# Causal Discovery for Cross-Sectional Data Based on Super-Structure and Divide-and-Conquer

1st Wenyu Wang
*School of Computer Science*
*University of South China*
Hengyang, China
20232008110502@stu.usc.edu.cn

2nd Yaping Wan
*School of Computer Science*
*University of South China*
Hengyang, China
ypwan@usc.edu.cn

***Abstract*—This paper tackles a critical bottleneck in Super-Structure-based divide-and-conquer causal discovery: the high computational cost of constructing accurate Super-Structures—particularly when conditional independence (CI) tests are expensive and domain knowledge is unavailable. We propose a novel, lightweight framework that relaxes the strict requirements on Super-Structure construction while preserving the algorithmic benefits of divide-and-conquer. By integrating weakly constrained Super-Structures with efficient graph partitioning and merging strategies, our approach substantially lowers CI test overhead without sacrificing accuracy. We instantiate the framework in a concrete causal discovery algorithm and rigorously evaluate its components on synthetic data. Comprehensive experiments on Gaussian Bayesian networks, including magic-NIAB, ECOLI70, and magic-IRRI, demonstrate that our method matches or closely approximates the structural accuracy of PC and FCI while drastically reducing the number of CI tests. Further validation on the real-world China Health and Retirement Longitudinal Study (CHARLS) dataset confirms its practical applicability. Our results establish that accurate, scalable causal discovery is achievable even under minimal assumptions about the initial Super-Structure, opening new avenues for applying divide-and-conquer methods to large-scale, knowledge-scarce domains such as biomedical and social science research.**

***Index Terms*—Bayesian networks, Computational complexity, Dimension reduction**

## I. Introduction

Causal discovery aims to recover a directed acyclic graph (DAG) from observational data, enabling mechanistic interpretation and downstream reasoning. In practice, however, learning causal structure at scale remains challenging. Mainstream paradigms include constraint-based methods [1], score-based search [2], and hybrid approaches [3]. When the number of variables grows, these methods often become dominated by conditional independence (CI) testing or combinatorial search over candidate graphs, leading to substantial computational overhead and degraded reliability [4].

To improve scalability, divide-and-conquer strategies have become increasingly attractive: they first partition variables into smaller subsets, learn local subgraphs, and then merge them into a global structure. A representative direction is to introduce a Super-Structure to constrain candidate edges and guide graph partitioning [5]. Yet existing Super-Structure-based pipelines typically require the Super-Structure to contain (or closely approximate) the true skeleton, which implicitly demands high recall. Constructing such a high-recall scaffold without domain knowledge is expensive, since it still requires many CI tests or heavy pre-processing, and the construction cost can erase the downstream gains.

This paper proposes a lightweight alternative by weakening the constraint on Super-Structures. Instead of requiring the true skeleton to be fully contained in the Super-Structure, we only require the Super-Structure to be a subgraph of the true skeleton. This shifts the design goal from high recall to high precision: the Super-Structure may miss edges, but the edges it includes are reliable and therefore safe to use for partitioning. Conceptually, this is analogous to a principle that has proven effective in other large-scale learning systems: injecting a small amount of reliable structure can dramatically reduce search complexity while keeping overall quality. For example, in controllable diffusion and conditional generation, explicitly enforcing reliable constraints (e.g., consistent quantity and layout) can stabilize optimization and reduce ambiguity in complex editing tasks [6]. Similarly, attribute-wise or modular conditioning can improve controllability and efficiency by restricting the hypothesis space to the most informative factors [7], [8]. Progressive conditioning and unified pose-guided formulations further illustrate how structured intermediate representations can improve robustness under challenging variations [9], [10]. Rich-context conditioning has also been shown to reduce long-range inconsistency in sequential generation by providing a more reliable contextual scaffold [11], while motion priors can stabilize long-horizon temporal coherence by constraining dynamics [12]. Inspired by these successes, we apply the same structural viewpoint to causal discovery: a weak but reliable graph scaffold can enable efficient decomposition without requiring an expensive, high-recall pre-estimation step.

We instantiate the proposed framework with a concrete algorithm and validate it on synthetic benchmarks and real-world datasets, including large-scale biomedical and social science data. Experiments demonstrate that our method substantially reduces the number of CI tests while maintaining competitive structural accuracy, showing that weakly constrained Super-

This research is supported by the Natural Science Foundation of Hunan Province,China(Grant No. 2024JJ7428)

Structures can deliver meaningful computational savings with minimal loss in fidelity.

Our contributions are:

- We introduce a causal partitioning framework that integrates weakly constrained Super-Structures with divide-and-conquer causal discovery, enabling efficient dimensionality reduction with a low-cost scaffold.
- We present an algorithmic realization that operationalizes this framework and provides practical efficiency improvements in high-dimensional settings.
- We provide empirical evidence on both synthetic and real datasets demonstrating large reductions in CI tests with minimal accuracy degradation.

## II. Related Work

Virtually all causal discovery algorithms operate by searching within a structured hypothesis space of directed acyclic graphs (DAGs). Divide-and-conquer approaches accelerate this search by partitioning the variable set into subsets, performing local causal discovery on each, and then merging the results to reconstruct the full causal graph. Causal discovery methods based on the divide-and-conquer approach can be categorized into three types:

*1) Recursive CI-based methods:* Cai et al. [13] introduced SADA, a recursive algorithm that, at each step, uses either multiple low-order or a single high-order CI test to infer a decomposition structure and derive variable subsets. However, the resulting subset sizes are uncontrolled and must be enumerated to meet target constraints. Zhang [14] proposed CAPA, another recursive method that partitions variables into two subsets per recursion using only low-order CI tests, combined with a regression-based CI test tailored for linear non-Gaussian additive noise models to preserve d-separation [15].

*2) Clustering-based approaches:* which partitions variables—treated as nodes—into clusters via hierarchical clustering, validated on continuous data. Huang & Zhou extended this with pHGS [16], adapting PEF for discrete data. Chen et al. [17] proposed CDSC, leveraging spectral clustering for dimensionality reduction in high-dimensional settings, while Wei et al. [18] developed CDKM, a K-means–based method using minimum correlation distance for variable clustering.

*3) Super-Structure–based methods:* By leveraging a Super-Structure, the causal partitioning problem can be naturally cast as graph partitioning. Zhang [19] proposed CPA, which reformulates causal partitioning as edge-cutting on the adjacency graph, stably splitting variables into two subsets. Building on the same foundation, Shah et al. [5] introduced the Expansive Causal Partition framework, which unifies multiple graph partitioning strategies to achieve flexible and scalable causal decomposition.

Perrier [20] introduced the Super-Structure—a scaffold for causal search constructed via low-order CI tests and/or domain knowledge—to reduce computational complexity. While Super-Structures can also be derived from CI testing or constraint-based methods, these alternatives incur substantial overhead in high-dimensional settings. To address this, we propose a weakly constrained Super-Structure, which relaxes construction requirements while retaining compatibility with the Super-Structure framework. Our approach supports diverse construction strategies—including domain knowledge, correlation measures, and score-based causal discovery—with details provided in **Section III**.

## III. Method

Following Perrier [20], a Super-Structure that contains the true skeleton is deemed sound; otherwise, it is incomplete. In this work, we adopt the converse convention for weakly constrained Super-Structures: those contained within the true skeleton are labeled sound, and others incomplete. Prior methods require highly sound Super-Structures to preserve d-separation across subgraphs and minimize redundant CI tests—yet their construction heavily relies on domain knowledge. In high-dimensional, mixed-distribution settings without such priors, achieving soundness becomes computationally prohibitive. To overcome this, we embrace a lightweight strategy that deliberately uses incomplete weakly constrained Super-Structures, sidestepping the burden of constructing fully sound ones. This approach reframes the problem as optimizing the algorithm's initial search point: as long as the weakly constrained Super-Structure is closer to the true graph than the naive starting point (e.g., a complete graph), it effectively reduces computational overhead. By leveraging such Super-Structures, we apply low-cost graph partitioning to decompose the problem into smaller subgraphs—dramatically lowering dimensionality, avoiding expensive high-order CI tests, and shrinking the search space. Building on this insight, we propose a divide-and-conquer causal discovery framework; an overview of the pipeline is shown in Fig. 1.

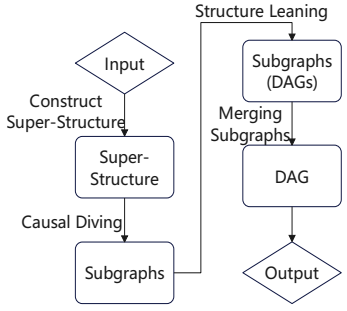


Fig. 1. Pipeline overview figure.

In the upcoming content of this section, we will demonstrate how to build a causal discovery algorithm based on the proposed framework.

### A. Super-Structure constructing module

This module takes a dataset as input and outputs a weakly constrained Super-Structure. To construct an extreme contrast to Perrier's "sound" Super-Structure [20], we employ the Chow–Liu algorithm [21], which builds a maximum spanning tree (MST) over a dependency matrix to yield a sparse,

tree-structured approximation of the underlying graph. The procedure consists of: (i) constructing a pairwise dependency matrix, and (ii) computing its MST.

We replace the original Pearson correlation with Copula entropy [22]—a model-free, non-parametric dependence measure—to better capture complex, non-Gaussian relationships. Formally, for random variables $X$ with marginal distributions and copula density $c(u)$ , Copula entropy quantifies dependence without distributional assumptions, enhancing robustness across diverse data types. The copula entropy of $X$ is

$$H_c(X) = -\int_u c(u) \log c(u) \mathrm{d}u \tag{1}$$

### B. Causal dividing module

This module takes a Super-Structure as input and outputs the segmented subgraphs, with the specific process shown in Fig. 2. Here we use Girvan's method [23] to implement graph

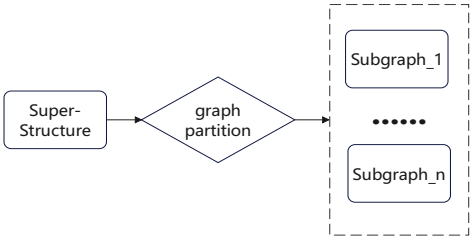


Fig. 2. Causal dividing module.

partitioning and use causal extension method [5] to process the subgraphs after partitioning. This module partitions nodes into subgraphs via the Super-Structure, reducing dimensionality and computational cost. However, the weakly constrained Super-Structure may omit the true skeleton, violating d-separation across subgraphs and necessitating additional CI tests during merging—partially eroding the initial computational gains.

### C. Subgraph Learning module

This module takes the subgraphs from the previous stage as input and outputs a skeleton graph. It employs a basic two-phase search strategy initialized from the subgraph as the starting point: (1) a **forward phase**, which iteratively tests all untested non-adjacent node pairs and adds edges for those exhibiting dependence; and (2) a **backward phase**, which re-evaluates all adjacent pairs and removes edges deemed conditionally independent.

### D. Subgraph merging module

The input of this module is the learning results of the previous module, and the merged DAG is the output. In this article, this module first completes the correction by traversing all pairs of untested nodes for CI using the same strategy as the Subgraph Learning module, and then implements the merging using Shah's method [5].The specific process is shown as follows:

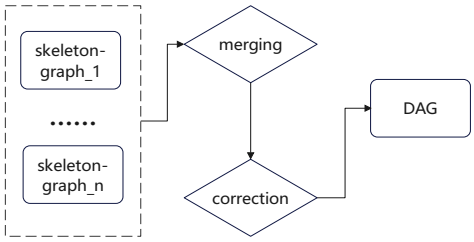


Fig. 3. Subgraph merging module.

## IV. Experiments

### A. Evaluation metrics

Causal discovery aims to recover causal relationships, and its evaluation metrics typically measure the accuracy of inferred causal structures. We adopt the following metrics: Structural Hamming Distance (SHD), F1-score, Precision, Recall, Accuracy, and CI test count—where CI test denotes the number of unique CI tests executed during skeleton learning. Under the three conditions outlined earlier, repeated CI tests are cached, rendering their time cost negligible; thus, skeleton learning constitutes the primary phase where constraint-based algorithms perform CI tests. Given that this study focuses on the skeleton learning phase, our experiments also concentrate on this aspect. The evaluation metrics above are applied to skeleton-graphs rather than DAGs.

### B. Simulation Experiment

*1) CD Module Ablation Experiment:* This experiment evaluates the impact of the Causal Dividing (CD) module on algorithm performance. We compare a partial variant (without subgraph partitioning) against the full algorithm on synthetic data, and further assess performance when initializing the Super-Structure with maximum spanning trees derived from weight matrices built using different dependence measures.Synthetic datasets are generated following Strobl's method [24]: for a given node count $p$ , we construct a random DAG with expected indegree $E(n) = 0.3795p$. Specifically, a lower-triangular adjacency matrix is sampled from $Bernoulli(0.075p/(p-1))$ ,then randomly permuted to break topological order. Edge weights are assigned i.i.d. from $Uniform[0.5, 0.9]$ to form the weighted adjacency matrix, which defines linear Gaussian SEM coefficients.To isolate the CD module's effect, we generate datasets with $p \in \{20, 22, \ldots, 40\}$ , each with 5,000 samples and Gaussian noise. Fisher's Z-test is used for CI testing. For each p , five graph instances are created; both algorithm variants are run four times per instance, and metrics are averaged. Results are summarized below:

Fig. 4∼Fig. 9: The causal dividing (CD) module ablation experiment results. The x-axis shows the number of nodes, and the y-axis corresponds to the indicators shown in the legend. The green legend indicates the results without using the CD module, and the blue legend indicates the results using the CD module.

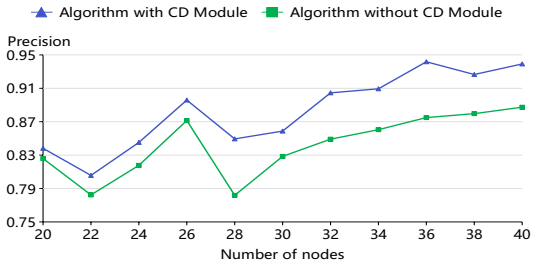


Fig. 4. Precision of ablation experiment.

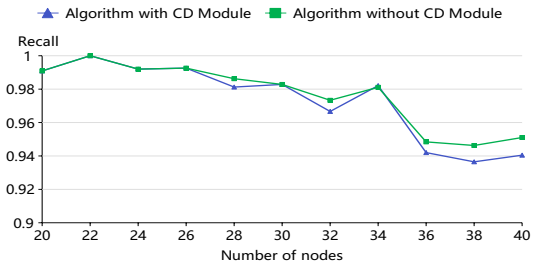


Fig. 5. Recall of ablation experiment.

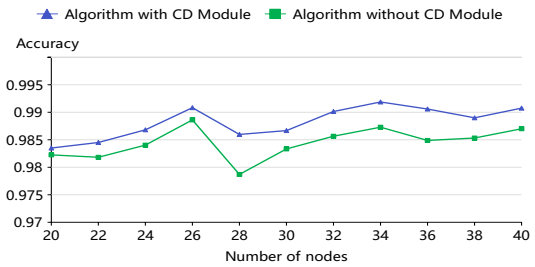


Fig. 6. Accuracy of ablation experiment.

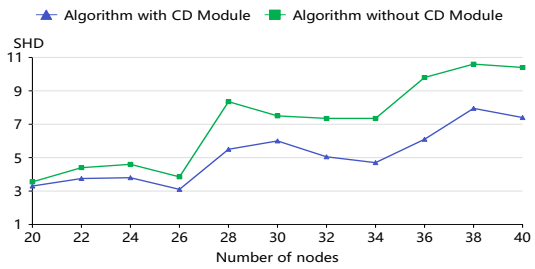


Fig. 7. Structural Hamming Distance of ablation experiment.

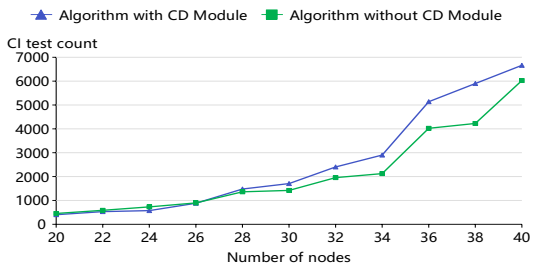


Fig. 8. CI test count of ablation experiment.

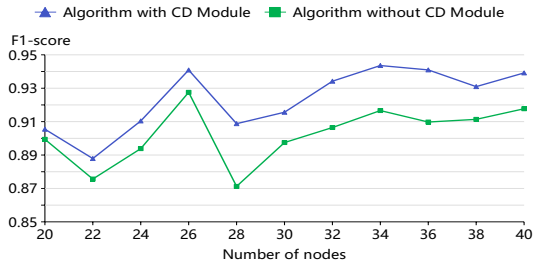


Fig. 9. F1-score of ablation experiment.

Experimental results show that the algorithm with the CD module outperforms the ablated version on most metrics—exhibiting only marginally higher CI test counts and slightly lower recall—demonstrating improved accuracy with minimal runtime overhead. This gain becomes more pronounced as the graph size increases, confirming that reduction in causal subset dimensionality through graph partitioning under a weakly constrained superstructure framework effectively enhances accuracy. The increased CI test count is primarily attributed to d-separation violations, which necessitate additional CI corrections. Given the current superstructure construction's limited fidelity, refining it to better approximate the true skeleton is expected to substantially reduce this computational overhead.

*2) Correlation Metrics Comparison Experiment:* This experiment evaluates the impact of different correlation metrics on algorithm performance when constructing weakly constrained Super-Structures. We generated synthetic datasets with 24 nodes and 5,000 samples under four noise distributions—Gaussian, exponential, gamma, and uniform—producing two datasets per distribution. Using Petersen's non-parametric CI test [25], we compared four correlation measures: Copula entropy, mutual information, Pearson, and Spearman coefficients. Each dataset was run twice, and results were averaged across runs. The findings are summarized below:

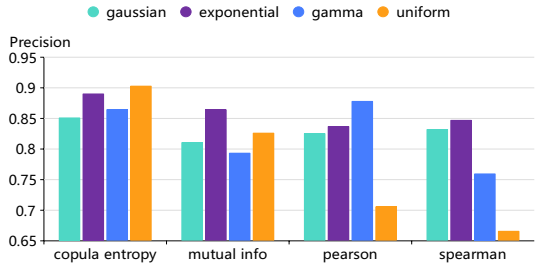


Fig. 10. Precision of correlation metrics comparison experiment.

Fig. 10∼Fig. 15: Experiment on effectiveness correlation metrics. The x-axis correspond to the metrics indicated in the legends, and the y-axes represents the correlation metrics from left to right: Copula Entropy, Mutual Information, Pearson coefficient, and Spearman coefficient. The green, purple,

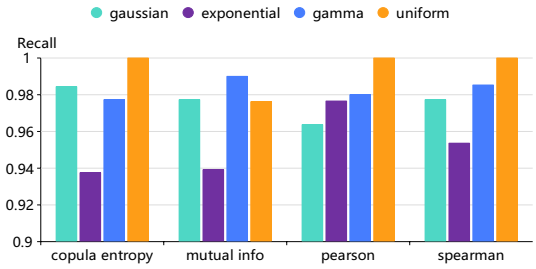


Fig. 11. Recall of correlation metrics comparison experiment.

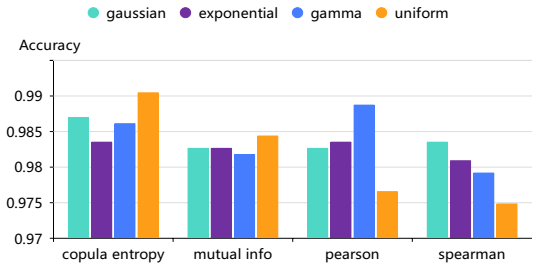


Fig. 12. Accuracy of correlation metrics comparison experiment.

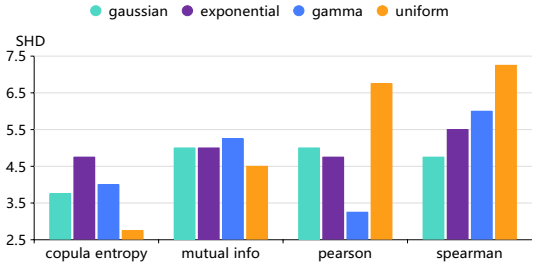


Fig. 13. Structural Hamming Distance of correlation metrics comparison experiment.

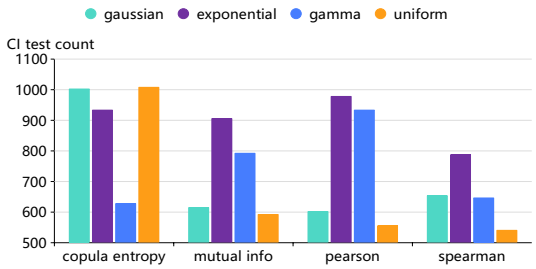


Fig. 14. CI test count of correlation metrics comparison experiment.

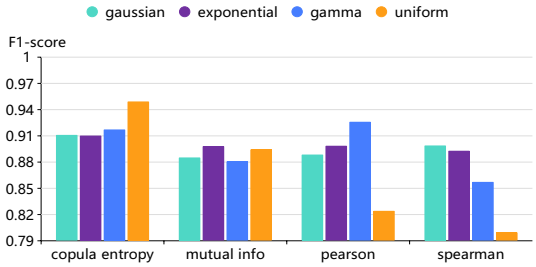


Fig. 15. F1-score of correlation metrics comparison experiment.

blue, and orange legends correspond to Gaussian, exponential, gamma, and uniform distributions, respectively.

The comparison experiment shows that, among correlation measures,**Copula entropy** yields the most robust Super-Structure construction—enabling our algorithm to better handle datasets with heterogeneous noise distributions. It outperforms alternatives on all metrics except recall and total CI tests, establishing it as the optimal choice for this task. Further experiments confirm that Copula entropy achieves the highest cross-distribution robustness and effectively captures dominant dependencies in mixed-distribution data.

*3) Algorithm Performance Comparison Experiment:* We use Gaussian Bayesian networks MAGIC-NIAB [26], ECOLI70 [27], and MAGIC-IRRI [26] to sample and generate simulated data to test the performance of our algorithm, with parameters as shown below.

TABLE I
PROPERTIES OF BAYESIAN NETWORKS

| **Bayesian networks** | MAGIC-NIAB | ECOLI70 | MAGIC-IRRI |
|---|---|---|---|
| **Number of nodes** | 44 | 46 | 64 |
| **Number of arcs** | 66 | 70 | 102 |
| **Average Markov blanket** | 9.88 | 3.65 | 9.97 |
| **Average degree** | 3 | 3.04 | 3.19 |
| **Maximum in-degree** | 9 | 4 | 11 |

Data were generated using the bnlearn R package by sampling Gaussian Bayesian networks from .rdn files. Each network produced 10 test sets of 5,000 samples each; every set was evaluated four times, and average performance metrics were reported. CI tests used Fisher's Z without prior knowledge. Baseline methods were FCI and PC; our method is denoted Ag. Results are presented below:

TABLE II
EXPERIMENTAL RESULTS FROM MAGIC-NIAB

| **Algorithm** | **Ag** | **FCI** | **PC** |
|---|---|---|---|
| **Accuracy** | 0.9840 | 0.9846 | 0.9841 |
| **Precision** | 0.8108 | 0.8176 | 0.8124 |
| **Recall** | 1.0000 | 1.0000 | 1.0000 |
| **F1-score** | 0.8952 | 0.8992 | 0.8961 |
| **SHD** | 15.5 | 14.9 | 15.4 |
| **CI test count** | 16366.4 | 94287.3 | 18893.8 |

TABLE III
EXPERIMENTAL RESULTS FROM ECOLI70

| **Algorithm** | **Ag** | **FCI** | **PC** |
|---|---|---|---|
| **Accuracy** | 0.9856 | 0.9863 | 0.9861 |
| **Precision** | 0.8672 | 0.8841 | 0.8803 |
| **Recall** | 0.9396 | 0.9314 | 0.9350 |
| **F1-score** | 0.8979 | 0.9018 | 0.9016 |
| **SHD** | 14.5 | 13.9 | 14.0 |
| **CI test count** | 11633.0 | 77255.9 | 27422.5 |

Experimental results show that, except on the largest network (MAGIC-IRRI), our method matches baseline accuracy—exhibiting only slightly lower precision and F1, and

TABLE IV
EXPERIMENTAL RESULTS FROM MAGIC-IRRI

| Algorithm | Ag | FCI | PC |
|---|---|---|---|
| **Accuracy** | 0.9856 | 0.9860 | 0.9858 |
| **Precision** | 0.8429 | 0.8544 | 0.8501 |
| **Recall** | 0.9475 | 0.9406 | 0.9429 |
| **F1-score** | 0.8888 | 0.8908 | 0.8897 |
| **SHD** | 19.5 | 19.1 | 19.4 |
| **CI test count** | 17303.4 | 199399.2 | 31921.0 |

modestly higher SHD and recall. All methods perform better on the sparser ECOLI70 than on MAGIC-NIAB. Crucially, our approach requires significantly fewer non-redundant CI tests across all networks, fulfilling its core design objective.These results indicate that even with a less accurate Super-Structure, overall causal learning performance remains robust. As a divide-and-conquer algorithm prioritizing CI test reduction over peak accuracy, our method successfully validates its central premise: relaxing strict Super-Structure constraints preserves correctness while enabling broader applicability.

*4) Real-world data:* We applied our method to cognitive function data from middle-aged and older adults in the China Health and Retirement Longitudinal Study (CHARLS) [28]. Inclusion criteria required participants to: (i) complete the global cognitive assessment and depression questionnaire in Wave 3; and (ii) have completed the cognitive test in Wave 2 and participated in the Wave 3 physical exam and second blood draw. Exclusion criteria were self-reported dementia, Parkinson's disease. Variable definitions are provided in the table below. We select variables from three heterogeneous sources—physical exams, blood tests, and scale assessments—each exhibiting distinct distributions. Standard CI tests struggle with such cross-distribution dependencies. To address this, we adopt Petersen's non-parametric CI test based on quantile regression and partial copulas [25], which excels at detecting CI relationships among continuous variables with arbitrary distributions. However, its high computational cost necessitates dimensionality reduction before applying it to structural learning on even moderately sized graphs—precisely the setting where divide-and-conquer causal discovery is most valuable. Conventional constraint-based divide-and-conquer methods first require costly CI tests to construct a Super-Structure or partition variables, especially in the absence of domain knowledge, rendering them impractical for such data. In contrast, our method constructs the Super-Structure **without any CI testing**, drastically reducing the total number of CI queries. This efficiency gain makes it feasible to deploy powerful but expensive non-parametric CI tests like Petersen's in real-world, heterogeneous datasets. After applying Z-score normalization to all input data and handling missing values using Strobl's method [24], the final results are shown in Fig. 16.Algorithmic analysis results indicate that cognitive scores are influenced by factors such as depressive symptoms, prior decline in cognitive scores, age, and lung capacity. These findings align with existing consensus in cognitive science: cognitive function tends to decline with age in middle-aged and older adults [29]. Since cognitive scores essentially reflect an individual's cognitive ability, which is closely related to educational attainment, and educational attainment influences the trajectory of cognitive decline with age in middle-aged and older populations [30]. Depressive symptoms show a significant association with cognitive function and are an important risk factor [31]. Lung capacity can partially reflect the physical activity levels of middle-aged and older adults, and regular exercise has been proven to slow the progression of Alzheimer's disease [32]. By comparing algorithmic learning outcomes with domain knowledge, it was found that the algorithm can, to some extent, reconstruct the causal relationships in the data.

TABLE V
VARIABLE NAMES AND DEFINITIONS

| Variable | Definition |
|---|---|
| X1 | Depression |
| X2 | Cognitive Function |
| X3 | White Blood Cell |
| X4 | Hemoglobin |
| X5 | Hematocrit |
| X6 | Mean Corpuscular Volume |
| X7 | Platelets |
| X8 | Triglycerides |
| X9 | Creatinine |
| X10 | Blood Urea Nitrogen |
| X11 | High Density Lipoprotein Cholesterol |
| X12 | Low Density Lipoprotein Cholesterol |
| X13 | Total Cholesterol |
| X14 | Glucose |
| X15 | Uric Acid |
| X16 | Cystatin C |
| X17 | C-Reactive Protein |
| X18 | Glycated Hemoglobin |
| X19 | Age |
| X20 | Glomerular Filtration Rate |
| X21 | Average Pulse Measure |
| X22 | Body Mass Index |
| X23 | Maximum Lung Function Peak Flow |
| X24 | Cognitive Function Decrease (from the previous peak) |

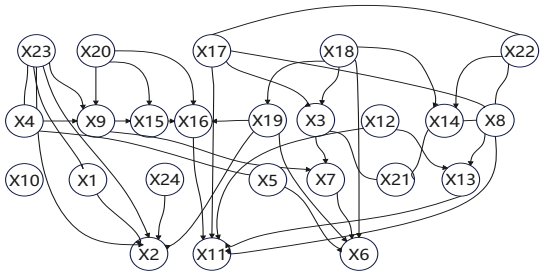


Fig. 16. Results of algorithm execution

## V. CONCLUSION

We present a lightweight causal discovery framework that integrates weakly constrained Super-Structures with a divide-and-conquer strategy, enabling low-cost construction without domain knowledge—making it well-suited for large-scale, exploratory medical research. Our work provides a clear

methodological contribution by formalizing the pipeline into three stages: Super-Structure construction, subgraph partitioning, and merging. Extensive experiments on synthetic, benchmark, and real-world datasets demonstrate its robustness and practical applicability. Although not yet state-of-the-art in accuracy, the framework offers significant computational advantages and opens new avenues for simplifying causal discovery and dimensionality reduction, encouraging a re-examination of the role of d-separation in divide-and-conquer approaches. Future work will focus on improving the efficiency and fidelity of weakly constrained Super-Structure construction and integrating advanced learning and merging strategies for broader applicability.

## REFERENCES


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

[2] D. M. Chickering, "Optimal Structure Identification With Greedy Search," *Journal of machine learning research*, vol. 3, no. Nov, pp. 507–554, Nov. 2005.

[3] I. Tsamardinos, L. E. Brown, and C. F. Aliferis, "The max-min hill-climbing Bayesian network structure learning algorithm," *Machine Learning*, vol. 65, no. 1, pp. 31–78, Oct. 2006. [Online]. Available: http://link.springer.com/10.1007/s10994-006-6889-7

[4] W. Bergsma, Pieter, "Testing conditional independence for continuous random variables," Eurandom, Tech. Rep., 2004.

[5] A. Shah, A. DePavia, N. Hudson, I. Foster, and R. Stevens, "Causal Discovery over High-Dimensional Structured Hypothesis Spaces with Causal Graph Partitioning," *Transactions on Machine Learning Research*, 2025. [Online]. Available: https://openreview.net/forum?id=FecsgPCOHk

[6] F. Shen, X. Du, Y. Gao, J. Yu, Y. Cao, X. Lei, and J. Tang, "Imagharmony: Controllable image editing with consistent object quantity and layout," *arXiv preprint arXiv:2506.01949*, 2025.

[7] F. Shen, J. Yu, C. Wang, X. Jiang, X. Du, and J. Tang, "Imaggarment-1: Fine-grained garment generation for controllable fashion design," *arXiv preprint arXiv:2504.13176*, 2025.

[8] F. Shen, X. Jiang, X. He, H. Ye, C. Wang, X. Du, Z. Li, and J. Tang, "Imagdressing-v1: Customizable virtual dressing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 7, 2025, pp. 6795–6804.

[9] F. Shen, H. Ye, J. Zhang, C. Wang, X. Han, and Y. Wei, "Advancing pose-guided image synthesis with progressive conditional diffusion models," in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: https://openreview.net/forum?id=rHzapPnCgT

[10] F. Shen and J. Tang, "Imagpose: A unified conditional framework for pose-guided person generation," *Advances in neural information processing systems*, vol. 37, pp. 6246–6266, 2024.

[11] F. Shen, H. Ye, S. Liu, J. Zhang, C. Wang, X. Han, and Y. Wei, "Boosting consistency in story visualization with rich-contextual conditional diffusion models," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 7, 2025, pp. 6785–6794.

[12] F. Shen, C. Wang, J. Gao, Q. Guo, J. Dang, J. Tang, and T.-S. Chua, "Long-term talkingface generation via motion-prior conditional diffusion model," in *Forty-second International Conference on Machine Learning*, 2025.

[13] R. Cai, Z. Zhang, and Z. Hao, "SADA: A General Framework to Support Robust Causation Discovery," in *JMLR: W&CP*, vol. 28. Atlanta: PMLR, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:15780170

[14] H. Zhang, S. Zhou, C. Yan, J. Guan, X. Wang, J. Zhang, and J. Huan, "Learning Causal Structures Based on Divide and Conquer," *IEEE Transactions on Cybernetics*, vol. 52, no. 5, pp. 3232–3243, May 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9165127/

[15] J. PEARL, *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press, 2000.

[16] J. Huang and Q. Zhou, "Partitioned hybrid learning of Bayesian network structures," *Machine Learning*, vol. 111, no. 5, pp. 1695–1738, May 2022. [Online]. Available: https://link.springer.com/10.1007/s10994-022-06145-4

[17] S. Chen, Y. Peng, G. He, H. Zhang, L. Cai, and C. Wei, "CDSC: Causal decomposition based on spectral clustering," *Information Sciences*, vol. 657, p. 119985, Feb. 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0020025523015700

[18] H. Wei, C. Wei, S. Chen, G. He, and Z. Li, "CDKM: Causal Decomposition Method Based on K-means Clustering," *Guangxi Sciences*, vol. 32, no. 1, pp. 121–131, 2025.

[19] H. Zhang, Y. Ren, Y. Xia, S. Zhou, and J. Guan, "Towards Effective Causal Partitioning by Edge Cutting of Adjoint Graph," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10 259–10 271, Dec. 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10614830/

[20] E. Perrier, S. Imoto, and S. Miyano, "Finding Optimal Bayesian Network Given a Super-Structure," *Journal of Machine Learning Research*, vol. 9, pp. 2251–2286, 2008.

[21] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, May 1968. [Online]. Available: https://ieeexplore.ieee.org/document/1054142/

[22] M. Jian, "Mutual Information Is Copula Entropy," *Tsinghua Science & Technology*, vol. 16, no. 1, p. 4, 2011.

[23] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002, arXiv:cond-mat/0112110. [Online]. Available: http://arxiv.org/abs/cond-mat/0112110

[24] E. V. Strobl, S. Visweswaran, and P. L. Spirtes, "Fast causal inference with non-random missingness by test-wise deletion," *International Journal of Data Science and Analytics*, vol. 6, no. 1, pp. 47–62, Aug. 2018. [Online]. Available: http://link.springer.com/10.1007/s41060-017-0094-6

[25] L. Petersen and N. R. Hansen, "Testing Conditional Independence via Quantile Regression Based Partial Copulas," *Journal of Machine Learning Research*, vol. 22, pp. 1–47, 2021.

[26] M. Scutari, P. Howell, D. J. Balding, and I. Mackay, "Multiple Quantitative Trait Analysis Using Bayesian Networks," *Genetics*, vol. 198, no. 1, pp. 129–137, Sep. 2014. [Online]. Available: https://academic.oup.com/genetics/article/198/1/129/6081358

[27] J. Schäfer and K. Strimmer, "A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, Jan. 2005.

[28] Y. Zhao, Y. Hu, J. P. Smith, J. Strauss, and G. Yang, "Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS)," *International Journal of Epidemiology*, vol. 43, no. 1, pp. 61–68, Feb. 2014. [Online]. Available: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dys203

[29] J. Su and X. Xiao, "Factors leading to the trajectory of cognitive decline in middle-aged and older adults using group-based trajectory modeling: A cohort study," *Medicine*, vol. 101, no. 47, p. e31817, Nov. 2022. [Online]. Available: https://journals.lww.com/10.1097/MD.0000000000031817

[30] H. Li, C. Li, A. Wang, Y. Qi, W. Feng, C. Hou, L. Tao, X. Liu, X. Li, W. Wang, D. Zheng, and X. Guo, "Associations between social and intellectual activities with cognitive trajectories in Chinese middle-aged and older adults: a nationally representative cohort study," *Alzheimer's Research & Therapy*, vol. 12, no. 1, p. 115, Dec. 2020. [Online]. Available: https://alzres.biomedcentral.com/articles/10.1186/s13195-020-00691-6

[31] Y. Sun, J. Feng, Z. Lei, G. Qu, X. Li, and Y. Gan, "Correlation between depressive symptoms and cognitive function in middle-aged and elderly population in China: an analysis of CHARLS baseline data," *Chinese Journal of Public Health*, vol. 40, no. 10, pp. 1206–1211, 2024.

[32] R.-x. Jia, J.-h. Liang, Y. Xu, and Y.-q. Wang, "Effects of physical activity and exercise on the cognitive function of patients with Alzheimer disease: a meta-analysis," *BMC Geriatrics*, vol. 19, no. 1, p. 181, Dec. 2019. [Online]. Available: https://bmcgeriatr.biomedcentral.com/articles/10.1186/s12877-019-1175-2