of X, then Y should be added to the PC of X.

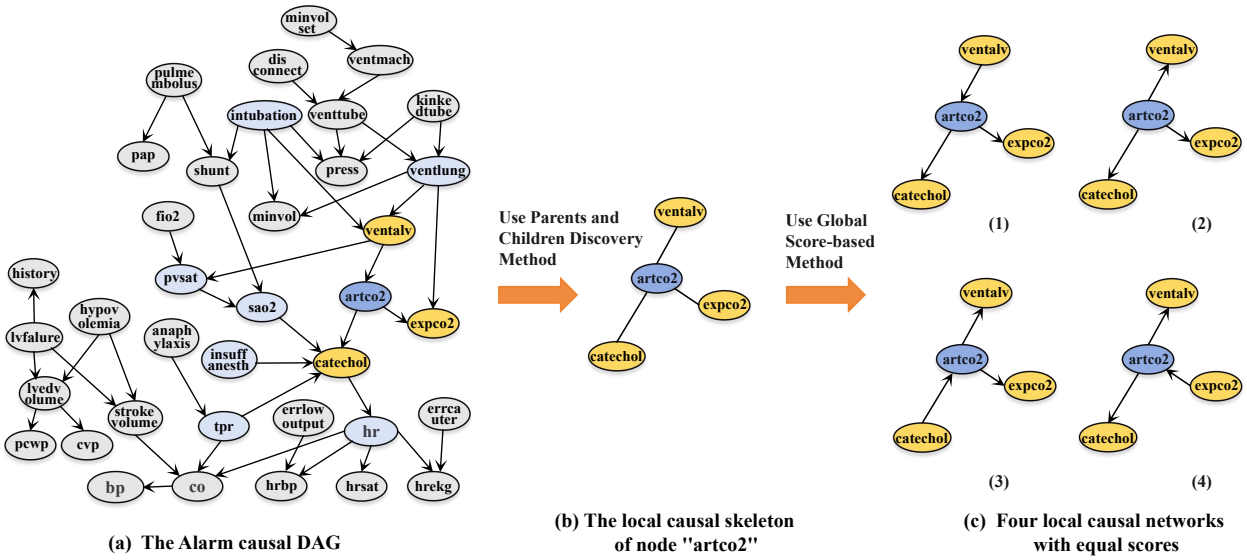

Figure 1: Directly using the search scoring algorithm to find the maximum score local network structure of node "artco2" will randomly return one of the four local structures in (c). It may depend on the order in which the variables in the dataset are encountered.

try tests introduce certain errors. The AND rule aims to rigorously eliminate all erroneous relationships, while the OR rule seeks to include as many true positives as possible, operating under a more lenient criterion [Guo *et al.*, 2024]. Empirical studies have provided evidence that approaches based on the AND rule achieve superior precision, whereas methods based on the OR rule exhibit better recall [Wu *et al.*, 2021; Guo *et al.*, 2023]. Consequently, neither approach yields completely accurate results. Additionally, the presence of data bias caused by inherent noise and incomplete local information further compounds the negative impact, exacerbating the potential for misleading results in local causal discovery.

To battle the challenge of insufficient observational samples in rare, costly, or privacy-sensitive events and the absence of global information due to unknown or unconsidered distant causal relationships in local causal discovery, a natural approach is to leverage a hybrid methodology for local causal discovery. This approach seeks to enhance performance by combining the strengths of constraint-based and score-based methods. While hybrid methods are commonly employed in global causal discovery research, their application in local causal discovery remains relatively unexplored. The complexity arises because a straightforward combination of these two methods inevitably faces the efficiency dilemma mentioned earlier in score-based approaches. Moreover, directly utilizing a global search scoring method to find the maximum score of local network structures may lead to incorrect local causal networks due to local equivalence class issues, as shown in Fig. 1. Furthermore, inaccuracies in CI tests resulting from information miss cascade into errors in the score-based causal discovery process. Consequently, effectively leveraging score information in local causal discovery poses a significant challenge.

In this paper, we introduce a novel hybrid method that identifies causal skeletons and V-structures by comparing scores among different local causal structures. Specifically, we employ a constraint-based approach for the initial causal skeleton, which uses symmetric tests with the OR rule to achieve a comprehensive yet less precise structure. On this basis, we demonstrate the identification and removal of redundant structures through specialized local structure scores between the target variable and its causes and effects. Additionally, we prove the discovery of V-structures using similar score information. Our main contributions are summarized as follows:

- We theoretically analyze the special local structure score relationships between the target variable and its causal variables, as well as different local structure score relationships between equivalence classes and V-structures.
- We propose a Hybrid Local Causal Discovery algorithm, HLCD. To the best of our knowledge, HLCD is the first work on hybrid local causal discovery. Based on our analysis, HLCD can effectively eliminate redundant causal skeletons and differentiate between V-structures and equivalence classes by scoring to avoid interference caused by score equivalence.
- We conducted extensive experiments against seven state-of-the-art local causal discovery algorithms on 14 benchmark BN datasets and two real datasets. The results show that our HLCD algorithm outperforms the compared methods, especially in the small sample case.

Section 2 reviews the related work. Section 3 describes the proposed HLCD algorithm in detail and Section 4 reports the experimental results. Section 5 summarizes the paper.

2 Related Work

Most local causal discovery algorithms are constraint-based and rely on CI tests to build and orient causal networks. Notable early approaches, such as Local Causal Discovery (LCD) [Cooper, 1997] and its variants, use CI tests to discover

causal relationships between sets of four variables. Bayesian Local Causal Discovery (BLCD) focuses on learning the Y-structure within the MB of a target variable [Mani and Cooper, 2004]. However, these LCD/BLCD algorithms aim to identify only a subset of causal edges, focusing on specific structural patterns among variables, without distinguishing the direct causal relationships for the target variable.

To address this problem, the state-of-the-art local causal discovery algorithms distinguish the direct causes and effects of the target variable directly. PCD-by-PCD (PCD means parents, children, and some descendants) [Yin *et al.*, 2008] uses the Max-Min Parents and Children (MMPC) algorithm [Tsamardinos *et al.*, 2003] to find PC and separating sets for V-structure identification, and then applies AND rule for local causal skeleton construction. Finally, the identified V-structures and Meek-rules [Meek, 1995] are applied to orient the edges in the local causal skeleton. MB-by-MB [Wang *et al.*, 2014] first finds a MB of the target node and constructs a local causal structure, and then sequentially finds MB of variables connected to the target and simultaneously constructs local structures along the paths starting from the target until the causes and effects of the target have been determined. Causal Markov Blanket (CMB) [Gao and Ji, 2015] initially applies the HITON-Markov Blanket (HITON-MB) algorithm [Aliferis *et al.*, 2003] to identify the target's MB, and then orients the edges by monitoring changes in conditional independence within the MBs. The Local Causal Structure learning by Feature Selection algorithm (LCS-FS) [Ling *et al.*, 2020] uses the mutual information-based feature selection method [Peng *et al.*, 2005] to discover the PC set of variables and construct the skeleton using OR rules, and then searches for separating sets from the learned PC sets and in turn uses the separating sets for edge orientation. Yang et al. [Yang *et al.*, 2021] proposed the concept of N-structures. By using N-structures, the Efficient Local Causal Structure Learning (ELCS) algorithm [Yang *et al.*, 2021] uncovers the local structure of the target variable while minimizing the number of MBs learned, thus reducing the number and influence of unreliable CI tests. The Partial Structure Learning (PSL) algorithm [Ling *et al.*, 2022b] is a partial causal discovery algorithm. It uses the OR rule to build the skeleton and finds two types of V-structures, Type-C and Type-NC, in the PC set of the current node, avoiding the false edge orientation problem of local causal discovery algorithms. Recently, Yang et al. proposed GraN-LCS (Gradient-based Local Causal Structure Learning) [Liang *et al.*, 2024], a gradient-based approach for learning local causal structures. This method builds a multilayer perceptron (MLP) to simultaneously model the relationships of all other variables with a target variable, and incorporates an acyclicity-constrained local recovery loss to encourage the discovery of local graphs and identify direct causes and effects.

3 The Proposed Method

This section presents our approach, including theoretical analysis and algorithmic details. For detailed proofs of the theorems and the analysis of the algorithm's time complexity, please refer to [Ling *et al.*, 2025a].

3.1 The local causal discovery strategy

In this section, we introduce the hybrid local causal discovery strategy for HLCD. This strategy is constructed based on the following two fundamental theorems.

Before introducing and proving the theorem, we need to account for the symbolic representation. Since the AIC score function (denoted as $\mathcal{S}_{\mathcal{A}}(\mathcal{G}, \mathcal{D})$) and BDue score function (denoted as $\mathcal{S}_{\mathcal{B}}(\mathcal{G}, \mathcal{D})$) are decomposable, they can be written as a sum of metrics, each of which is a function of only one node and its parent node (i.e. $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(\mathcal{G}, \mathcal{D}) = \sum_{i=1}^{n} \mathcal{S}_{\mathcal{A}/\mathcal{B}}(X_i, Pa_i^{\mathcal{G}})$, where $Pa_i^{\mathcal{G}}$ denotes the parent of X_i in $\mathcal{G}$). Therefore, we use the symbols $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to X, \mathcal{D})$ and $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(X \to Y, \mathcal{D})$ to denote the AIC score or BDue score of node X and Y, respectively. Where $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to X, \mathcal{D})$ denotes the score of node X when the empty set is the parent of X, and $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(X \to Y, \mathcal{D})$ denotes the score of node Y when X is the parent of Y. Moreover, $\mathbf{U}$ denotes the set of variables in the dataset, and $\mathbf{PC}_T$ refers to the parents and children nodes of the target variable T.

Theorem 1. Let T be any variable in $\mathbf{U}$, and X be a variable in $\mathbf{PC}_T$. Assume that the score function maintains local score consistency within the data $\mathcal{D}$. When node X is treated as a parent of T, in the local structure $X \to T$, the score of node T will increase. Conversely, when node T is treated as a parent of X, in the local structure $T \to X$, the score of node X will increase. Moreover, the score gains in both cases are identical. i.e. $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(X \to T, \mathcal{D}) - \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to T, \mathcal{D}) = \mathcal{S}_{\mathcal{A}/\mathcal{B}}(T \to X, \mathcal{D}) - \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to X, \mathcal{D}) > 0$ holds.

Theorem 1 shows that when X is a causal node of T, the local score relationship between X and T will always satisfy $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(X \to T, \mathcal{D}) - \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to T, \mathcal{D}) = \mathcal{S}_{\mathcal{A}/\mathcal{B}}(T \to X, \mathcal{D}) - \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to X, \mathcal{D}) > 0$. Therefore, using Theorem 1, we can first employ existing parent and child discovery algorithms and the OR rule to construct a comprehensive but redundant causal skeleton. Then, the nodes in the skeleton that do not satisfy Theorem 1 are deleted, thus removing the redundant skeleton structure and providing a more precise causal structure search space for the subsequent causal orientation.

Since all structures within the equivalence class share the same score, we consider $X \to T \to Y$ as the representative structure of the equivalence class, and $X \to T \leftarrow Y$ as the V-structure.

Theorem 2. Let $X, Y, T \in \mathbf{U}$ and T be a target node with no edge connected between X and Y, and $X, Y \in \mathbf{PC}_T$. Assume that the score function maintains score consistency within the data $\mathcal{D}$. Then, when the score of local structures $X \to T \leftarrow Y$ is greater than the score of local structures $X \to T \to Y$, i.e., $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to X, \mathcal{D}) + \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to Y, \mathcal{D}) + \mathcal{S}_{\mathcal{A}/\mathcal{B}}(X, Y \to T, \mathcal{D}) > \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to X, \mathcal{D}) + \mathcal{S}_{\mathcal{A}/\mathcal{B}}(X \to T, \mathcal{D}) + \mathcal{S}_{\mathcal{A}/\mathcal{B}}(T \to Y, \mathcal{D})$, there exists a V-structure in variables X, Y, T, and T is a collision node.

With Theorem 2, we can identify V-structures through score comparison, thereby avoiding interference caused by equivalence classes, and determine the final local causal directions by incorporating the Meek rule.

Algorithm 1 Hybrid Local Causal Discovery

Input: $\mathcal{D}$: Data, T: The target variable
Output: Parents of T: Direct causes of T, Children of T: Direct effects of T
Initialize: $\mathbf{V} = \emptyset$, Q (a regular queue) = $\{T\}$
repeat
  /* Step 1: Hybrid local causal skeleton construction */
  $Z = Q.pop$;
  if $Z \notin$ ***V*** **then**
    $\mathbf{PC}_Z$ = getPC($\mathcal{D}$,Z);
    $\mathbf{V} = \mathbf{V} \cup \{Z\}$;
  end
  for each $X \in \mathbf{PC_Z}$ **do**
    if The local score of $X \to Z$ is not equal to $Z \to X$ or $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(X \to Z, \mathcal{D}) - \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to Z, \mathcal{D}) < 0$ **then**
      $\mathbf{PC}_Z = \mathbf{PC}_Z \backslash \{X\}$;
    end
  end
  $Q = Q.push(\mathbf{PC}_Z \backslash \{\mathbf{V}\})$;
  /* Step 2: Hybrid local causal orientation */
  for each X,$Y \in \mathbf{PC}_Z$ **do**
    if The local score of $X \to Z \leftarrow Y$ is greater than $X \to Z \to Y$ **then**
      The X, Y, Z form a V-structure, and Z is the collision node;
    end
  end
  Using Meek-rules to orient edge orientations between variables in $\mathbf{V}$;
until All causal orientations of T is determined, or $Q = \emptyset$, or $\mathbf{V}$ contains all variables;
Return Parents of T, Children of T;

3.2 Detailed descriptions of the HLCD algorithm

In this section, we describe the details of the HLCD algorithm implementation. It consists of the following two steps.

Step 1. Hybrid local causal skeleton construction: The HLCD begins by removing a variable from the front of the queue Q and assigning it as the current iteration node Z (starting with Z as the given target node T). It then applies the constraint-based parent and child discovery algorithm to determine the $\mathbf{PC}_Z$ and constructs a local causal skeleton using the OR rule. The HLCD can use any of the state-of-the-art parent and child discovery algorithms, such as MMPC, FCBF, etc. Then, the HLCD stores Z into $\mathbf{V}$ to prevent repeated learning of the PC of variables. At this point, the HLCD builds an initial local causal skeleton from the OR rule and learned PC sets.

As the OR rule can generate a comprehensive but potentially redundant causal skeleton, the HLCD incorporates the score-based method to eliminate redundant causal skeletons, ensuring they don't interfere with subsequent causal orientations. With the analysis of Theorem 1, if node $X \in \mathbf{PC}_Z$, then the following equation will hold: $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(X \to Z, \mathcal{D}) - \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to Z, \mathcal{D}) = \mathcal{S}_{\mathcal{A}/\mathcal{B}}(Z \to X, \mathcal{D}) - \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to X, \mathcal{D}) > 0$. The HLCD does this by testing each variable X in $\mathbf{PC}_Z$ to see if it satisfies Theorem 1, and removing it from $\mathbf{PC}_Z$ if it does not satisfy. Next, the HLCD adds all variables in $\mathbf{PC}_Z \backslash \{\mathbf{V}\}$ to the queue Q, allowing it to recursively determine the PC of each node in $\mathbf{PC}_Z$ in subsequent iterations for further expansion. At the end of step 1, the HLCD obtains a refined local causal skeleton consisting of all nodes in the set $\mathbf{V}$ and their PC nodes.

Step 2. Hybrid local causal orientation: To avoid the effect of the score equivalence, the HLCD distinguishes between V-structures and equivalence class structures by employing the score-based method. Specifically, the HLCD identifies the V-structures in the causal skeleton by comparing the two local structure scores of each tuple X, Y and Z ($X, Y \in \mathbf{PC}_Z$) in the causal skeleton obtained in step 1. If $\mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to X, \mathcal{D}) + \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to Y, \mathcal{D}) + \mathcal{S}_{\mathcal{A}/\mathcal{B}}(X, Y \to Z, \mathcal{D}) > \mathcal{S}_{\mathcal{A}/\mathcal{B}}(\emptyset \to X, \mathcal{D}) + \mathcal{S}_{\mathcal{A}/\mathcal{B}}(X \to Z, \mathcal{D}) + \mathcal{S}_{\mathcal{A}/\mathcal{B}}(Z \to Y, \mathcal{D})$, then the edge $X - Z$ and edge $Y - Z$ will be oriented as $X \to Z$ and $Y \to Z$. At this point, the HLCD orients the causal orientations of all V-structures in the current causal skeleton and does not orient the causal orientations of equivalent class structures.

Finally, The HLCD uses the constraint-based Meek-rule[4] as well as the discovered V-structure to orient the causal orientations of the nodes in the set $\mathbf{V}$. If all causal orientations of T are recognized in the current $\mathbf{V}$, learning stops, otherwise it continues to expand outward until it recognizes between the causes and effects of T. If the set $\mathbf{V}$ includes all variables, and there are still nodes in $\mathbf{PC}_T$ that have not been directed as parents or children, then these nodes are considered undirected. In this case, the HLCD also outputs the undirected causal nodes. That is, if there are undirected causal orientations, the HLCD outputs the local completed partially directed acyclic graph (CPDAG) of T.

Theorem 3 (Correctness of HLCD). Given a set of i.i.d data $\mathcal{D}$, and samples from some distribution $\mathcal{P}$. As the size of $\mathcal{D}$ goes to infinity, HLCD correctly identifies all the causes and effects of a given variable.

Theorem 3 guarantees the correctness of Algorithm 1 under the given assumptions, showing that HLCD can accurately recover the target variable's causal structure without ambiguity from equivalence classes.

4 Experiments

We conducted experiments on 14 benchmark BN datasets, where each BN dataset generated samples of sizes 500 and 1000, respectively. Furthermore, we used two real datasets. The first was a well-known dataset from [Sachs *et al.*, 2005], which captures the varying expression levels of proteins and phospholipids in human cells. This dataset's ground truth causal graph consists of 11 nodes and 17 edges, and we tested it with observational data comprising 853 samples. The second dataset was a pseudo-real dataset generated using the SynTReN generator [Van den Bulcke *et al.*, 2006], which simulates synthetic transcriptional regulatory networks to approximate experimental gene expression data. We generated a dataset with 20 nodes and 500 samples, using the default parameters. Finally, we compare our approaches to HLCD

[4]The Meek-rule consist of three main principles that, when combined with collider information, allow for the orientation of the remaining edges without introducing directed cycles.

Metrics	Algorithm	Alarm	Child	Barley	Hailfinder3	Link	Pigs	Gene
F1	GraN-LCS	0.37±0.03	0.30±0.04	0.20±0.02	0.10±0.01	-	0.43±0.01	-
	HLCD-M	**0.64±0.05**	**0.57±0.04**	**0.30±0.02**	**0.36±0.02**	**0.24±0.01**	**0.98±0.01**	**0.84±0.01**
	LCS-FS	0.44±0.05	0.30±0.20	0.24±0.02	0.32±0.02	0.18±0.01	0.92±0.01	0.91±0.01
	HLCD-FS	**0.58±0.02**	**0.68±0.11**	**0.29±0.05**	**0.43±0.04**	**0.20±0.02**	**0.96±0.01**	**0.94±0.01**
	ELCS	0.44±0.04	0.53±0.10	0.21±0.01	0.31±0.02	0.19±0.02	0.90±0.01	0.70±0.01
	HLCD-H	**0.64±0.05**	**0.66±0.19**	**0.28±0.02**	**0.37±0.02**	**0.24±0.01**	**0.98±0.00**	**0.83±0.01**
	PSL	0.50±0.07	0.56±0.10	0.19±0.02	0.27±0.02	0.17±0.01	0.94±0.01	0.85±0.01
	HLCD-P	**0.60±0.06**	**0.70±0.05**	**0.29±0.03**	**0.34±0.03**	**0.23±0.01**	**0.99±0.00**	**0.91±0.01**
SHD	GraN-LCS	2.57±0.24	2.41±0.16	5.39±0.21	9.40±0.44	-	1.81±0.05	-
	HLCD-M	**1.29±0.16**	**1.39±0.10**	**4.93±0.18**	**4.43±0.08**	**4.05±0.09**	**0.10±0.02**	**0.48±0.03**
	LCS-FS	1.75±0.11	1.85±0.53	4.33±0.10	3.02±0.06	4.29±0.29	0.47±0.08	0.30±0.04
	HLCD-FS	**1.43±0.11**	**0.99±0.23**	**3.05±0.15**	**2.89±0.14**	**4.09±0.25**	**0.26±0.06**	**0.25±0.05**
	ELCS	1.76±0.14	1.48±0.22	9.31±0.44	5.39±0.07	4.53±0.09	0.36±0.03	0.83±0.02
	HLCD-H	**1.27±0.15**	**1.20±0.46**	**5.14±0.23**	**4.43±0.07**	**4.08±0.10**	**0.10±0.02**	**0.53±0.02**
	PSL	1.62±0.18	1.43±0.26	11.52±0.69	5.67±0.12	4.41±0.08	0.20±0.03	0.41±0.03
	HLCD-P	**1.34±0.22**	**1.11±0.15**	**4.97±0.18**	**4.52±0.09**	**4.09±0.08**	**0.05±0.01**	**0.29±0.03**

Table 1: The experimental results of F1 and SHD for HLCD and its competitors on a partial BN dataset (7 out of 14 datasets) with a sample size of 500.

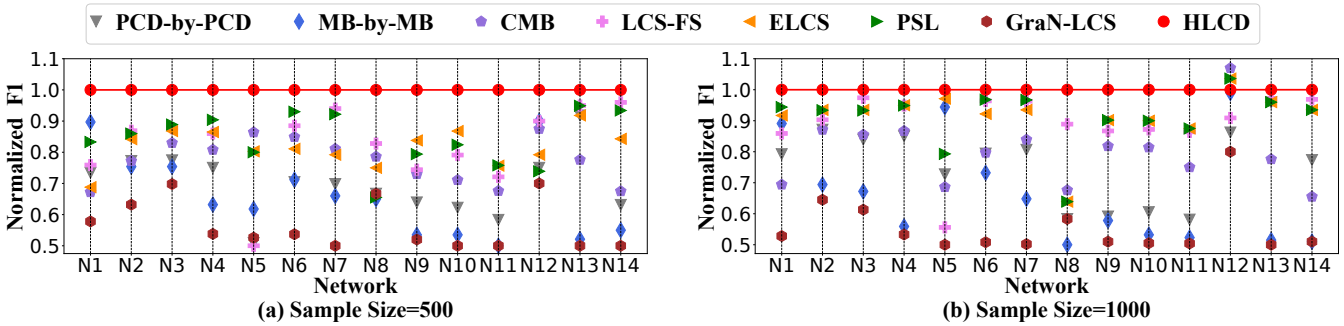

Figure 2: The experimental results of normalized F1, where the normalized value is the result of the comparison algorithm divided by the result of the HLCD. The larger the normalized F1, the better (the x-axis labels from N1 to N14 represent the Bayesian networks. N1: Alarm. N2: Alarm3. N3: Alarm5. N4: Alarm10. N5: Child. N6: Insurance3. N7: Insurance5. N8: Barley. N9: Hailfinder3. N10: Hailfinder5. N11: Hailfinder10. N12: Link. N13: Pigs. N14: Gene).

with seven state-of-the-art local causal discovery algorithms [Ling *et al.*, 2022c], including PCD-by-PCD [Yin *et al.*, 2008], MB-by-MB [Wang *et al.*, 2014], CMB [Gao and Ji, 2015], LCS-FS [Ling *et al.*, 2020], ELCS [Yang *et al.*, 2021], PSL [Ling *et al.*, 2022b],and GraN-LCS [Liang *et al.*, 2024].

In the evaluation of the quality of local causal graph learning, we utilized two metrics: F1 and SHD [Tsamardinos *et al.*, 2006]. The F1 is the harmonic mean of precision and recall, weighted accordingly. The SHD represents the Structural Hamming Distance. A higher F1 score indicates better performance, while lower SHD values is preferable.

Different local discovery methods adopt various parent and child discovery algorithms to identify PC set when constructing the skeleton structure. To eliminate discrepancies arising from these differences, we ensure that HLCD and the comparison methods employ the same parent-child identification approach and conduct separate comparisons of their experimental results.

Furthermore, for more details on the experimental setup, more comprehensive experimental data, and additional experimental results, please refer to [Ling *et al.*, 2025a].

4.1 Synthetic data experiment

Table 1 shows the results of the F1 and SHD experiments of HLCD with state-of-the-art local causal discovery algorithms over the past four years on a partial BN dataset (7 out of 14 datasets) with a sample size of 500. The Figs. 2 and 3 present the normalized F1 and SHD results of HLCD and its competitors across 14 BN datasets. Specifically, among 28 BN sample datasets, HLCD achieves the highest F1 score in 27 datasets and the lowest SHD value in 24 datasets. Compared to PCD-by-PCD and GraN-LCS, HLCD-M (using MMPC as the PC learning algorithm) achieves an 8% to 27% improvement in F1 scores and a 5% to 20% reduction in SHD values on networks such as Alarm, Child, Insurance, Hailfinder, and Gene. Compared to MB-by-MB and LCS-FS, HLCD-FS (using FCBF as the PC learning algorithm) achieves a 3% to 24% improvement in F1 scores and a 2% to 28% reduction in SHD values on Alarm, Insurance, Barley, Hailfinder, and Pigs networks. Compared to CMB, ELCS, and PSL,

Metrics	Algorithm	Size=500	Size=1000	Size=5000	Size=10000	Size=15000	Size=20000
F1	GraN-LCS	0.20±0.02	0.21±0.03	0.22±0.01	0.21±0.01	0.23±0.01	0.26±0.01
	HLCD-M	**0.30±0.02**	**0.36±0.02**	**0.51±0.03**	**0.52±0.01**	**0.52±0.01**	**0.55±0.00**
	LCS-FS	0.24±0.02	0.32±0.02	0.42±0.03	0.44±0.02	0.44±0.02	0.44±0.01
	HLCD-FS	**0.29±0.05**	**0.36±0.02**	**0.46±0.04**	**0.49±0.01**	**0.51±0.02**	**0.50±0.01**
	ELCS	0.21±0.01	0.22±0.01	0.39±0.02	0.38±0.02	0.38±0.01	0.43±0.03
	HLCD-H	**0.28±0.02**	**0.34±0.02**	**0.51±0.03**	**0.52±0.01**	**0.54±0.01**	**0.55±0.00**
	PSL	0.19±0.02	0.23±0.02	0.41±0.02	0.42±0.02	0.42±0.02	0.47±0.02
	HLCD-P	**0.29±0.03**	**0.36±0.03**	**0.53±0.01**	**0.54±0.00**	**0.54±0.00**	**0.57±0.00**
SHD	GraN-LCS	5.39±0.21	**5.22±0.23**	4.32±0.05	4.43±0.17	4.18±0.17	3.62±0.14
	HLCD-M	**4.93±0.18**	5.26±0.22	**3.57±0.14**	**4.04±0.09**	**3.78±0.10**	**3.33±0.07**
	LCS-FS	4.33±0.10	3.77±0.11	2.90±0.13	2.72±0.07	2.75±0.08	2.80±0.05
	HLCD-FS	**3.05±0.15**	**2.98±0.08**	**2.73±0.10**	**2.60±0.05**	**2.56±0.06**	**2.63±0.03**
	ELCS	9.31±0.44	9.79±0.26	4.81±0.10	5.31±0.13	4.92±0.05	4.35±0.10
	HLCD-H	**5.14±0.23**	**5.28±0.21**	**3.58±0.15**	**4.05±0.11**	**3.66±0.05**	**3.30±0.08**
	PSL	11.52±0.69	11.06±0.34	4.56±0.05	4.88±0.06	4.53±0.07	3.82±0.13
	HLCD-P	**4.97±0.18**	**5.23±0.24**	**3.49±0.09**	**3.89±0.07**	**3.56±0.09**	**3.09±0.04**

Table 2: Results of F1 and SHD experiments between HLCD and its competitors on Barley's network with different sample size dimensions (Sample size: 500 $\sim$ 20000)

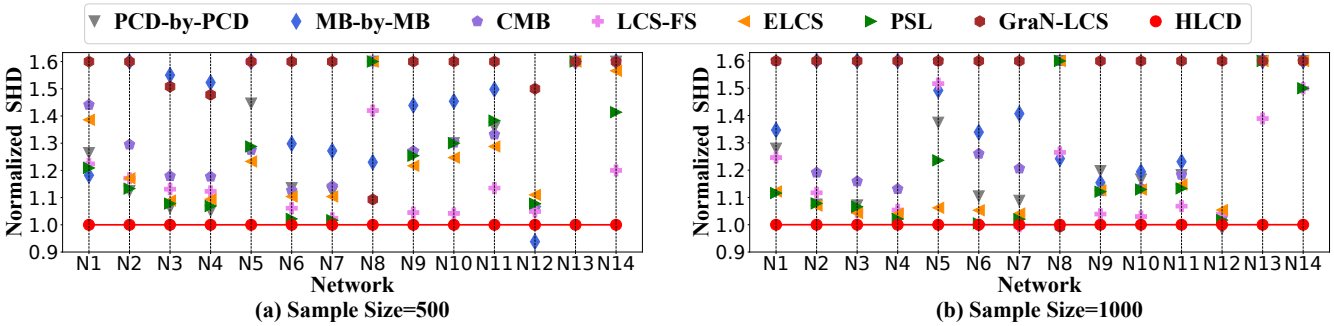

Figure 3: The experimental results of normalized SHD. The lower the normalized SHD, the better (the x-axis labels from N1 to N14 are identical to those in Figure 2).

HLCD-H (using HITON-PC as the PC learning algorithm) and HLCD-P (using PCsimple as the PC learning algorithm) achieve a 4% to 22% improvement in F1 scores and a 6% to 28% reduction in SHD values on all networks. In general, as the average number of conditional probability parameters Θ increases, the performance of local causal discovery declines due to the larger state space requiring more samples. Conversely, fewer parameters lead to better results. HLCD alleviates this issue through a scoring-based approach, achieving superior performance in both simple and complex networks.

4.2 Performance evaluation of HLCD with different sample sizes

To evaluate the effect of sample size on the algorithm, we assessed the performance of HLCD and its competitors (state-of-the-art algorithms from the last four years) on barley networks with sample sizes ranging from 500 to 20,000.

Table 2 summarizes the experimental results of HLCD and its competitors on the Barley network across sample sizes ranging from 500 to 20,000. As the sample size increases, the F1 score and SHD score of all algorithms generally improve. HLCD consistently outperforms other methods across most sample sizes. For example, compared to GraN-LCS, HLCD-M improves the F1 score by 8–11% and reduces the SHD by 8–10%. Against LCS-FS, HLCD-FS increases the F1 score by 4–6% and decreases the SHD by 4%–18%. When compared to ELCS, HLCD-H boosts the F1 score by 5%–10% and lowers the SHD by 13%–22%. Finally, compared to PSL, HLCD-P improves the F1 score by 10%–12% and reduces the SHD by about 20%.

Further analysis reveals that methods relying on CI tests or mutual information perform poorly with smaller sample sizes but improve significantly as the sample size grows, though they still lag behind HLCD. This is because HLCD leverages score information from data to enhance performance.

4.3 Real data experiment

Tables 3-4 summarize the experimental results of HLCD and its competitors on two real datasets. On the Sachs dataset, GraN-LCS achieved the highest F1 score of 42%, followed by HLCD-M/FS/H/P at 37%. For the SHD metric, HLCD-FS recorded the lowest value of 2.36. On the SynTReN dataset, HLCD-FS achieved the best F1 score of 25% and the second-lowest SHD of 2.20. Furthermore, certain algorithms fail to

Algorithm	F1	Precision	Recall	SHD	Time
PCD-by-PCD	0.13	0.15	0.13	3.09	**0.01**
GraN-LCS	**0.42**	0.45	**0.46**	**2.55**	75.30
HLCD-M	0.37	**0.55**	0.32	**2.55**	**0.01**
MB-by-MB	0.13	0.20	0.12	3.27	**0.01**
LCS-FS	0.00	0.00	0.00	3.09	**0.01**
HLCD-FS	**0.37**	**0.55**	**0.32**	**2.36**	**0.01**
CMB	0.00	0.00	0.00	3.36	**0.01**
ELCS	0.00	0.00	0.00	3.45	**0.01**
HLCD-H	**0.37**	**0.55**	**0.32**	**2.55**	**0.01**
PSL	0.00	0.00	0.00	3.45	**0.01**
HLCD-P	**0.37**	**0.55**	**0.32**	**2.55**	**0.01**

Table 3: Experimental results of HLCD and its competitors on Sachs dataset

Algorithm	F1	Precision	Recall	SHD	Time
PCD-by-PCD	**0.08**	**0.13**	0.07	**2.10**	**0.01**
GraN-LCS	0.01	0.02	0.01	3.40	3.46
HLCD-M	**0.08**	0.12	**0.10**	2.40	**0.01**
MB-by-MB	0.14	0.15	0.15	2.50	**0.01**
LCS-FS	0.00	0.00	0.00	2.90	**0.01**
HLCD-FS	**0.25**	**0.30**	**0.26**	**2.20**	**0.01**
CMB	0.03	0.03	0.03	2.70	**0.01**
ELCS	0.00	0.00	0.00	2.75	**0.01**
HLCD-H	**0.08**	**0.12**	**0.10**	**2.40**	**0.01**
PSL	0.00	0.00	0.00	2.75	**0.01**
HLCD-P	**0.08**	**0.12**	**0.10**	**2.40**	**0.01**

Table 4: Experimental results of HLCD and its competitors on SyNTReN dataset

recover correct local causal structures on the Sachs and SynTReN datasets, possibly because the learned PC or MB sets lack key collider nodes needed for accurate orientation.

To illustrate the detailed learning of local causal structures by the HLCD algorithm on the Sachs network, we present the experimental results in Fig. 4. The Sachs network contains 11 nodes, shown in dark blue in the innermost circle. Each branch extending from these nodes represents their parent or child nodes. Blue edges indicate that HLCD correctly identified a parent or child node, while red edges signify unsuccessful identification. From Fig. 4, it is evident that the HLCD algorithm performs well when a node's PC set is small but struggles as the PC set size increases. This decline in performance can be attributed to two factors: 1) Insufficient recall of the parent and child discovery algorithms: A smaller PC set may exclude correct causal nodes, causing V-structures (e.g., Raf and Mek) to be missed due to undetected parent nodes. 2) Nodes as colliders: Some nodes (e.g., PKA and PKC) are inherently collider nodes, lacking identifiable V-structures. In these cases, causal directions can only be inferred based on whether their child nodes are colliders.

Purely CI test or mutual information methods often fail to identify local causal networks in real-world datasets accurately. GraN-LCS iteratively refines the local causal graph using

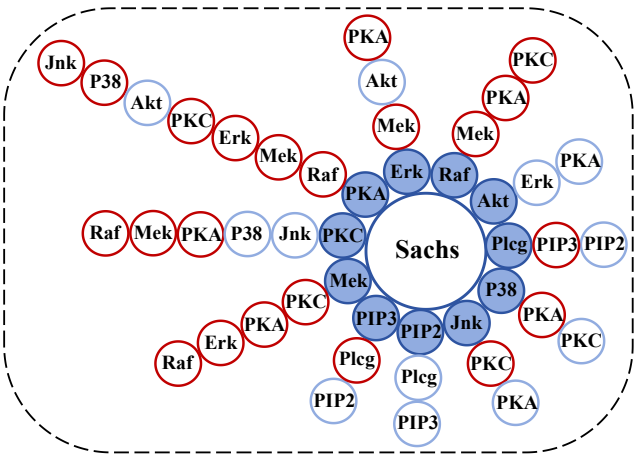

Figure 4: The identification results of the local causal structure for each node by the HLCD algorithm on the Sachs real network. Blue edges indicate that HLCD correctly identified a parent or child node, while red edges signify unsuccessful identification.

an MLP but suffers from low time efficiency due to extensive matrix computations. In contrast, HLCD integrates both constraint-based and score-based methods, striking an effective balance between accuracy and efficiency.

5 Conclusion

In this paper, we discuss the limitations of AND and OR rules in constructing exact local causal skeletons, and the problem of global causal discovery methods randomly returning incorrect local causal networks due to equivalence classes ambiguities. To address the challenges, we propose a novel hybrid local causal discovery (HLCD) method. Specifically, During the skeleton construction phase, HLCD uses maximized local scores to eliminate redundant causal skeleton structures, thereby providing a more precise causal network space. In the skeleton orientation phase, HLCD employs an innovative score-based V-structure identification approach to avoid interference caused by equivalence classes. The experimental results show that the quality of local causal discovery of HLCD is significantly better than existing methods. In future work, we aim to pursue two directions: 1) investigating the reasons behind HLCD's superior performance in small-sample settings, and 2) extending HLCD to dynamic or time-series data analysis, as well as exploring its potential in discovering more complex causal structures.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (under grant 2021ZD0111801), the National Natural Science Foundation of China (under grant 62306002, 62272001, 62376001, 62376087, and 62120106008), the Natural Science Foundation of Anhui Province of China under Grant 2108085QF270, and the Xunfei Zhiyuan Digital Transformation Innovation Research Special for Universities (2023ZY001).

References

[Aliferis *et al.*, 2003] Constantin F Aliferis, Ioannis Tsamardinos, and Alexander Statnikov. Hiton: a novel markov blanket algorithm for optimal variable selection. In *AMIA annual symposium proceedings*, volume 2003, pages 21–25. American Medical Informatics Association, 2003.

[Chen *et al.*, 2023] Zhengming Chen, Feng Xie, Jie Qiao, Zhifeng Hao, and Ruichu Cai. Some general identification results for linear latent hierarchical causal structure. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 3568–3576, 2023.

[Chickering *et al.*, 2004] Max Chickering, David Heckerman, and Chris Meek. Large-sample learning of bayesian networks is np-hard. *Journal of Machine Learning Research*, 5:1287–1330, 2004.

[Cooper, 1997] Gregory F Cooper. A simple constraint-based algorithm for efficiently mining observational databases for causal relationships. *Data Mining and Knowledge Discovery*, 1:203–224, 1997.

[Cui *et al.*, 2020] Peng Cui, Zheyan Shen, Sheng Li, Liuyi Yao, Yaliang Li, Zhixuan Chu, and Jing Gao. Causal inference meets machine learning. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3527–3528, 2020.

[Gao and Ji, 2015] Tian Gao and Qiang Ji. Local causal discovery of direct causes and effects. In *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2*, pages 2512–2520, 2015.

[Gao *et al.*, 2017] Tian Gao, Kshitij Fadnis, and Murray Campbell. Local-to-global bayesian network structure learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1193–1202, 2017.

[Guo *et al.*, 2023] Xianjie Guo, Kui Yu, Lin Liu, Peipei Li, and Jiuyong Li. Adaptive skeleton construction for accurate dag learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(10):10526–10539, 2023.

[Guo *et al.*, 2024] Xianjie Guo, Kui Yu, Lin Liu, Fuyuan Cao, and Jiuyong Li. Causal feature selection with dual correction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):938–951, 2024.

[Huang *et al.*, 2019] Biwei Huang, Kun Zhang, Mingming Gong, and Clark Glymour. Causal discovery and forecasting in nonstationary environments with state-space models. In *Proceedings of the 36th International Conference on Machine Learning*, pages 2901–2910, 2019.

[Liang *et al.*, 2024] Jiaxuan Liang, Jun Wang, Guoxian Yu, Carlotta Domeniconi, Xiangliang Zhang, and Maozu Guo. Gradient-based local causal structure learning. *IEEE transactions on cybernetics*, 54(1):486–495, 2024.

[Ling *et al.*, 2020] Zhaolong Ling, Kui Yu, Hao Wang, Lei Li, and Xindong Wu. Using feature selection for local causal structure learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(4):530–540, 2020.

[Ling *et al.*, 2022a] Zhaolong Ling, Ying Li, Yiwen Zhang, Kui Yu, Peng Zhou, Bo Li, and Xindong Wu. A light causal feature selection approach to high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering*, 35(8):7639–7650, 2022.

[Ling *et al.*, 2022b] Zhaolong Ling, Kui Yu, Lin Liu, Jiuyong Li, Yiwen Zhang, and Xindong Wu. Psl: An algorithm for partial bayesian network structure learning. *ACM Transactions on Knowledge Discovery from Data*, 16(5):1–25, 2022.

[Ling *et al.*, 2022c] Zhaolong Ling, Kui Yu, Yiwen Zhang, Lin Liu, and Jiuyong Li. Causal learner: A toolbox for causal structure and markov blanket learning. *Pattern Recognition Letters*, 163:92–95, 2022.

[Ling *et al.*, 2025a] Zhaolong Ling, Honghui Peng, Yiwen Zhang, Debo Cheng, Xingyu Wu, Peng Zhou, and Kui Yu. Hybrid local causal discovery. *arXiv preprint arXiv:2412.19507*, 2025.

[Ling *et al.*, 2025b] Zhaolong Ling, Jingxuan Wu, Yiwen Zhang, Peng Zhou, Xingyu Wu, Kui Yu, and Xindong Wu. Label-aware causal feature selection. *IEEE Transactions on Knowledge and Data Engineering*, 37(3):1268–1281, 2025.

[Ling *et al.*, 2025c] Zhaolong Ling, Jiale Yu, Yiwen Zhang, Debo Cheng, Peng Zhou, Xingyu Wu, Bingbing Jiang, and Kui Yu. Local causal discovery without causal sufficiency. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence (AAAI'25)*, volume 39, pages 18737–18745, 2025.

[Mani and Cooper, 2004] Subramani Mani and Gregory F Cooper. Causal discovery using a bayesian local causal discovery algorithm. In *MEDINFO 2004*, pages 731–735. IOS Press, 2004.

[Meek, 1995] Christopher Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 403–410, 1995.

[Peng *et al.*, 2005] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238, 2005.

[Prosperi *et al.*, 2020] Mattia Prosperi, Yi Guo, Matt Sperrin, James S Koopman, Jae S Min, Xing He, Shannan Rich, Mo Wang, Iain E Buchan, and Jiang Bian. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7):369–375, 2020.

[Sachs *et al.*, 2005] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

[Scutari *et al.*, 2019] Marco Scutari, Claudia Vitolo, and Allan Tucker. Learning bayesian networks from big data with greedy search: computational complexity and efficient implementation. *Statistics and Computing*, 29:1095–1108, 2019.

[Spirtes *et al.*, 2000] Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000.

[Tsamardinos *et al.*, 2003] Ioannis Tsamardinos, Constantin F Aliferis, and Alexander Statnikov. Time and sample efficient discovery of markov blankets and direct causal relations. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 673–678. ACM, 2003.

[Tsamardinos *et al.*, 2006] Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65:31–78, 2006.

[Van den Bulcke *et al.*, 2006] Tim Van den Bulcke, Koenraad Van Leemput, Bart Naudts, Piet van Remortel, Hongwu Ma, Alain Verschoren, Bart De Moor, and Kathleen Marchal. Syntren: a generator of synthetic gene expression data for design and analysis of structure learning algorithms. *BMC bioinformatics*, 7:1–12, 2006.

[Wang *et al.*, 2014] Changzhang Wang, You Zhou, Qiang Zhao, and Zhi Geng. Discovering and orienting the edges connected to a target variable in a dag via a sequential local learning approach. *Computational Statistics & Data Analysis*, 77:252–266, 2014.

[Wu *et al.*, 2021] Xingyu Wu, Bingbing Jiang, Kui Yu, and Huanhuan Chen. Separation and recovery markov boundary discovery and its application in eeg-based emotion recognition. *Information Sciences*, 571:262–278, 2021.

[Wu *et al.*, 2022] Xingyu Wu, Bingbing Jiang, Yan Zhong, and Huanhuan Chen. Multi-target markov boundary discovery: Theory, algorithm, and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4964–4980, 2022.

[Wu *et al.*, 2023] Xingyu Wu, Bingbing Jiang, Xiangyu Wang, Taiyu Ban, and Huanhuan Chen. Feature selection in the data stream based on incremental markov boundary learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(10):6740–6754, 2023.

[Wu *et al.*, 2024] Xingyu Wu, Yan Zhong, Zhaolong Ling, Jie Yang, Li Li, Weiguo Sheng, and Bingbing Jiang. Nonlinear learning method for local causal structures. *Information Sciences*, 654:119789, 2024.

[Xie *et al.*, 2020] Feng Xie, Ruichu Cai, Biwei Huang, Clark Glymour, Zeng Hao, and Kun Zhang. Generalized independent noise condition for estimating latent variable causal graphs. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, pages 14891–14902, 2020.

[Yang *et al.*, 2021] Shuai Yang, Hao Wang, Kui Yu, Fuyuan Cao, and Xindong Wu. Towards efficient local causal structure learning. *IEEE Transactions on Big Data*, 8(6):1592–1609, 2021.

[Yin *et al.*, 2008] Jianxin Yin, You Zhou, Changzhang Wang, Ping He, Cheng Zheng, and Zhi Geng. Partial orientation and local structural learning of causal networks for prediction. In *Causation and Prediction Challenge*, pages 93–105. PMLR, 2008.

[You *et al.*, 2023] Dianlong You, Siqi Dong, Shina Niu, Huigui Yan, Zhen Chen, Shunfu Jin, Di Wu, and Xindong Wu. Local causal structure learning for streaming features. *Information Sciences*, 647:119502, 2023.

[Yu *et al.*, 2020] Kui Yu, Xianjie Guo, Lin Liu, Jiuyong Li, Hao Wang, Zhaolong Ling, and Xindong Wu. Causality-based feature selection: Methods and evaluations. *ACM Computing Surveys (CSUR)*, 53(5):1–36, 2020.

[Yu *et al.*, 2023] Kui Yu, Zhaolong Ling, Lin Liu, Peipei Li, Hao Wang, and Jiuyong Li. Feature selection for efficient local-to-global bayesian network structure learning. *ACM Transactions on Knowledge Discovery from Data*, 18(2):1–27, 2023.

[Zeng *et al.*, 2021] Yan Zeng, Zhifeng Hao, Ruichu Cai, Feng Xie, Libo Huang, and Shohei Shimizu. Nonlinear causal discovery for high-dimensional deterministic data. *IEEE Transactions on Neural Networks and Learning Systems*, 34(5):2234–2245, 2021.

[Zhang *et al.*, 2021] Chen Zhang, Haodi Zhang, Weiteng Xie, Nan Liu, Kaishun Wu, and Lei Chen. Where to: Crowd-aided path selection by selective bayesian network. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):1072–1087, 2021.