

**IMPERIAL**

KCL-ICL Causality Workshop  
24<sup>th</sup> March 2025

# Causal Discovery for Trustworthy AI

From Contestable Neural Networks to Transparent Causal Discovery

Fabrizio Russo

---

## Outline

- ① Task and Assumptions
  - ② Causal Graphs for Contestable Neural Networks
  - ③ Causal Discovery with Shapley Values
  - ④ Argumentative Causal Discovery
  - ⑤ Conclusion and Future Work
-

# Introduction

## Tasks and Assumptions

- Causal discovery from observational data
- Predictions for tabular data with neural networks

## Tasks and Assumptions

- Causal discovery from observational data
- Predictions for tabular data with neural networks
- **Desiderata:** transparency, robustness, accountability

## Tasks and Assumptions

- Causal discovery from observational data
- Predictions for tabular data with neural networks
- **Desiderata:** transparency, robustness, accountability
- **Assumptions:** Sufficiency, Acyclicity, Faithfulness

# Causal Graphs for Contestable Neural Networks

Russo & Toni. *Causal Discovery and Knowledge Injection for Contestable Neural Networks*. ECAI 2023

---

## Motivation

- Neural networks are black boxes

## Motivation

- Neural networks are black boxes
- Contestability has been advocated by major AI ethics frameworks and regulations (e.g. OECD, ACM and GDPR)

## Motivation

- Neural networks are black boxes
- Contestability has been advocated by major AI ethics frameworks and regulations (e.g. OECD, ACM and GDPR)

*Can causal discovery make NNs more transparent, human-aligned and robust?*

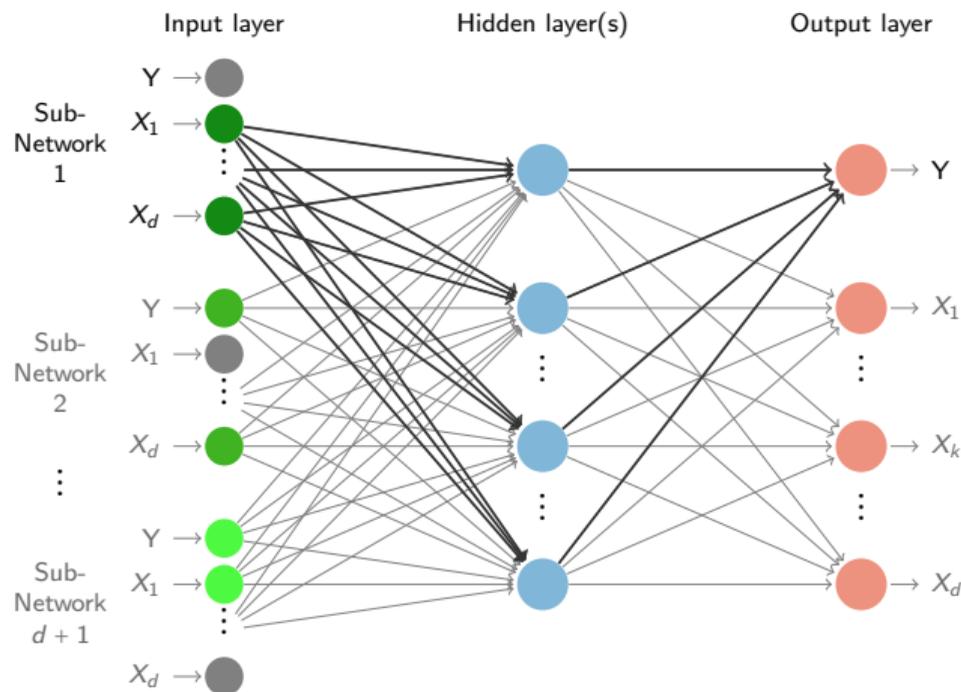
## Contribution

### Contestable Neural Networks

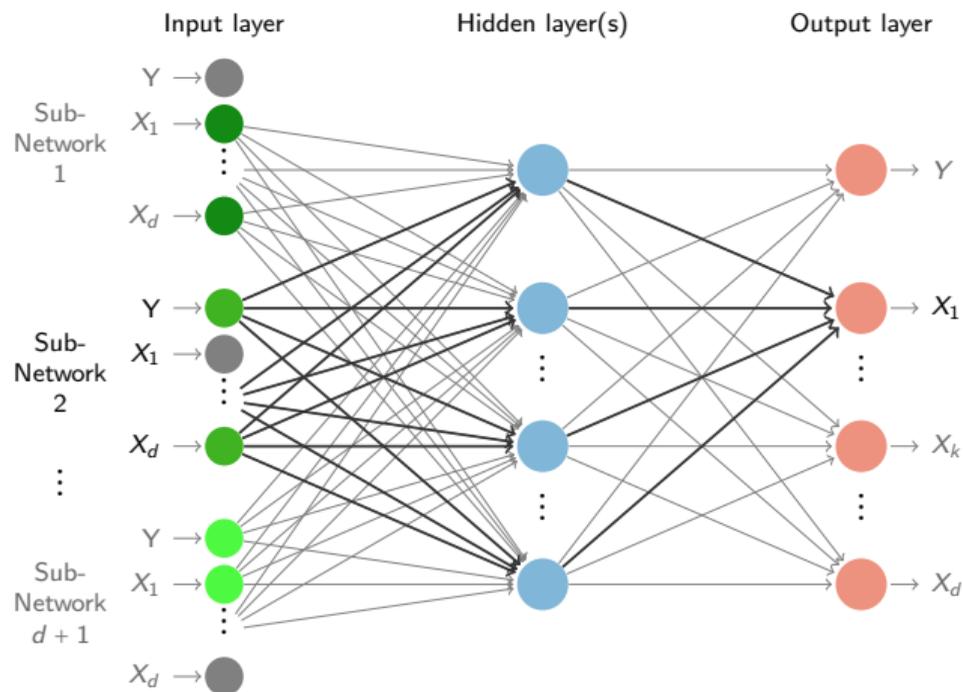
We propose *Knowledge Injection* empowered by *Causal Discovery* as a mean to make neural networks *Contestable* and improve them through experts' feedback.

This enables human-in-the-loop debugging and increased transparency.

# Joint Network Structure (Kyono, Zhang and van der Schaar 2020)

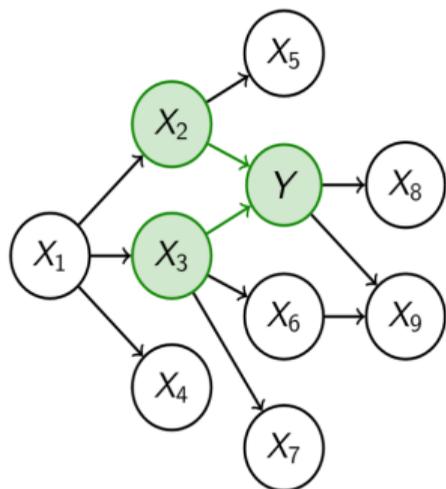


## Joint Network Structure (Kyono, Zhang and van der Schaar 2020)



## The Intuition

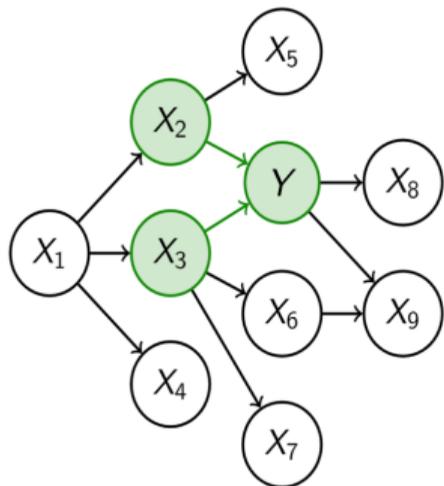
- **Objective:** have the network use only the relationships contained in the DAG i.e. predict each of the features using only its parents.



	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
Y	0.0	0.005	0.017	0.008	0.002	0.042	0.02	0.005	0.059	0.05
X <sub>1</sub>	0.006	0.0	0.063	0.054	0.068	0.009	0.006	0.013	0.006	0.008
X <sub>2</sub>	0.088	0.036	0.0	0.022	0.019	0.124	0.008	0.011	0.006	0.008
X <sub>3</sub>	0.087	0.034	0.021	0.0	0.024	0.005	0.107	0.104	0.006	0.009
X <sub>4</sub>	0.009	0.032	0.02	0.023	0.0	0.01	0.013	0.01	0.005	0.005
X <sub>5</sub>	0.026	0.006	0.017	0.004	0.004	0.0	0.012	0.002	0.005	0.018
X <sub>6</sub>	0.025	0.006	0.008	0.011	0.005	0.017	0.0	0.014	0.002	0.114
X <sub>7</sub>	0.029	0.003	0.007	0.011	0.002	0.024	0.029	0.0	0.011	0.01
X <sub>8</sub>	0.036	0.002	0.004	0.003	0.004	0.006	0.009	0.006	0.0	0.006
X <sub>9</sub>	0.024	0.003	0.003	0.004	0.003	0.005	0.079	0.01	0.004	0.0

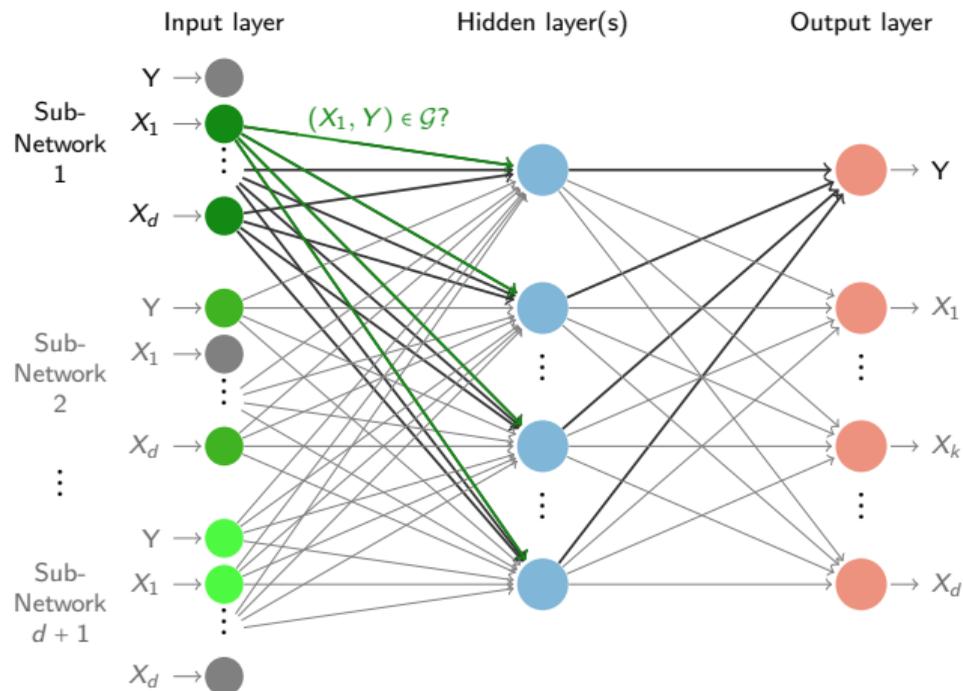
## The Intuition

- **Objective:** have the network use only the relationships contained in the DAG i.e. predict each of the features using only its parents.

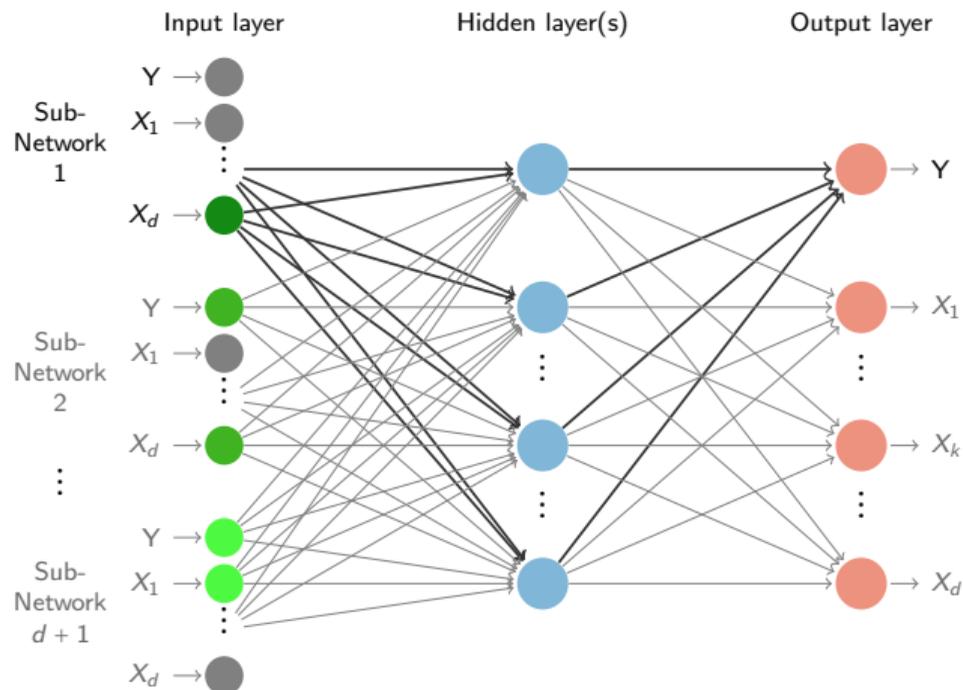


	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
Y	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.059	0.05
X <sub>1</sub>	0.0	0.0	0.063	0.054	0.068	0.0	0.0	0.0	0.0	0.0
X <sub>2</sub>	0.088	0.0	0.0	0.0	0.0	0.124	0.0	0.0	0.0	0.0
X <sub>3</sub>	0.087	0.0	0.0	0.0	0.0	0.0	0.107	0.104	0.0	0.0
X <sub>4</sub>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
X <sub>5</sub>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
X <sub>6</sub>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.114
X <sub>7</sub>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
X <sub>8</sub>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
X <sub>9</sub>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

# Inject Causal Knowledge



## Inject Causal Knowledge



## Summary of Results

- Injected Networks can perform significantly better than causal regularisation

## Summary of Results

- Injected Networks can perform significantly better than causal regularisation
- Injecting a set of partial constraints does not generally hurt performance

## Summary of Results

- Injected Networks can perform significantly better than causal regularisation
- Injecting a set of partial constraints does not generally hurt performance
- Injecting part of the causal relations helps the causal discovery task

## Limitations

- Contestable Neural Networks are not a tool for Causal Inference

## Limitations

- Contestable Neural Networks are not a tool for Causal Inference
- Masking is only applied to direct effects

## Limitations

- Contestable Neural Networks are not a tool for Causal Inference
- Masking is only applied to direct effects
- Causal Graphs make Networks more transparent, what about transparency and assurance around the Causal Graphs themselves?

# Causal Discovery with Shapley Values

Russo & Toni. *Shapley-PC: Constraint-based Causal Structure Learning with a Shapley Inspired Framework*. CLear 2025. (Forthcoming)

## Shapley Values

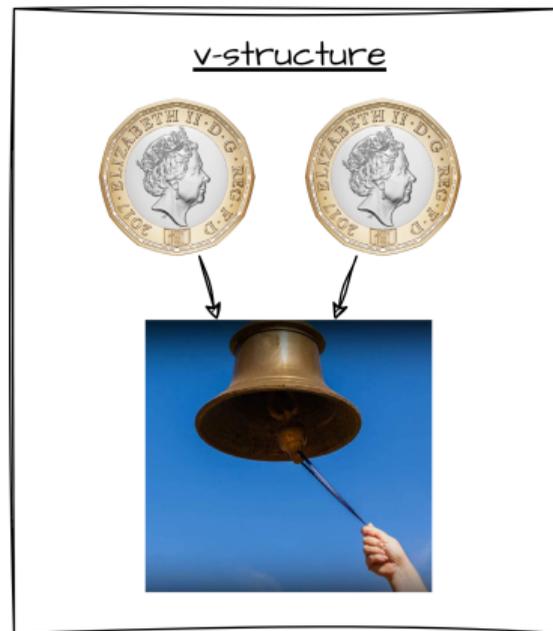
- Shapley value in cooperative game theory (Shapley 1953)
- Suppose a **team**  $N = 1, \dots, n$  of players cooperates to earn **value**  $v(N)$ :

$$\phi_v(i) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} [v(S \cup \{i\}) - v(S)] \quad (1)$$

- Represents a **player**  $i$ 's marginal value-added upon joining a team

## V-structures or Immoralities

- Two variables are marginally independent  $Coin1 \perp\!\!\!\perp Coin2$
- But they are dependent when conditioned on a common descendant  $Coin1 \not\perp\!\!\!\perp Coin2 \mid Bell$



## Shapley Independence Values (SIV)

- Suppose we are assessing whether  $X_i - X_j - X_k$  in  $\mathcal{C}$  is a v-structure

## Shapley Independence Values (SIV)

- Suppose we are assessing whether  $X_i - X_j - X_k$  in  $\mathcal{C}$  is a v-structure
- For a given skeleton  $\mathcal{C}$  our **team** of variables is:

$$\mathbf{N} = \{\mathbf{S} : \mathbf{S} \subseteq \text{adj}(\mathcal{C}, X_i) \setminus \{X_j\} \vee \mathbf{S} \subseteq \text{adj}(\mathcal{C}, X_k) \setminus \{X_j\}\} \quad (2)$$

## Shapley Independence Values (SIV)

- Suppose we are assessing whether  $X_i - X_j - X_k$  in  $\mathcal{C}$  is a v-structure
- For a given skeleton  $\mathcal{C}$  our **team** of variables is:

$$\mathbf{N} = \{\mathbf{S} : \mathbf{S} \subseteq \text{adj}(\mathcal{C}, X_i) \setminus \{X_j\} \vee \mathbf{S} \subseteq \text{adj}(\mathcal{C}, X_k) \setminus \{X_j\}\} \quad (2)$$

- Our **value** function is the  $p$ -value returned by the conditional independence test  $I(X_i, X_k | \mathbf{S})$

## Shapley Independence Values (SIV)

### *Shapley Independence Value (SIV)*

$$\phi_I(X_j, \{X_i, X_k\}) = \sum_{\mathbf{S} \in \mathbf{N}} \frac{|\mathbf{S}|!(n - |\mathbf{S}| - 1)!}{n!} [I(X_i, X_k | \mathbf{S} \cup \{X_j\}) - I(X_i, X_k | \mathbf{S})] \quad (3)$$

## Shapley Independence Values (SIV)

### Shapley Independence Value (SIV)

$$\phi_I(X_j, \{X_i, X_k\}) = \sum_{\mathbf{S} \in \mathbf{N}} \frac{|\mathbf{S}|!(n - |\mathbf{S}| - 1)!}{n!} [I(X_i, X_k | \mathbf{S} \cup \{X_j\}) - I(X_i, X_k | \mathbf{S})] \quad (3)$$

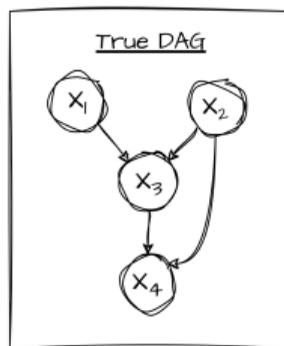
The higher  $\phi_I(X_j, \{X_i, X_k\})$  the higher is the  $X_j$ 's contribution the independence between  $X_i$  and  $X_k$ .

## Contribution

### Shapley Independence Values (SIV)

We prove that, given adjacency faithfulness and perfect conditional independence information, negative SIVs correctly identify v-structures.

## PC-Algorithm (Spirtes, Glymour and Scheines 2000)

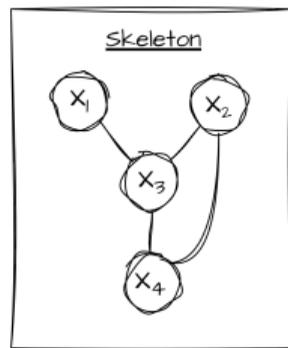


**Input:** Conditional Independence Information for all variables

- 1: Build skeleton  $\mathcal{C}$  via adjacency search (save separating sets)
- 2: Orient v-structures in  $\mathcal{C}$  (using separating sets from step 1)
- 3: Propagate *d-separation* via Meek rules

**Return:** CPDAG

## PC-Algorithm (Spirtes, Glymour and Scheines 2000)

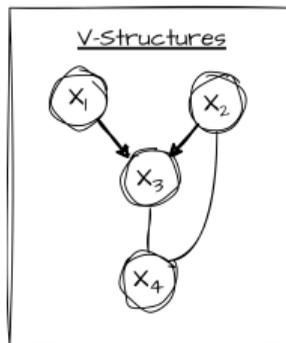


**Input:** Conditional Independence Information for all variables

- 1: **Build skeleton  $\mathcal{C}$  via adjacency search**
- 2: Orient v-structures in  $\mathcal{C}$
- 3: Propagate orientations via Meek rules

**Return:** CPDAG

## PC-Algorithm (Spirtes, Glymour and Scheines 2000)

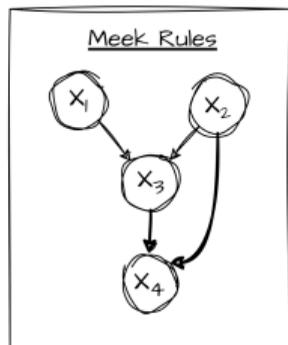


**Input:** Conditional Independence Information for all variables

- 1: Build skeleton  $\mathcal{C}$  via adjacency search
- 2: **Orient v-structures in  $\mathcal{C}$**
- 3: Propagate orientations via Meek rules

**Return:** CPDAG

## PC-Algorithm (Spirtes, Glymour and Scheines 2000)



**Input:** Conditional Independence Information for all variables

- 1: Build skeleton  $\mathcal{C}$  via adjacency search
- 2: Orient v-structures in  $\mathcal{C}$
- 3: **Propagate orientations via Meek rules**

**Return:** CPDAG

## Summary of Results

- Keep the theoretical soundness and asymptotic consistency of PC
- No computational overhead (compared to modifications of PC)
- Remove the dependency on the usual significance threshold  $\alpha$

## Summary of Results

- Keep the theoretical soundness and asymptotic consistency of PC
- No computational overhead (compared to modifications of PC)
- Remove the dependency on the usual significance threshold  $\alpha$
- Outperforms all previously proposed variations of PC using the same conditional independence tests - for "close to unfaithful" data (Ramsey, Spirtes and J. Zhang 2006)

## Limitations

- Assumes Sufficiency and Acyclicity

## Limitations

- Assumes Sufficiency and Acyclicity
- Only used for v-structure discovery, could it be extended to the other discovery phases?

## Limitations

- Assumes Sufficiency and Acyclicity
- Only used for v-structure discovery, could it be extended to the other discovery phases?
- Liable to consistency errors due to its sequential nature

# Argumentative Causal Discovery

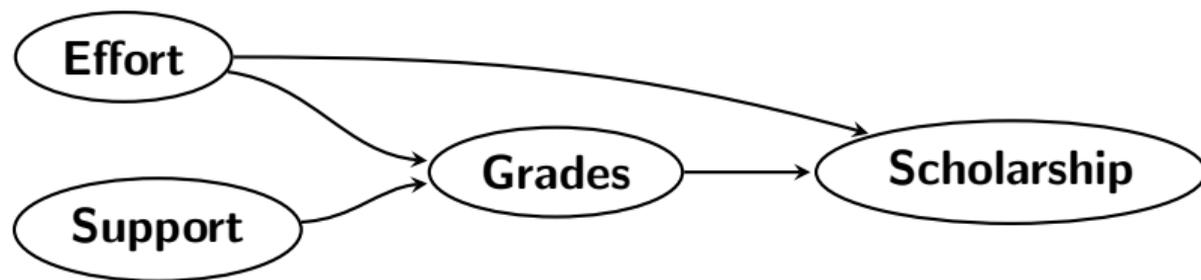
Russo, Rapberger & Toni. *Argumentative Causal Discovery*. KR 2024.

---

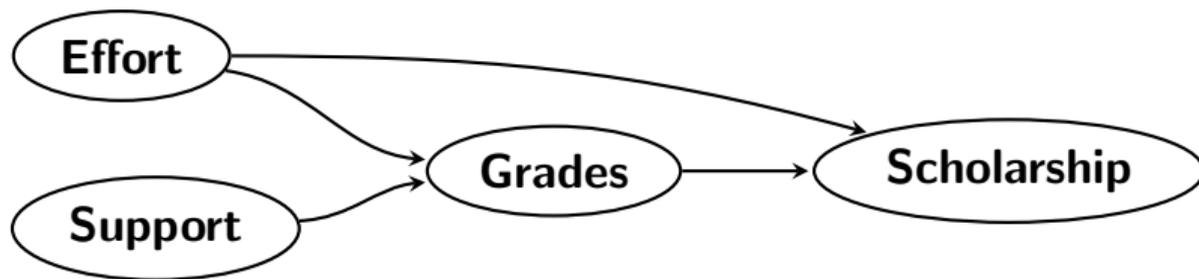
## Student's Achievement Example



## Student's Achievement Example

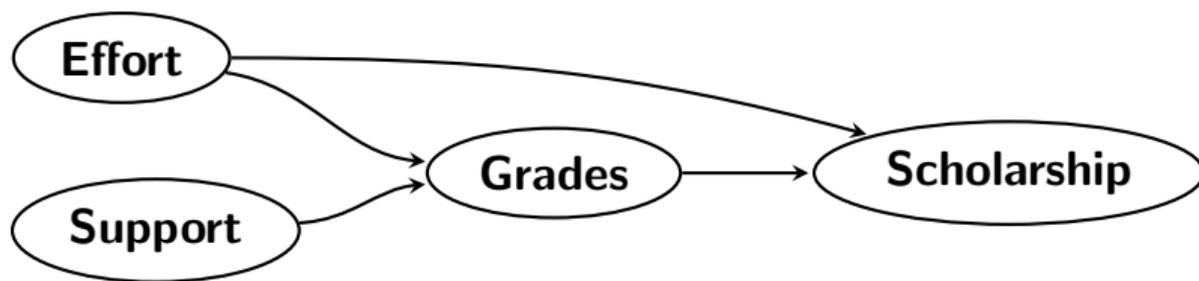


## Student's Achievement Example



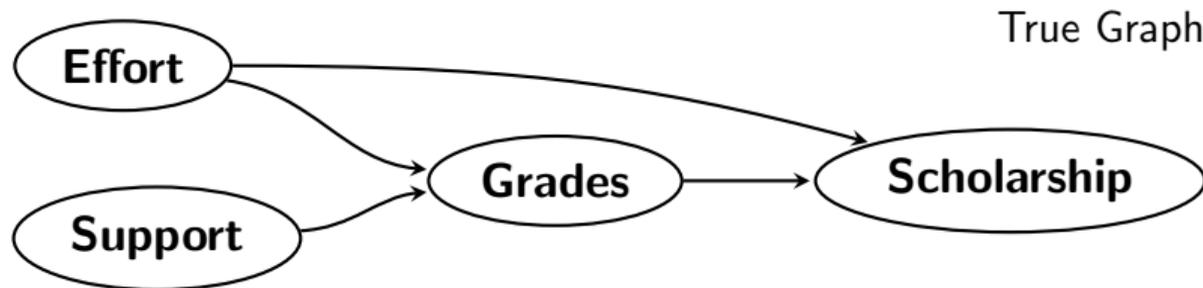
- Effort  $\perp\!\!\!\perp$  Support
- Effort  $\perp\!\!\!\perp$  Support | Grades
- Effort  $\perp\!\!\!\perp$  Support | Scholar
- Effort  $\perp\!\!\!\perp$  Support | {Grades, Scholar}
- Support  $\not\perp\!\!\!\perp$  Scholar
- Support  $\perp\!\!\!\perp$  Scholar | Effort
- Support  $\not\perp\!\!\!\perp$  Scholar | Grades
- Support  $\not\perp\!\!\!\perp$  Scholar | {Grades, Effort}

## Student's Achievement Example

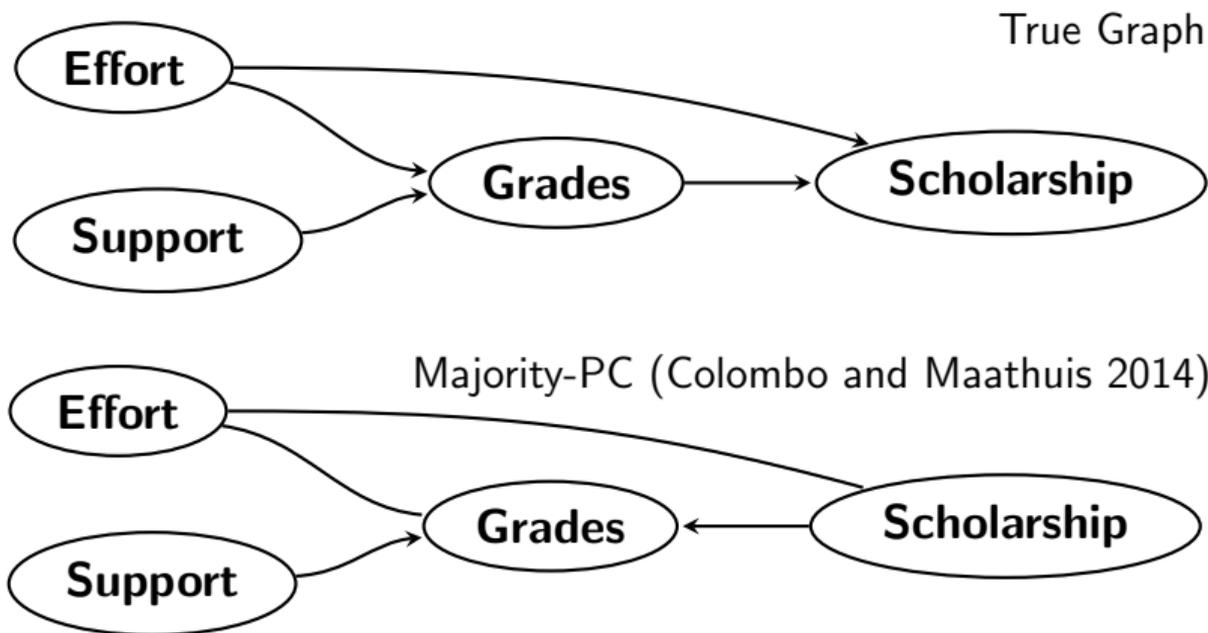


- Effort  $\perp\!\!\!\perp$  Support
- Effort  $\perp\!\!\!\perp$  Support | Grades
- Effort  $\perp\!\!\!\perp$  Support | Scholar
- Effort  $\perp\!\!\!\perp$  Support | {Grades, Scholar}
- Support  $\not\perp\!\!\!\perp$  Scholar
- Support  $\perp\!\!\!\perp$  Scholar | Effort
- Support  $\not\perp\!\!\!\perp$  Scholar | Grades
- Support  $\not\perp\!\!\!\perp$  Scholar | {Grades, Effort}

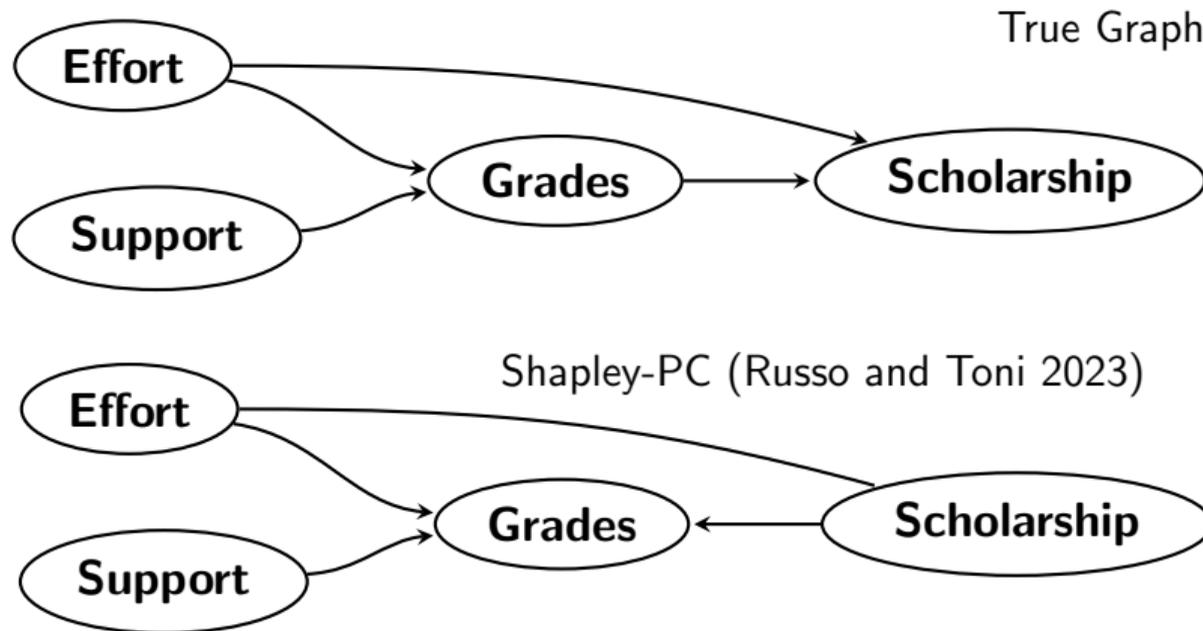
## Student's Achievement Example



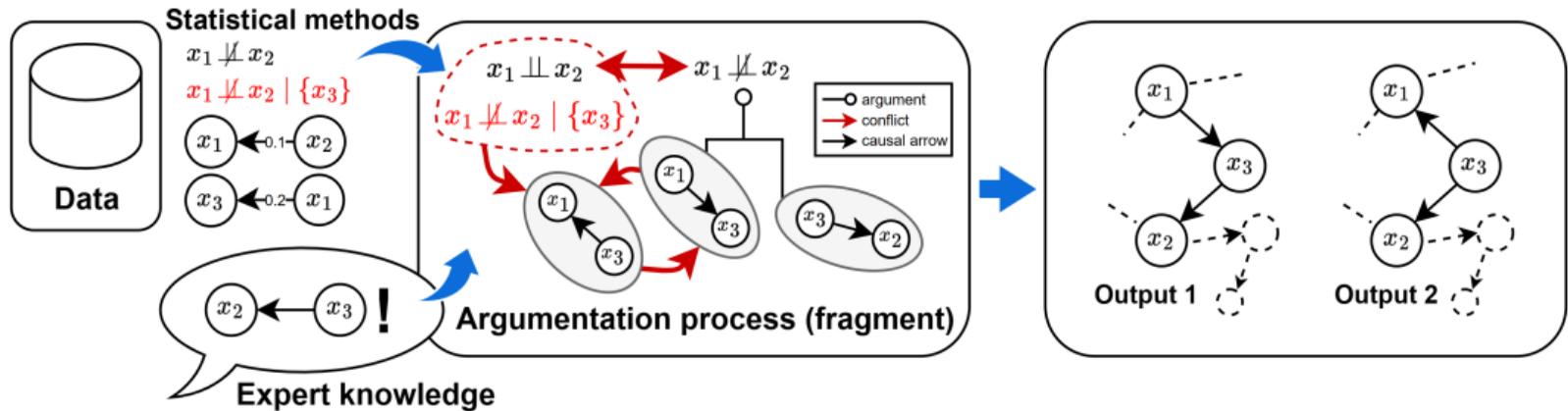
## Student's Achievement Example



## Student's Achievement Example



## Causal Assumption-Based Argumentation



## Summary of Results

- Ensure the consistency between a subset of the relations provided in input and the output causal graph
- Strong empirical performance against state-of-the-art baselines on bnlearn data
- Opens up avenues of future work on interpretability and contestability

## Limitations

- Assumes Sufficiency and Acyclicity

## Limitations

- Assumes Sufficiency and Acyclicity
  
- Scalability - Can handle max 11 variables

## Conclusion & Future Work

## Conclusions & Future Work

- We discussed three methods to
  - Employ Causal Graphs as means to contest and improve Neural Networks
  - Discover Causal Graphs more robustly
  - Discover and potentially explain and contest causal graphs

## Conclusions & Future Work

- We discussed three methods to
  - Employ Causal Graphs as means to contest and improve Neural Networks
  - Discover Causal Graphs more robustly
  - Discover and potentially explain and contest causal graphs
- Guide human input into construction of causal models
- Improve scalability while maintaining reliability

Thanks for your attention!  
Questions?

---

## References I

- Chickering, David Maxwell (Mar. 2002). 'Learning equivalence classes of bayesian-network structures'. In: *Journal of Machine Learning Research* 2, pp. 445–498. ISSN: 1532-4435. DOI: 10.1162/153244302760200696. URL: <https://doi.org/10.1162/153244302760200696>.
- Colombo, Diego and Marloes H. Maathuis (Jan. 2014). 'Order-Independent Constraint-Based Causal Structure Learning'. In: *Journal of Machine Learning Research* 15.1, pp. 3741–3782. ISSN: 1532-4435.
- Dai, Haoyue et al. (2023). 'MI4c: Seeing causality through latent vicinity'. In: *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*. SIAM, pp. 226–234.
- Harris, Naftali and Mathias Drton (Jan. 2013). 'PC algorithm for nonparanormal graphical models'. In: *Journal of Machine Learning Research* 14.1, pp. 3365–3383. ISSN: 1532-4435.
- Hung, H. M. James et al. (1997). 'The Behavior of the P-Value When the Alternative Hypothesis is True'. In: *Biometrics* 53.1, pp. 11–22. ISSN: 0006341X, 15410420. URL: <http://www.jstor.org/stable/2533093> (visited on 09/10/2023).
- Kalisch, Markus and Peter Bühlman (2007). 'Estimating high-dimensional directed acyclic graphs with the PC-algorithm.'. In: *Journal of Machine Learning Research* 8.22, pp. 613–636.

## References II

- Kyono, Trent, Zhang and Mihaela van der Schaar (2020). 'CASTLE: Regularization via Auxiliary Causal Graph Discovery'. In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. URL: <https://proceedings.neurips.cc/paper/2020/hash/1068bceb19323fe72b2b344ccf85c254-Abstract.html>.
- Meek, Christopher (1995). 'Causal Inference and Causal Explanation with Background Knowledge'. In: *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*. UAI 1995. Montréal, Qué, Canada: Morgan Kaufmann Publishers Inc., pp. 403–410. ISBN: 1558603859.
- Peters, Jonas and Peter Bühlmann (2015). 'Structural intervention distance for evaluating causal graphs'. In: *Neural computation* 27.3, pp. 771–799.
- Ramsey, Joseph (2016). *Improving accuracy and scalability of the pc algorithm by maximizing p-value*.
- Ramsey, Joseph, Madelyn Glymour et al. (2017). 'A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images'. In: *International journal of data science and analytics* 3, pp. 121–129.
- Ramsey, Joseph, Peter Spirtes and Jiji Zhang (2006). 'Adjacency-Faithfulness and Conservative Causal Inference'. In: *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*. UAI 2006. Cambridge, MA, USA: AUAI Press, pp. 401–408. ISBN: 0974903922.

## References III

- Russo, Fabrizio and Francesca Toni (2023). 'Shapley-PC: Constraint-based Causal Structure Learning with Shapley Values'. In: *CoRR* abs/2312.11582. DOI: 10.48550/ARXIV.2312.11582. arXiv: 2312.11582. URL: <https://doi.org/10.48550/arXiv.2312.11582>.
- Shapley, Lloyd S (1953). 'A value for n-person games (1953)'. In: *Contribution to the Theory of Games*.
- Spirtes, Peter, Glymour and Richard Scheines (2000). *Causation, prediction, and search*. MIT press.
- Zheng, Xun et al. (Nov. 2020). 'Learning Sparse Nonparametric DAGs'. In: *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. Vol. 108. Proceedings of Machine Learning Research. PMLR, pp. 3414–3425. URL: <https://proceedings.mlr.press/v108/zheng20a.html>.

## Appendices

# Contestable NNs

## Contested Network - Results

Table 1: Average AUC over 25 runs for the Adult dataset in four experimental scenarios.

Adult Dataset ( $ V  = 14$ )				
Data ( $N$ )	CASTLE $ E  = 210$	Injected $ E  = 46$	Partial $ E  = 116$	Contested $ E  = 30$
100	0.67 (0.03)	<b>0.69 .</b>	0.66	<b>0.69 .</b>
500	0.72 (0.04)	<b>0.74 *</b>	0.71	<b>0.74 *</b>
1000	0.75 (0.03)	<b>0.76</b>	0.74	<b>0.76</b>
2000	0.74 (0.03)	<b>0.77 ***</b>	<b>0.76 *</b>	<b>0.77 ***</b>
5000	0.75 (0.03)	<b>0.79 ***</b>	<b>0.76</b>	<b>0.79 ***</b>
10000	0.75 (0.02)	<b>0.85 ***</b>	<b>0.76 .</b>	<b>0.85 ***</b>
20000	0.76 (0.02)	<b>0.86 ***</b>	<b>0.77 .</b>	<b>0.86 ***</b>

Significance levels against CASTLE (Kyono, Zhang and van der Schaar 2020): 0 '\*\*\*\*' 0.001 '\*\*\*' 0.01 '\*\*' 0.05 '.' 0.1 '' 1.

# Shapley-PC

## PC-based Algorithms

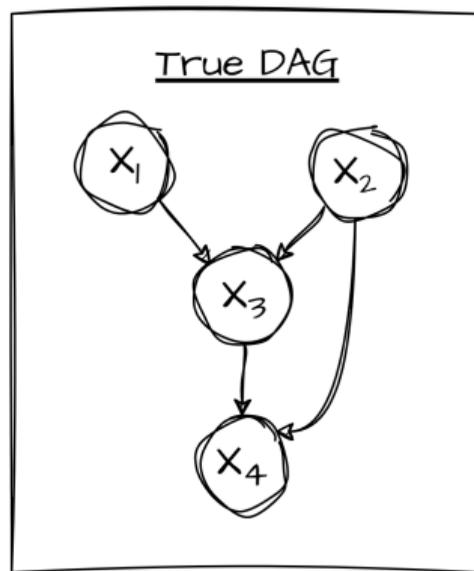
- ❶ Conservative-PC (Ramsey, Spirtes and J. Zhang 2006)
  - Orient v-structure only if the candidate collider renders dependence in **all** tests when in the conditioning set
- ❷ Majority-PC (Colombo and Maathuis 2014)
  - Orient v-structure only if the candidate collider renders dependence in the **majority** of the tests when in the conditioning set
- ❸ Max-PC (Ramsey 2016)
  - Orient v-structure only if the test with the **maximum p-value** does not contain the candidate collider
- ❹ ML4C (Dai et al. 2023)
  - Build a machine learning model to **predict** v-structures
- ❺ Shapley-PC (Russo and Toni 2023)
  - Orient v-structure only if the candidate collider has the **lowest Shapley Independence Value**

## PC-Algorithm (Propagation, Meek 1995)



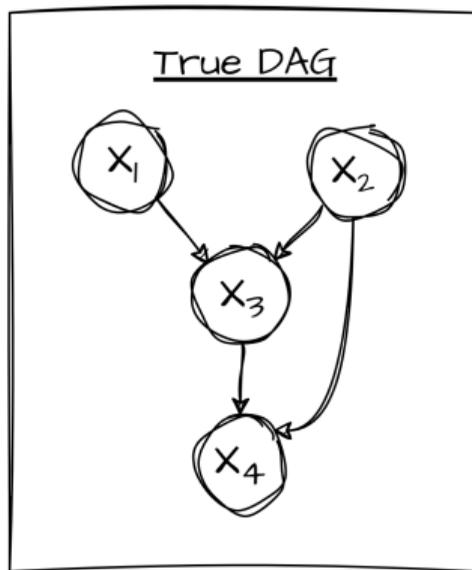
Figure 1: Orientation rules for patterns

## Example I (Perfect Independence Information)



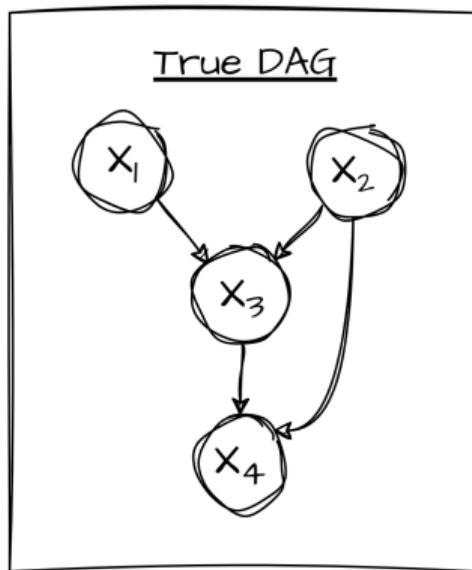
- Assess whether  $X_1 - X_3 - X_2$  is a v-structure
- Perfect Conditional Independence Information
  - 1  $I(X_1, X_2) = 1 \rightarrow X_1 \perp\!\!\!\perp X_2$
  - 2  $I(X_1, X_2 | X_3) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_3$
  - 3  $I(X_1, X_2 | X_4) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_4$
  - 4  $I(X_1, X_2 | \{X_3, X_4\}) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | \{X_3, X_4\}$

## Example I (Perfect Independence Information)



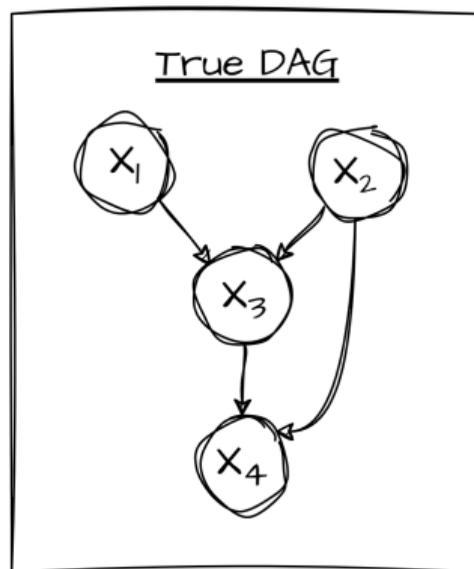
- Assess whether  $X_1 - X_3 - X_2$  is a v-structure
- Perfect Conditional Independence Information
  - ①  $I(X_1, X_2) = 1 \rightarrow X_1 \perp\!\!\!\perp X_2$
  - ②  $I(X_1, X_2 | X_3) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_3$
  - ③  $I(X_1, X_2 | X_4) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_4$
  - ④  $I(X_1, X_2 | \{X_3, X_4\}) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | \{X_3, X_4\}$
- $\phi_I(X_3, \{X_1, X_2\}) = \phi_I(X_4, \{X_1, X_2\}) = -0.5$

## Example I (Perfect Independence Information)



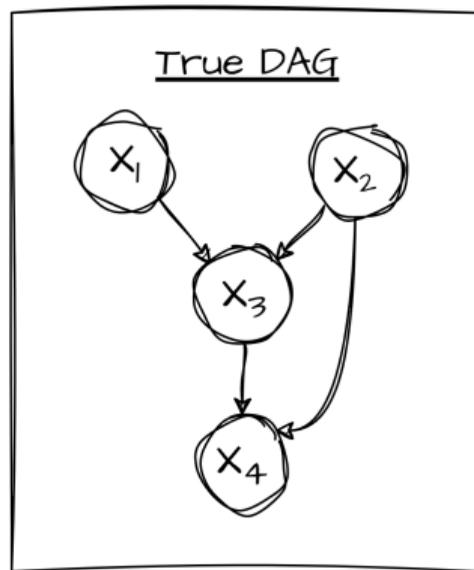
- Assess whether  $X_1 - X_3 - X_2$  is a v-structure
- Perfect Conditional Independence Information
  - 1  $I(X_1, X_2) = 1 \rightarrow X_1 \perp\!\!\!\perp X_2$
  - 2  $I(X_1, X_2 | X_3) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_3$
  - 3  $I(X_1, X_2 | X_4) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_4$
  - 4  $I(X_1, X_2 | \{X_3, X_4\}) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | \{X_3, X_4\}$
- $\phi_I(X_3, \{X_1, X_2\}) = \phi_I(X_4, \{X_1, X_2\}) = -0.5$
- All PC-based algorithms can recover the true DAG

## Example II (Imperfect Independence Information)



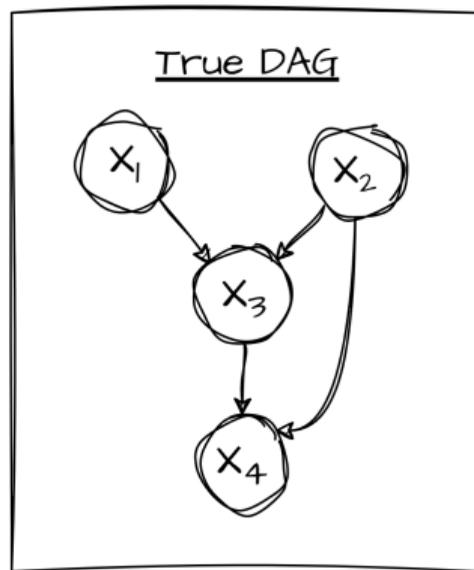
- Assess whether  $X_1 - X_3 - X_2$  is a v-structure
- Finite Sample Conditional Independence Tests
  - 1  $I(X_1, X_2) = 0.7 \geq \alpha \rightarrow X_1 \perp\!\!\!\perp X_2$
  - 2  $I(X_1, X_2 | X_3) = 0.01 < \alpha \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_3$
  - 3  $I(X_1, X_2 | X_4) = 0.1 \geq \alpha \rightarrow X_1 \perp\!\!\!\perp X_2 | X_4$
  - 4  $I(X_1, X_2 | \{X_3, X_4\}) = 0.75 \geq \alpha \rightarrow X_1 \perp\!\!\!\perp X_2 | \{X_3, X_4\}$

## Example II (Imperfect Independence Information)



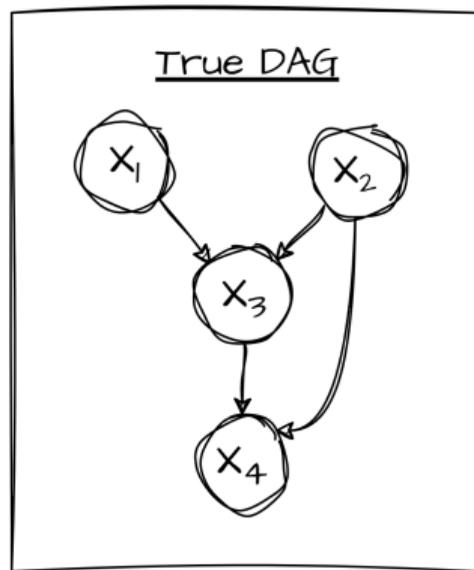
- Assess whether  $X_1 - X_3 - X_2$  is a v-structure
- Finite Sample Conditional Independence Tests
  - ①  $I(X_1, X_2) = 0.7 \geq \alpha \rightarrow X_1 \perp\!\!\!\perp X_2$
  - ②  $I(X_1, X_2 | X_3) = 0.01 < \alpha \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_3$
  - ③  $I(X_1, X_2 | X_4) = 0.1 \geq \alpha \rightarrow X_1 \perp\!\!\!\perp X_2 | X_4$
  - ④  $I(X_1, X_2 | \{X_3, X_4\}) = 0.75 \geq \alpha \rightarrow X_1 \perp\!\!\!\perp X_2 | \{X_3, X_4\}$
- $\phi_I(X_3, \{X_1, X_2\}) = -0.03, \phi_I(X_4, \{X_1, X_2\}) = 0.08$

## Example II (Imperfect Independence Information)



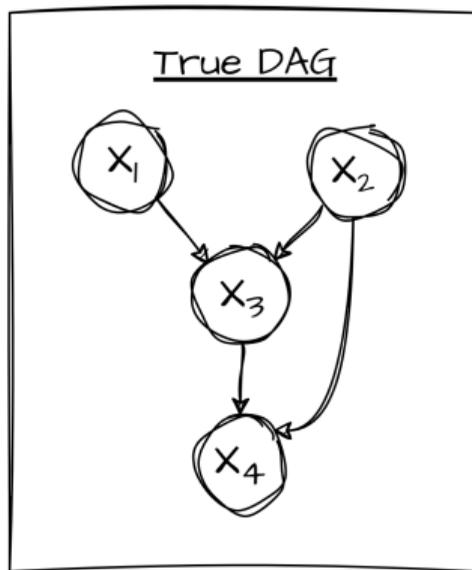
- Assess whether  $X_1 - X_3 - X_2$  is a v-structure
- Finite Sample Conditional Independence Tests
  - ①  $I(X_1, X_2) = 0.7 \geq \alpha \rightarrow X_1 \perp\!\!\!\perp X_2$
  - ②  $I(X_1, X_2 | X_3) = 0.01 < \alpha \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_3$
  - ③  $I(X_1, X_2 | X_4) = 0.1 \geq \alpha \rightarrow X_1 \perp\!\!\!\perp X_2 | X_4$
  - ④  $I(X_1, X_2 | \{X_3, X_4\}) = 0.75 \geq \alpha \rightarrow X_1 \perp\!\!\!\perp X_2 | \{X_3, X_4\}$
- $\phi_I(X_3, \{X_1, X_2\}) = -0.03$ ,  $\phi_I(X_4, \{X_1, X_2\}) = 0.08$
- Only Shapley-PC correctly orients the v-structure

## Example III (Little imperfect Independence Information)



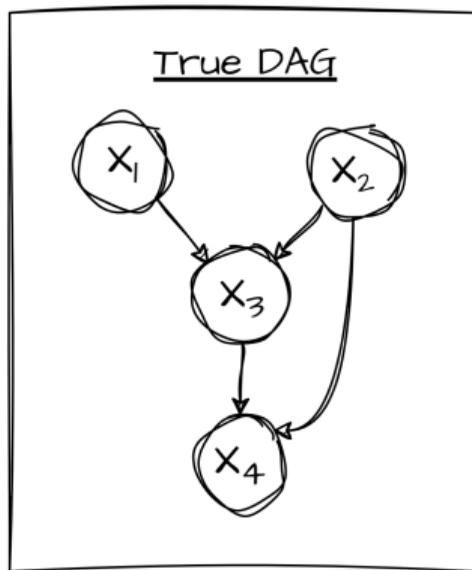
- Assess whether  $X_1 - X_3 - X_2$  is a v-structure
- Finite Sample Conditional Independence Tests
  - 1  $I(X_1, X_2) = 1 \rightarrow X_1 \perp\!\!\!\perp X_2$
  - 2  $I(X_1, X_2 | X_3) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_3$
  - 3  $I(X_1, X_2 | X_4) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_4$
  - 4  $I(X_1, X_2 | \{X_3, X_4\}) = 0.1 \rightarrow X_1 \perp\!\!\!\perp X_2 | \{X_3, X_4\}$

## Example III (Little imperfect Independence Information)



- Assess whether  $X_1 - X_3 - X_2$  is a v-structure
- Finite Sample Conditional Independence Tests
  - 1  $I(X_1, X_2) = 1 \rightarrow X_1 \perp\!\!\!\perp X_2$
  - 2  $I(X_1, X_2 | X_3) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_3$
  - 3  $I(X_1, X_2 | X_4) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_4$
  - 4  $I(X_1, X_2 | \{X_3, X_4\}) = 0.1 \rightarrow X_1 \perp\!\!\!\perp X_2 | \{X_3, X_4\}$
- $\phi_I(X_3, \{X_1, X_2\}) = \phi_I(X_4, \{X_1, X_2\}) = -0.45$

## Example III (Little imperfect Independence Information)



- Assess whether  $X_1 - X_3 - X_2$  is a v-structure
- Finite Sample Conditional Independence Tests
  - 1  $I(X_1, X_2) = 1 \rightarrow X_1 \perp\!\!\!\perp X_2$
  - 2  $I(X_1, X_2 | X_3) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_3$
  - 3  $I(X_1, X_2 | X_4) = 0 \rightarrow X_1 \not\perp\!\!\!\perp X_2 | X_4$
  - 4  $I(X_1, X_2 | \{X_3, X_4\}) = 0.1 \rightarrow X_1 \perp\!\!\!\perp X_2 | \{X_3, X_4\}$
- $\phi_I(X_3, \{X_1, X_2\}) = \phi_I(X_4, \{X_1, X_2\}) = -0.45$
- Only Shapley-PC and PC-max correctly orient the v-structure

## Soundness

### Definition (Perfect Conditional Independence Test (CIT))

$$I(X_i, X_j \mid \mathbf{S}) = \begin{cases} 1 & \text{iff } X_i \perp\!\!\!\perp X_j \mid \mathbf{S} \\ 0 & \text{otherwise} \end{cases}$$

### Lemma

Given a skeleton  $\mathcal{C}$ , a UT  $X_i - X_j - X_k \in \mathcal{C}$ ,  $X_i, X_j, X_k \in \mathbf{V}$ , and a perfect CIT  $I_\infty$ , the SIV of variable  $X_j$   $\phi_{I_\infty}(X_j, \{X_i, X_k\}) < 0$  if and only if  $X_j$  is a collider for  $X_i$  and  $X_k$ .

## Soundness

### Theorem (Correctness of Shapley-PC)

*Let  $\mathcal{P}(\mathbf{V})$  be a joint distribution faithful to a DAG  $\mathcal{G} = (\mathbf{V}, E)$ , and assume access to perfect conditional independence information for all pairs  $(X_i, X_j) \in \mathbf{V}$  given subsets  $\mathbf{S} \subseteq \mathbf{V} \setminus \{X_i, X_j\}$ . Then the output of Shapley-PC is the CPDAG representing the MEC of  $\mathcal{G}$ .*

## Asymptotic Consistency

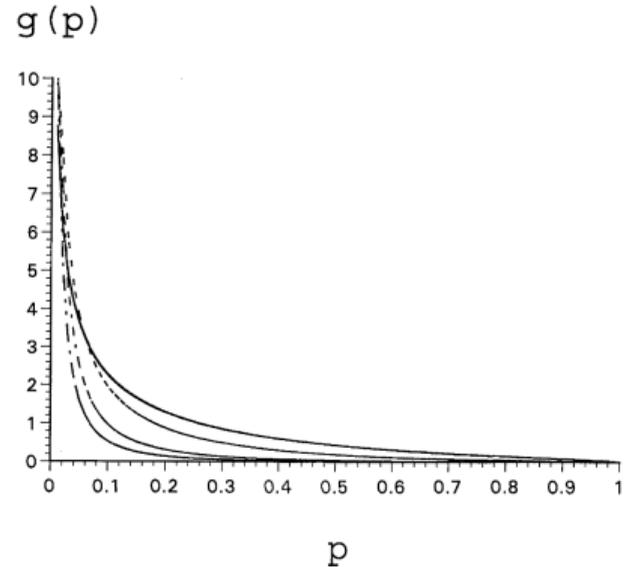
- In the sample limit, the original PC-algorithm has been shown to be consistent for sparse graphs and multivariate Gaussian distributions (Kalisch and Bühlman 2007) or Gaussian copulas (Harris and Drton 2013)
- The results are contingent on PC only performing CIT between pairs of variables, with the size of the conditioning sets  $\mathbf{S}$  less or equal to the degree of the graph
- Our proposed method has the same characteristic, hence the consistency results are equally applicable

## Complexity

- As in PC, the complexity of our algorithm depends on the number of vertices and their maximal degree (Spirtes, Glymour and Scheines 2000, p.85)
- As in CPC, MPC and PC-Max, we perform additional tests compared to the original PC that used the separating sets derived from the adjacency test in Step 1
- The number of tests though still depends on the degree as we only add the tests about the adjacency set to calculate  $\phi_I(X_j)$
- The majority of the testing is still done in the adjacency search of Step 1 of PC (Ramsey, Spirtes and J. Zhang 2006)

## Why the minimum? (Hung et al. 1997)

- Under the alternative hypothesis of dependence
  - $g(p)$  depends on the sample size and the value of the parameter in the alternative hypothesis
  - $g(p)$  decrease monotonically and concentrate around 0, the further towards 1 the lower is the probability of dependence

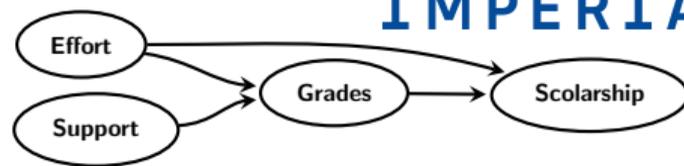


## Evaluation Metrics

- Structural Hamming Distance (SHD) is a purely graphical metric summing up the number of changes to be made to the estimated graph to match the true one (SHD = Extra + Missing + Reversed)
- Structural Intervention Distance (SID) (Peters and Bühlmann 2015) quantifies the closeness between two DAGs in terms of their corresponding causal inference statements

# Causal ABA

## Student's Achievement Example



$$Ef \perp\!\!\!\perp Sc \quad p = 0.00 \quad S = 1.00$$

$$Ef \perp\!\!\!\perp Gr \quad p = 0.00 \quad S = 1.00$$

$$Su \perp\!\!\!\perp Gr \quad p = 0.00 \quad S = 1.00$$

$$Gr \perp\!\!\!\perp Sc \quad p = 0.00 \quad S = 1.00$$

$$Su \perp\!\!\!\perp Ef \quad p = 0.46 \quad S = 0.71$$

$$Su \perp\!\!\!\perp Sc \quad p = 0.05 \quad S = 0.52$$

$$Ef \perp\!\!\!\perp Sc \mid \{Su\} \quad p = 0.00 \quad S = 0.50$$

$$Ef \perp\!\!\!\perp Sc \mid \{Gr\} \quad p = 0.00 \quad S = 0.50$$

$$Ef \perp\!\!\!\perp Gr \mid \{Su\} \quad p = 0.00 \quad S = 0.50$$

$$Su \perp\!\!\!\perp Gr \mid \{Sc\} \quad p = 0.00 \quad S = 0.50$$

$$Ef \perp\!\!\!\perp Gr \mid \{Sc\} \quad p = 0.00 \quad S = 0.50$$

$$Su \perp\!\!\!\perp Gr \mid \{Ef\} \quad p = 0.00 \quad S = 0.50$$

$$Gr \perp\!\!\!\perp Sc \mid \{Ef\} \quad p = 0.00 \quad S = 0.50$$

$$Gr \perp\!\!\!\perp Sc \mid \{Su\} \quad p = 0.00 \quad S = 0.50$$

$$Su \perp\!\!\!\perp Ef \mid \{Gr\} \quad p = 0.53 \quad S = 0.38$$

$$Su \perp\!\!\!\perp Sc \mid \{Gr\} \quad p = 0.03 \quad S = 0.35$$

$$Su \perp\!\!\!\perp Ef \mid \{Gr\} \quad p = 0.33 \quad S = 0.32$$

$$Su \perp\!\!\!\perp Ef \mid \{Sc\} \quad p = 0.05 \quad S = 0.25$$

$$Su \perp\!\!\!\perp Ef \mid \{Gr, Sc\} \quad p = 0.39 \quad S = 0.00$$

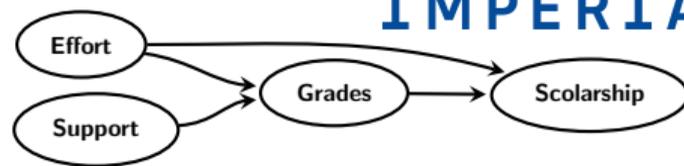
$$Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\} \quad p = 0.00 \quad S = 0.00$$

$$Ef \perp\!\!\!\perp Gr \mid \{Su, Sc\} \quad p = 0.00 \quad S = 0.00$$

$$Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\} \quad p = 0.00 \quad S = 0.00$$

$$Su \perp\!\!\!\perp Gr \mid \{Ef, Sc\} \quad p = 0.03 \quad S = 0.00$$

## Student's Achievement Example



$$Ef \perp\!\!\!\perp Sc \quad p = 0.00 \quad S = 1.00$$

$$Ef \perp\!\!\!\perp Gr \quad p = 0.00 \quad S = 1.00$$

$$Su \perp\!\!\!\perp Gr \quad p = 0.00 \quad S = 1.00$$

$$Gr \perp\!\!\!\perp Sc \quad p = 0.00 \quad S = 1.00$$

$$Su \perp\!\!\!\perp Ef \quad p = 0.46 \quad S = 0.71$$

$$Su \perp\!\!\!\perp Sc \quad p = 0.05 \quad S = 0.52$$

$$Ef \perp\!\!\!\perp Sc \mid \{Su\} \quad p = 0.00 \quad S = 0.50$$

$$Ef \perp\!\!\!\perp Sc \mid \{Gr\} \quad p = 0.00 \quad S = 0.50$$

$$Ef \perp\!\!\!\perp Gr \mid \{Su\} \quad p = 0.00 \quad S = 0.50$$

$$Su \perp\!\!\!\perp Gr \mid \{Sc\} \quad p = 0.00 \quad S = 0.50$$

$$Ef \perp\!\!\!\perp Gr \mid \{Sc\} \quad p = 0.00 \quad S = 0.50$$

$$Su \perp\!\!\!\perp Gr \mid \{Ef\} \quad p = 0.00 \quad S = 0.50$$

$$Gr \perp\!\!\!\perp Sc \mid \{Ef\} \quad p = 0.00 \quad S = 0.50$$

$$Gr \perp\!\!\!\perp Sc \mid \{Su\} \quad p = 0.00 \quad S = 0.50$$

$$Su \perp\!\!\!\perp Ef \mid \{Gr\} \quad p = 0.53 \quad S = 0.38$$

$$Su \perp\!\!\!\perp Sc \mid \{Gr\} \quad p = 0.03 \quad S = 0.35$$

$$Su \perp\!\!\!\perp Ef \mid \{Gr\} \quad p = 0.33 \quad S = 0.32$$

$$Su \perp\!\!\!\perp Ef \mid \{Sc\} \quad p = 0.05 \quad S = 0.25$$

$$Su \perp\!\!\!\perp Ef \mid \{Gr, Sc\} \quad p = 0.39 \quad S = 0.00$$

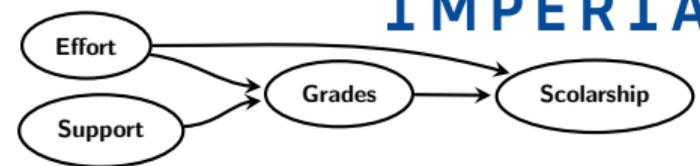
$$Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\} \quad p = 0.00 \quad S = 0.00$$

$$Ef \perp\!\!\!\perp Gr \mid \{Su, Sc\} \quad p = 0.00 \quad S = 0.00$$

$$Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\} \quad p = 0.00 \quad S = 0.00$$

$$Su \perp\!\!\!\perp Gr \mid \{Ef, Sc\} \quad p = 0.03 \quad S = 0.00$$

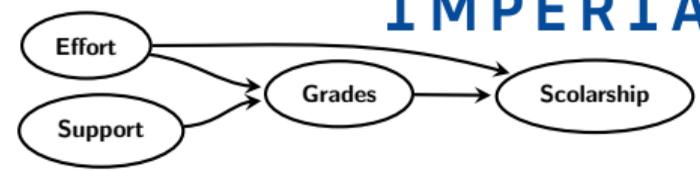
## Student's Achievement Example



$Ef \perp\!\!\!\perp Sc$	$p = 0.00$	$S = 1.00$
$Ef \perp\!\!\!\perp Gr$	$p = 0.00$	$S = 1.00$
$Su \perp\!\!\!\perp Gr$	$p = 0.00$	$S = 1.00$
$Gr \perp\!\!\!\perp Sc$	$p = 0.00$	$S = 1.00$
$Su \perp\!\!\!\perp Ef$	$p = 0.46$	$S = 0.71$
$Su \perp\!\!\!\perp Sc$	$p = 0.05$	$S = 0.52$
$Ef \perp\!\!\!\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Ef \perp\!\!\!\perp Sc \mid \{Gr\}$	$p = 0.00$	$S = 0.50$
$Ef \perp\!\!\!\perp Gr \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Ef \perp\!\!\!\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Gr \mid \{Ef\}$	$p = 0.00$	$S = 0.50$

$Gr \perp\!\!\!\perp Sc \mid \{Ef\}$	$p = 0.00$	$S = 0.50$
$Gr \perp\!\!\!\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Ef \mid \{Gr\}$	$p = 0.53$	$S = 0.38$
$Su \perp\!\!\!\perp Sc \mid \{Gr\}$	$p = 0.03$	$S = 0.35$
$Su \perp\!\!\!\perp Ef \mid \{Gr\}$	$p = 0.33$	$S = 0.32$
$Su \perp\!\!\!\perp Ef \mid \{Sc\}$	$p = 0.05$	$S = 0.25$
<del><math>Su \perp\!\!\!\perp Ef \mid \{Gr, Sc\}</math></del>	<del><math>p = 0.39</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Ef \perp\!\!\!\perp Gr \mid \{Su, Sc\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \perp\!\!\!\perp Gr \mid \{Ef, Sc\}</math></del>	<del><math>p = 0.03</math></del>	<del><math>S = 0.00</math></del>

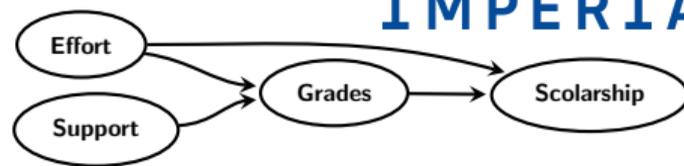
## Student's Achievement Example



$Ef \perp\!\!\!\perp Sc$	$p = 0.00$	$S = 1.00$
$Ef \perp\!\!\!\perp Gr$	$p = 0.00$	$S = 1.00$
$Su \perp\!\!\!\perp Gr$	$p = 0.00$	$S = 1.00$
$Gr \perp\!\!\!\perp Sc$	$p = 0.00$	$S = 1.00$
$Su \perp\!\!\!\perp Ef$	$p = 0.46$	$S = 0.71$
$Su \perp\!\!\!\perp Sc$	$p = 0.05$	$S = 0.52$
$Ef \perp\!\!\!\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Ef \perp\!\!\!\perp Sc \mid \{Gr\}$	$p = 0.00$	$S = 0.50$
$Ef \perp\!\!\!\perp Gr \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Ef \perp\!\!\!\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Gr \mid \{Ef\}$	$p = 0.00$	$S = 0.50$

$Gr \perp\!\!\!\perp Sc \mid \{Ef\}$	$p = 0.00$	$S = 0.50$
$Gr \perp\!\!\!\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Ef \mid \{Gr\}$	$p = 0.53$	$S = 0.38$
$Su \perp\!\!\!\perp Sc \mid \{Gr\}$	$p = 0.03$	$S = 0.35$
$Su \perp\!\!\!\perp Ef \mid \{Gr\}$	$p = 0.33$	$S = 0.32$
<del><math>Su \perp\!\!\!\perp Ef \mid \{Sc\}</math></del>	<del><math>p = 0.05</math></del>	<del><math>S = 0.25</math></del>
<del><math>Su \perp\!\!\!\perp Ef \mid \{Gr, Sc\}</math></del>	<del><math>p = 0.39</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Ef \perp\!\!\!\perp Gr \mid \{Su, Sc\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \perp\!\!\!\perp Gr \mid \{Ef, Sc\}</math></del>	<del><math>p = 0.03</math></del>	<del><math>S = 0.00</math></del>

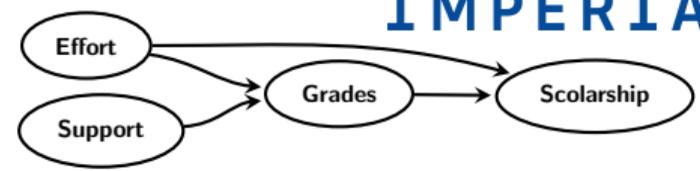
## Student's Achievement Example



$Ef \perp\!\!\!\perp Sc$	$p = 0.00$	$S = 1.00$
$Ef \perp\!\!\!\perp Gr$	$p = 0.00$	$S = 1.00$
$Su \perp\!\!\!\perp Gr$	$p = 0.00$	$S = 1.00$
$Gr \perp\!\!\!\perp Sc$	$p = 0.00$	$S = 1.00$
$Su \perp\!\!\!\perp Ef$	$p = 0.46$	$S = 0.71$
$Su \perp\!\!\!\perp Sc$	$p = 0.05$	$S = 0.52$
$Ef \perp\!\!\!\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Ef \perp\!\!\!\perp Sc \mid \{Gr\}$	$p = 0.00$	$S = 0.50$
$Ef \perp\!\!\!\perp Gr \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Ef \perp\!\!\!\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Gr \mid \{Ef\}$	$p = 0.00$	$S = 0.50$

$Gr \perp\!\!\!\perp Sc \mid \{Ef\}$	$p = 0.00$	$S = 0.50$
$Gr \perp\!\!\!\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Ef \mid \{Gr\}$	$p = 0.53$	$S = 0.38$
$Su \perp\!\!\!\perp Sc \mid \{Gr\}$	$p = 0.03$	$S = 0.35$
<del><math>Su \perp\!\!\!\perp Ef \mid \{Gr\}</math></del>	<del><math>p = 0.33</math></del>	<del><math>S = 0.32</math></del>
<del><math>Su \perp\!\!\!\perp Ef \mid \{Sc\}</math></del>	<del><math>p = 0.05</math></del>	<del><math>S = 0.25</math></del>
<del><math>Su \perp\!\!\!\perp Ef \mid \{Gr, Sc\}</math></del>	<del><math>p = 0.39</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Ef \perp\!\!\!\perp Gr \mid \{Su, Sc\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \perp\!\!\!\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \perp\!\!\!\perp Gr \mid \{Ef, Sc\}</math></del>	<del><math>p = 0.03</math></del>	<del><math>S = 0.00</math></del>

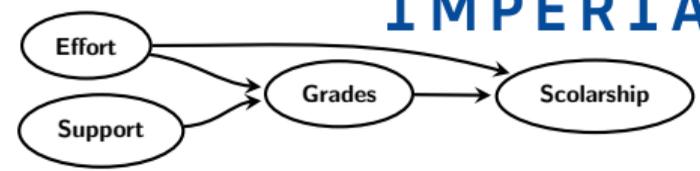
## Student's Achievement Example



$Ef \not\perp Sc$	$p = 0.00$	$S = 1.00$
$Ef \not\perp Gr$	$p = 0.00$	$S = 1.00$
$Su \not\perp Gr$	$p = 0.00$	$S = 1.00$
$Gr \not\perp Sc$	$p = 0.00$	$S = 1.00$
$Su \perp\!\!\!\perp Ef$	$p = 0.46$	$S = 0.71$
$Su \not\perp Sc$	$p = 0.05$	$S = 0.52$
$Ef \not\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Ef \not\perp Sc \mid \{Gr\}$	$p = 0.00$	$S = 0.50$
$Ef \not\perp Gr \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Su \not\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Ef \not\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Su \not\perp Gr \mid \{Ef\}$	$p = 0.00$	$S = 0.50$

$Gr \not\perp Sc \mid \{Ef\}$	$p = 0.00$	$S = 0.50$
$Gr \not\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Su \perp\!\!\!\perp Ef \mid \{Gr\}$	$p = 0.53$	$S = 0.38$
<del><math>Su \not\perp Sc \mid \{Gr\}</math></del>	<del><math>p = 0.03</math></del>	<del><math>S = 0.35</math></del>
<del><math>Su \perp\!\!\!\perp Ef \mid \{Gr\}</math></del>	<del><math>p = 0.33</math></del>	<del><math>S = 0.32</math></del>
<del><math>Su \perp\!\!\!\perp Ef \mid \{Sc\}</math></del>	<del><math>p = 0.05</math></del>	<del><math>S = 0.25</math></del>
<del><math>Su \perp\!\!\!\perp Ef \mid \{Gr, Sc\}</math></del>	<del><math>p = 0.39</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \not\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Ef \not\perp Gr \mid \{Su, Sc\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \not\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \not\perp Gr \mid \{Ef, Sc\}</math></del>	<del><math>p = 0.03</math></del>	<del><math>S = 0.00</math></del>

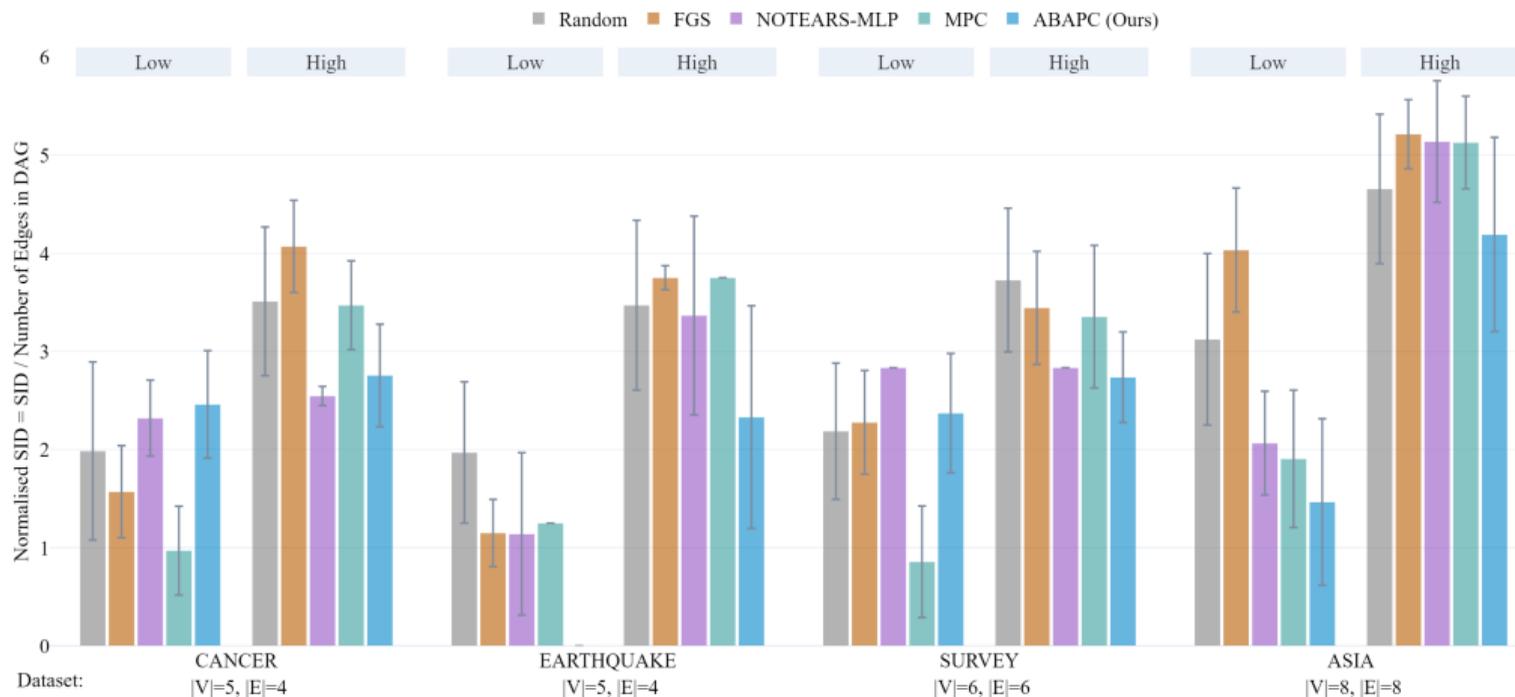
## Student's Achievement Example



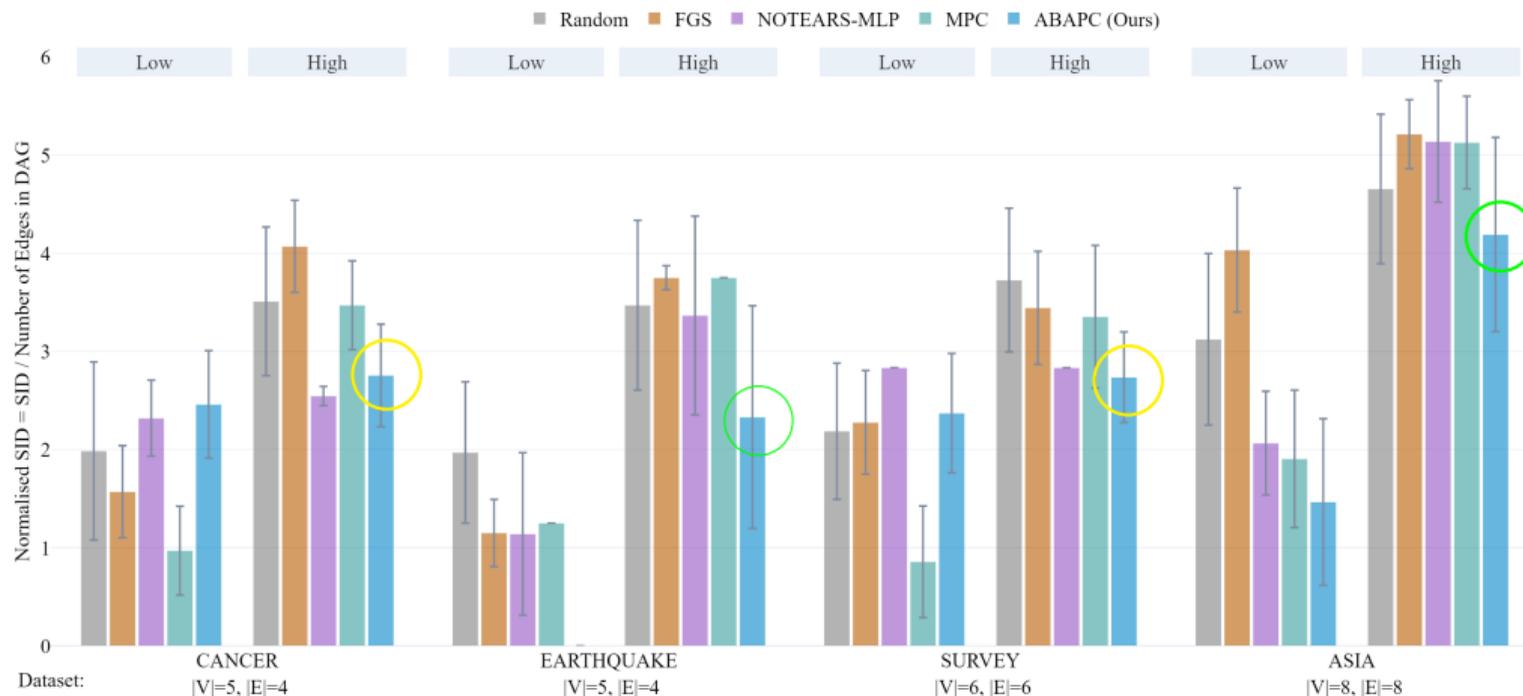
$Ef \not\perp Sc$	$p = 0.00$	$S = 1.00$
$Ef \not\perp Gr$	$p = 0.00$	$S = 1.00$
$Su \not\perp Gr$	$p = 0.00$	$S = 1.00$
$Gr \not\perp Sc$	$p = 0.00$	$S = 1.00$
$Su \perp\!\!\!\perp Ef$	$p = 0.46$	$S = 0.71$
$Su \not\perp Sc$	$p = 0.05$	$S = 0.52$
$Ef \not\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Ef \not\perp Sc \mid \{Gr\}$	$p = 0.00$	$S = 0.50$
$Ef \not\perp Gr \mid \{Su\}$	$p = 0.00$	$S = 0.50$
$Su \not\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Ef \not\perp Gr \mid \{Sc\}$	$p = 0.00$	$S = 0.50$
$Su \not\perp Gr \mid \{Ef\}$	$p = 0.00$	$S = 0.50$

$Gr \not\perp Sc \mid \{Ef\}$	$p = 0.00$	$S = 0.50$
$Gr \not\perp Sc \mid \{Su\}$	$p = 0.00$	$S = 0.50$
<del><math>Su \perp\!\!\!\perp Ef \mid \{Gr\}</math></del>	<del><math>p = 0.53</math></del>	<del><math>S = 0.38</math></del>
<del><math>Su \not\perp Sc \mid \{Gr\}</math></del>	<del><math>p = 0.03</math></del>	<del><math>S = 0.35</math></del>
<del><math>Su \perp\!\!\!\perp Ef \mid \{Gr\}</math></del>	<del><math>p = 0.33</math></del>	<del><math>S = 0.32</math></del>
<del><math>Su \perp\!\!\!\perp Ef \mid \{Sc\}</math></del>	<del><math>p = 0.05</math></del>	<del><math>S = 0.25</math></del>
<del><math>Su \perp\!\!\!\perp Ef \mid \{Gr, Sc\}</math></del>	<del><math>p = 0.39</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \not\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Ef \not\perp Gr \mid \{Su, Sc\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \not\perp Sc \mid \{Ef, Gr\}</math></del>	<del><math>p = 0.00</math></del>	<del><math>S = 0.00</math></del>
<del><math>Su \not\perp Gr \mid \{Ef, Sc\}</math></del>	<del><math>p = 0.03</math></del>	<del><math>S = 0.00</math></del>

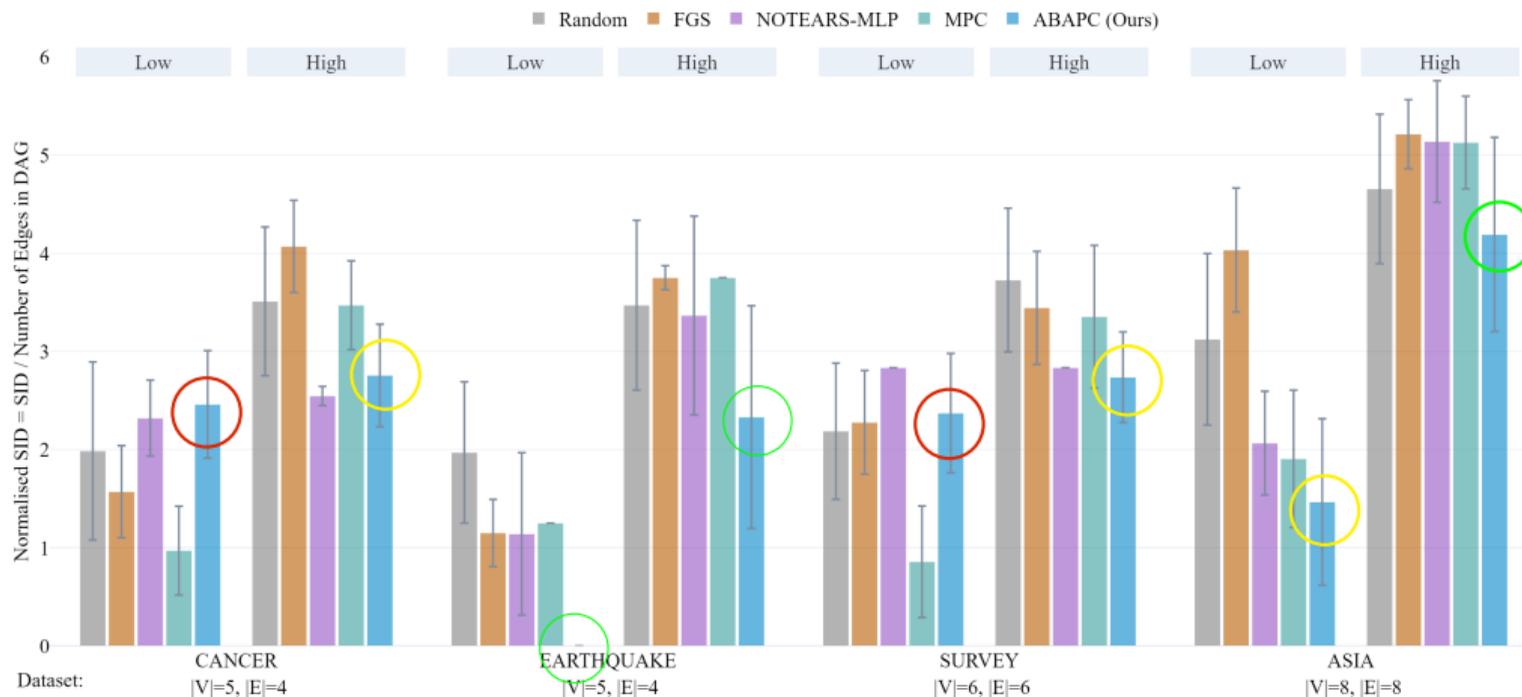
## Structural Interventional Distance (Peters and Bühlmann 2015)



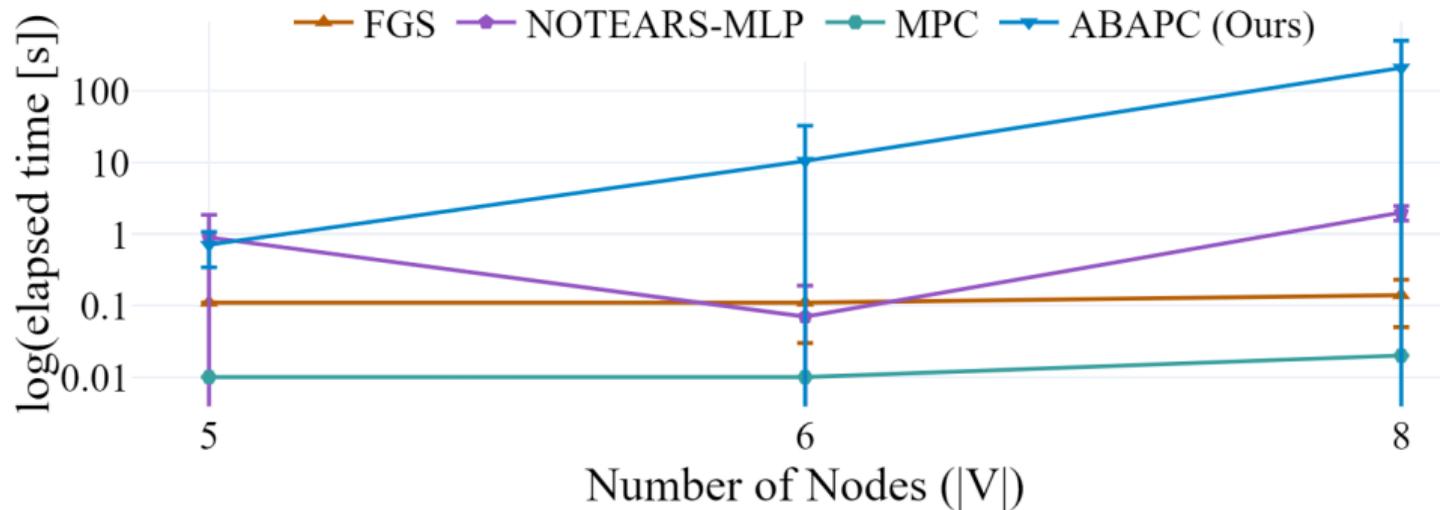
## Structural Interventional Distance (Peters and Bühlmann 2015)



## Structural Interventional Distance (Peters and Bühlmann 2015)



## Run Time Comparison



## Baselines

- 1 Majority-PC (Colombo and Maathuis 2014) is a modification of the PC-algorithm
- 2 Fast Greedy Equivalence Search (FGS) (Ramsey, M. Glymour et al. 2017) is a fast implementation of GES (Chickering 2002) where graphs are evaluated using the Bayesian Information Criterion
- 3 NOTEARS-MLP (Zheng et al. 2020) learns a non-linear SEM via continuous optimisation