Verification and search algorithms for causal DAGs

Davin Choo[1]*, Kirankumar Shiragur[2]*, Arnab Bhattacharyya[1]

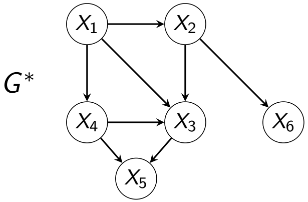
[1]National University of Singapore

[2]Stanford University

* Equal contribution

Motivation

Underlying data
generation process
(modelled as a DAG)



e.g. $X_4 = f_4(X_1, \varepsilon_4)$

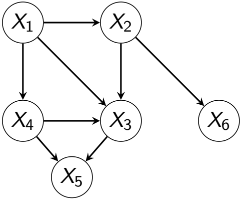
specific to node X_4

Motivation

Underlying data
generation process
(modelled as a DAG)

Observational data $\mathcal{D}$

G^*



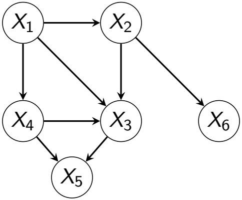
	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.3	0.4	0.1	-0.5	0.2	-0.3
Sample 2	0.1	1.2	0.6	-0.2	-0.1	-0.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮

e.g. $X_4 = f_4(X_1, \varepsilon_4)$

specific to node X_4

Motivation

G^*

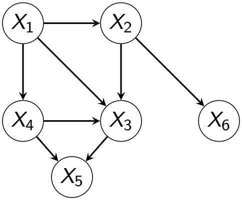


$\mathcal{D}$

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.3	0.4	0.1	-0.5	0.2	-0.3
Sample 2	0.1	1.2	0.6	-0.2	-0.1	-0.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Motivation

G^*

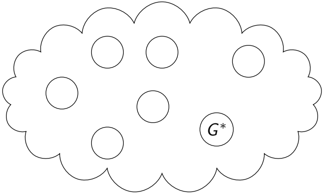


$\mathcal{D}$

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.3	0.4	0.1	-0.5	0.2	-0.3
Sample 2	0.1	1.2	0.6	-0.2	-0.1	-0.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮

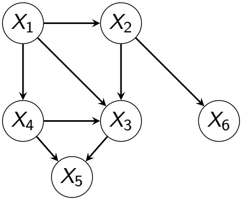
Identify possible graphs
via conditional dependencies
(e.g. PC [SGSH00], GES [Chi02])

Equivalence
class of DAGs



Motivation

G^*

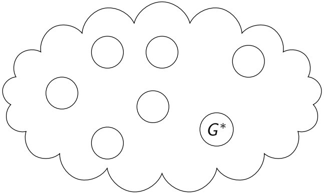


$\mathcal{D}$

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.3	0.4	0.1	-0.5	0.2	-0.3
Sample 2	0.1	1.2	0.6	-0.2	-0.1	-0.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Identify possible graphs
via conditional dependencies
(e.g. PC [SGSH00], GES [Chi02])

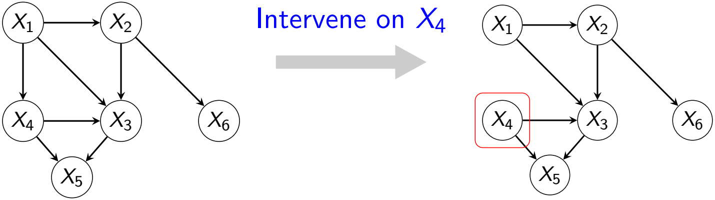
Equivalence
class of DAGs



Which is G^*?

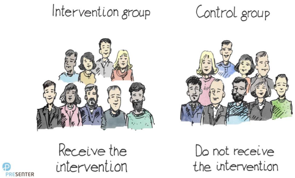
Two ways forward

1. Make model assumptions
   e.g. $X_4 = f_4(X_1, \varepsilon_4) = \alpha X_1 + \varepsilon_4$, where ε_4 is non-Gaussian
2. Perform interventions (Our focus)
   e.g. set $X_4 = 0.5$, then draw samples from the resulting intervened causal graph

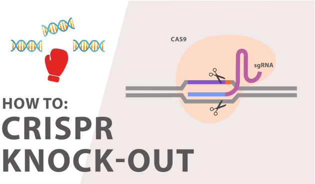


Interventions in real-life

- Randomized controlled trials



- Gene knockout experiments

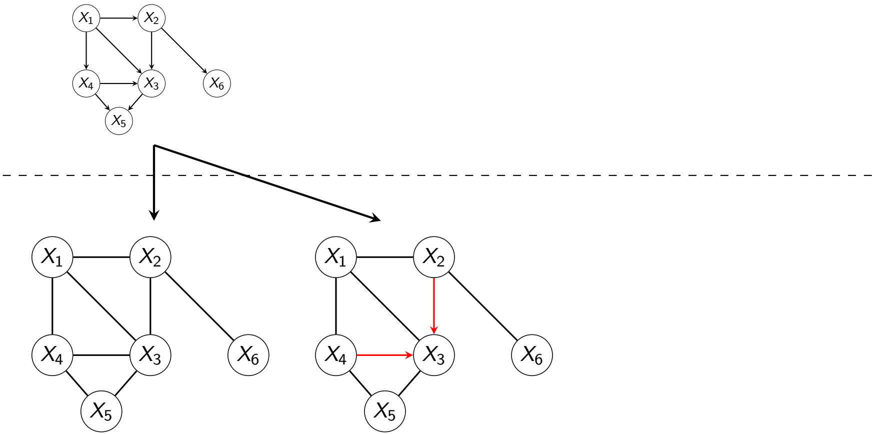


Can be expensive to perform ⇒ Minimize number of interventions!

Image credit: https://i.ytimg.com/vi/vyVyg4Bf2JY/maxresdefault.jpg
https://i.ytimg.com/vi/1BSiFfy0bTY/maxresdefault.jpg

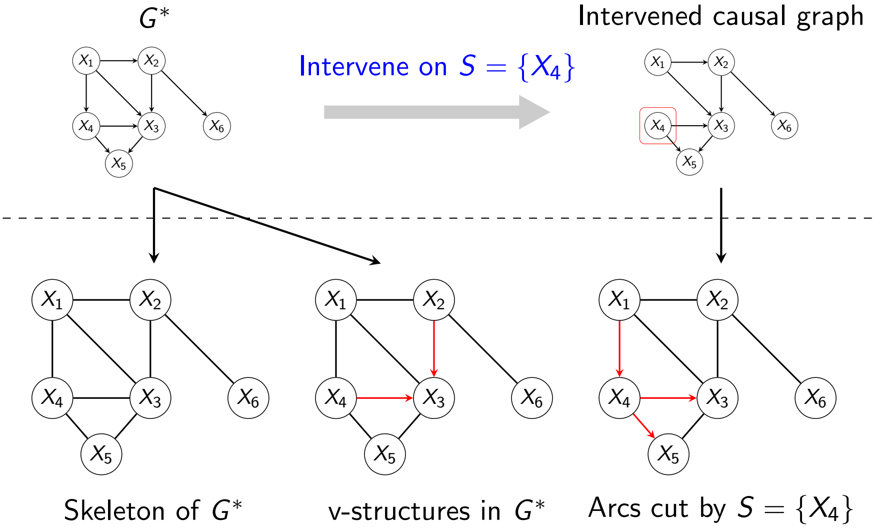
What can we learn?

G^*



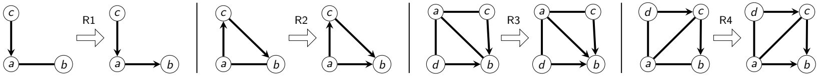
Skeleton of G^* v-structures in G^*

What can we learn?



Meek rules

Meek rules [Mee95]:
A set of 4 arc orientation rules that are *sound* and *complete*
(with respect to arc orientations with acyclic completion)



If $b \leftarrow a$,
then v-structure

If $b \leftarrow a$,
then cycle

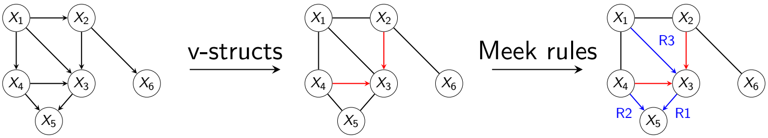
If $b \leftarrow a$, then the unoriented arcs would have been *oriented in the same way in all DAGs within the equivalence class* (via R2)
(See next slide on essential graphs)

Meek rules converge in polynomial time [WBL21, Algorithm 2].

Essential graph

We can represent an equivalence class with a partially oriented DAG called the *essential graph*

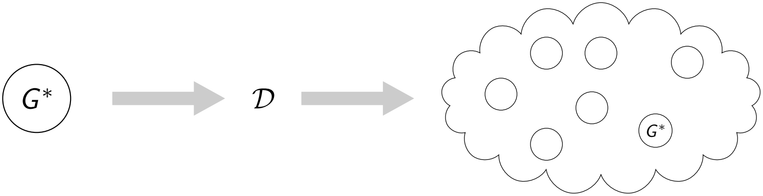
- Orient $u \to v$ if *all* DAGs agree on this direction
- An unoriented arc if there are two distinct DAGs G_1 and G_2 in the equivalence class orient it differently



G^* Essential graph of G^*

Problem setup

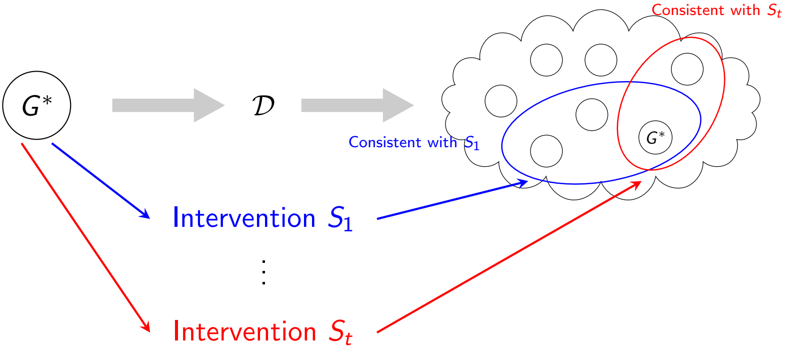
Identify G^*



Can be represented by
an essential graph
(partially oriented DAG)

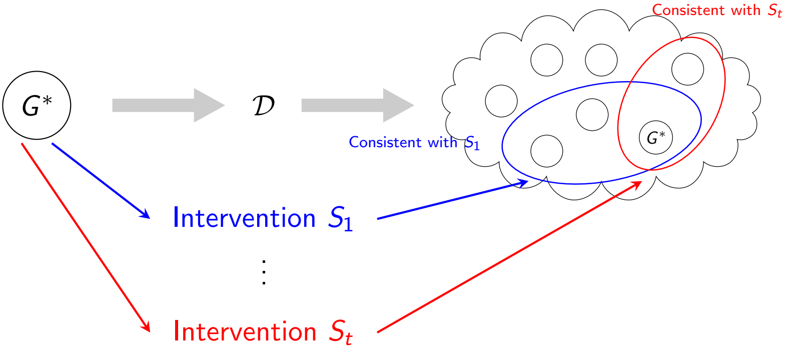
Problem setup (using interventions)

Identify G^* using as few interventions as possible (minimize t)



Problem setup (using **atomic** interventions)

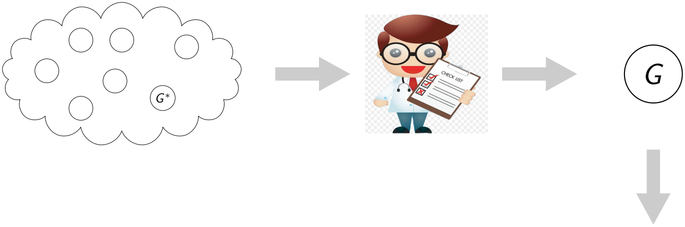
Identify G^* using as few interventions as possible (minimize t)



Simplifying assumption for this talk:
Each intervention is on a single node, i.e. $|S_1| = \ldots = |S_t| = 1$

Wait a minute... we have domain experts!

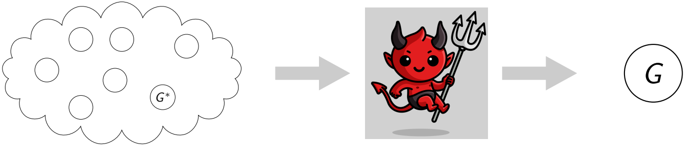
Problem solved with zero interventions!



Do stuff with
discovered causal graph G

Wait a minute... we have domain experts!

Problem solved with zero interventions!

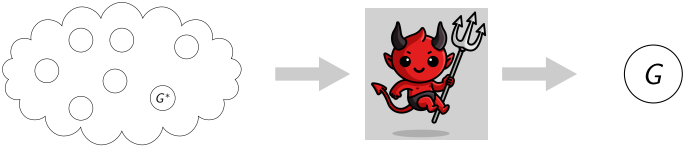


Experts can be wrong...

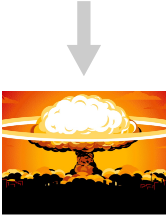
Do stuff with
discovered causal graph G

Wait a minute... we have domain experts!

~~Problem solved with zero interventions!~~



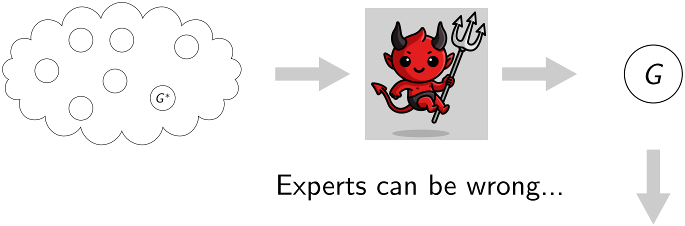
Experts can be wrong...



Wait a minute... we have domain experts!

How do we even check if $G = G^*$?

~~Problem solved with zero interventions!~~

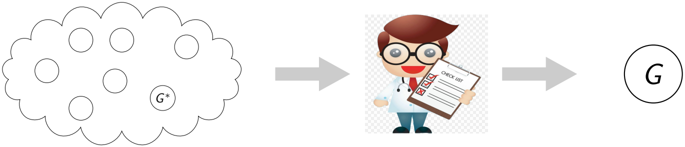


Experts can be wrong...

Image credit:
https://dribbble.com/shots/14489872-Devil
https://dribbble.com/shots/3759014-Atomic-Illustrations/attachments/3759014-Atomic-Illustrations?mode=media
https://img.favpng.com/23/12/11/questionnaire-survey-methodology-png-favpng-CW1Hb5zY6b47rPbAnvWgwEHPK.jpg

The verification problem

Goal: Determine if $G = G^*$
$\nu(G)$ = minimum number of interventions to answer $G \stackrel{?}{=} G^*$

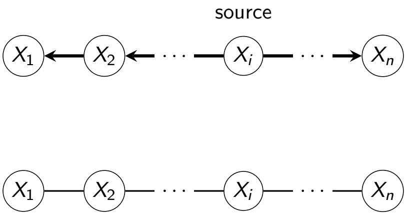


- We know: Intervening on v orients all arcs incident to v
- Trivial solution: Compute minimum vertex cover (MVC) on unoriented arcs! i.e. $\nu(G) \leq \texttt{MVC}(\text{unoriented})$

Verification: Problem with trivial solution

G

Essential graph

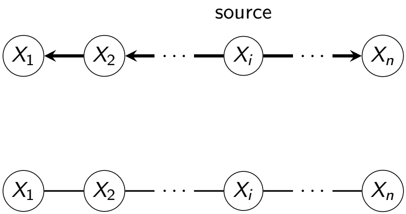


- Any line graph with a single source ($i \in \{1, \ldots, n\}$) has the same essential graph above of an unoriented line
- $\texttt{MVC}$(unoriented in essential graph) $= \left\lfloor \frac{n}{2} \right\rfloor$

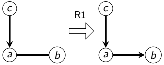
Verification: Problem with trivial solution

G

Essential graph



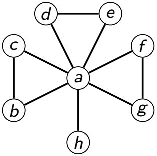
- Any line graph with a single source ($i \in \{1, \ldots, n\}$) has the same essential graph above of an unoriented line
- $\texttt{MVC}$(unoriented in essential graph) $= \left\lfloor \frac{n}{2} \right\rfloor$
- Optimal: Just 1 intervention needed!
  - Intervene on $X_i \Rightarrow$ Orient $X_{i-1} \leftarrow X_i$ and $X_i \rightarrow X_{i+1}$
  - Apply Meek R1 to orient the rest
  - If not fully oriented or $\exists$ disagreeing arcs, then $G \neq G^*$



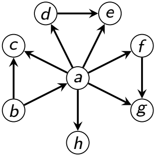
Verification: What was known

(Simplifying assumption: single undirected connected component)

1. [SMG+20]: $\nu(G) \geq \left\lfloor \frac{\omega(G)}{2} \right\rfloor$
2. [PSS22]: $\left\lceil \frac{n-r}{2} \right\rceil \leq \nu(G) \leq n - r$ (Note: 2-apx gap)



Essential graph

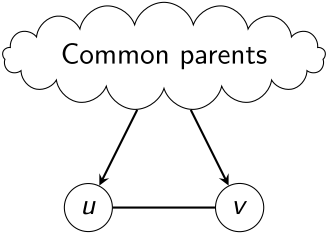


One possible DAG

- $n = 8$ nodes; $r = 4$ maximal cliques; largest clique $\omega(G) = 3$
- [SMG+20]: $\nu(G) \geq 1$; [PSS22]: $2 \leq \nu(G) \leq 4$
- Can we do better?

Verification: A complete characterization via covered edges

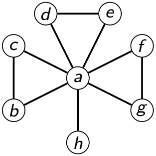
- Meek rules $\Rightarrow$ Outperform $\texttt{MVC}$(unoriented)
- Surprisingly, enough to compute $\texttt{MVC}$ on a *subset of edges*
- Covered edges [Chi95]:
  $u \sim v$ is covered edge $\iff$ $\text{Pa}(u) \setminus \{v\} = \text{Pa}(v) \setminus \{u\}$



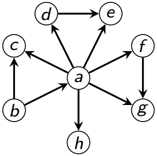
Claim: Necessary and sufficient to intervene on $\texttt{MVC}$(covered)

Proof: Simple (but subtle) using the notion of covered edges

Verification: Comparing to prior work



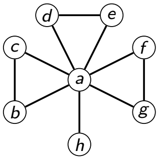
Essential graph



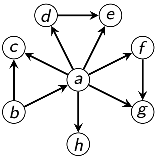
DAG

- $\omega(G) = 3,\ n = 8,\ r = 4 \Rightarrow \nu(G) \geq 1;\ 2 \leq \nu(G) \leq 4$
- Can we do better?

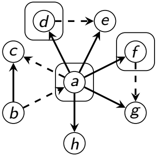
Verification: Comparing to prior work



Essential graph



DAG



Covered edges

- $\omega(G) = 3,\ n = 8,\ r = 4 \Rightarrow \nu(G) \geq 1;\ 2 \leq \nu(G) \leq 4$
- Can we do better? **Yes**
- We get *exact* $\nu(G)$ for each G in the equivalence class
- In fact, every DAG in this equivalence class needs 3 or 4 interventions, so the existing bounds on $\nu(G)$ are *not tight*.

Verification: Efficient computation

In general, MVC is NP-hard and we can only get a 2-approximation in polynomial time... [PSS22] also has a 2-approximation to $\nu(G)$.

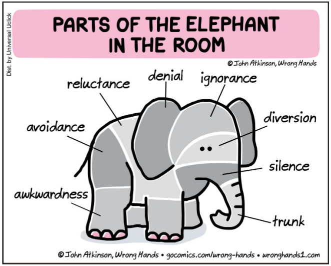
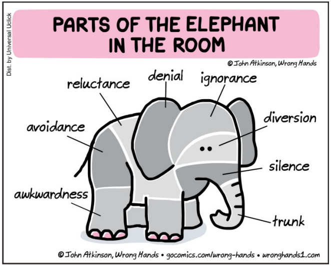


Image credit: https://wronghands1.com/2018/01/12/parts-of-the-elephant-in-the-room/

Verification: Efficient computation

In general, MVC is NP-hard and we can only get a 2-approximation in polynomial time... [PSS22] also has a 2-approximation to $\nu(G)$.



Claim: Covered edges form a forest.

Implication: MVC(covered) can be computed *exactly* in *linear time*.

Image credit: https://wronghands1.com/2018/01/12/parts-of-the-elephant-in-the-room/

Easy re-interpretation of known facts via covered edges

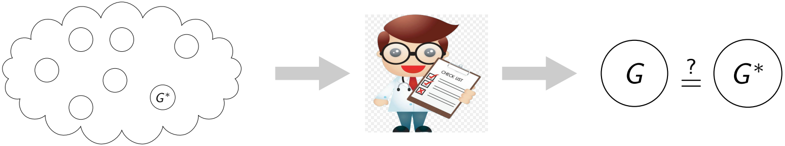
- Covered edges of clique K_n: $v_1 \to v_2, \ldots, v_{n-1} \to v_n$
- Covered edges of a tree: incident edges to root vertex
- Necessity of separating system for non-adaptive interventions
  - [Chi95]: Two graphs are equivalent $\iff$ there is a sequence of covered edge reversals to transform between them.
  - Unoriented edge $\Rightarrow$ Covered edge for *some* DAG in eq. class.
  - Conclusion: any *non-adaptive* search must cut *all* edges.
- Covered edge *cannot* have both endpoints as a sink of any maximal clique $\Rightarrow \nu(G) \leq n - r$ (result of [PSS22]).

(Slide catering to domain experts. If interested, pause to read; Else, skip)

The verification problem ✓

Can determine $G \stackrel{?}{=} G^*$

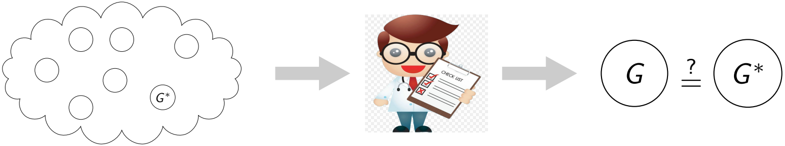
- Using $\nu(G) = \texttt{MVC}(\text{covered})$ interventions
- Computable in polynomial time



The verification problem ✓

Can determine $G \stackrel{?}{=} G^*$

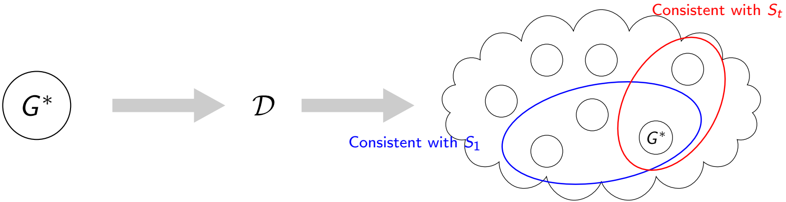
- Using $\nu(G) = \texttt{MVC}(\text{covered})$ interventions
- Computable in polynomial time



What about actually searching for G^* without the expert?

The adaptive search problem

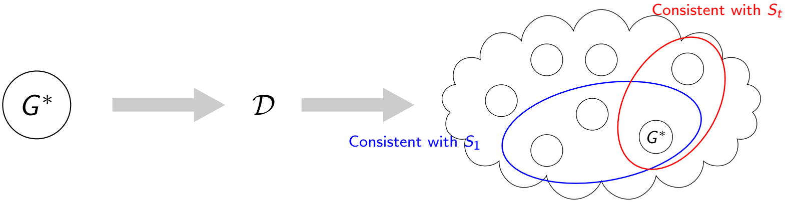
Goal: Identify G^* using as few interventions as possible



- We know that at least $\nu(G^*)$ interventions is *necessary*

The adaptive search problem

Goal: Identify G^* using as few interventions as possible

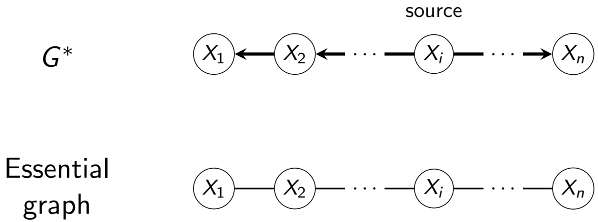


- We know that at least $\nu(G^*)$ interventions is *necessary*
- Punchline: $\mathcal{O}(\log n \cdot \nu(G^*))$ interventions suffice

"Search is almost as easy as verification"

Adaptive search: $\mathcal{O}(\log n \cdot \nu(G^*))$ interventions suffice

- Prior works only have theoretical guarantees for special classes of graphs: cliques, trees, intersection incomparable graphs, ...
- $\Omega(\log n \cdot \nu(G^*))$ interventions are necessary in the worst case

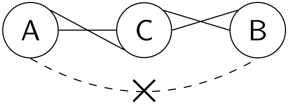


$\nu(G^*) = 1$ and identifying G^* is equivalent to binary search

- Covered edges are $X_{i-1} \leftarrow X_i \rightarrow X_{i+1} \Rightarrow$ Need to "hit" X_i
- Each intervention orients "one side" via Meek rule R1

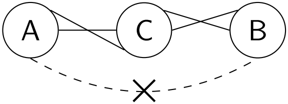
Adaptive search: How it works

- [HB12]: Intervene and remove oriented arcs $\Rightarrow$ *Chordal* graph
- [GRE84]: For any chordal graph, there exists a clique separator C such that
  - $|A|, |B| \leq n/2$
  - C is a clique, i.e. $|C| \leq \omega(G)$
  - Computable in polynomial time



Adaptive search: How it works

- [HB12]: Intervene and remove oriented arcs $\Rightarrow$ *Chordal* graph
- [GRE84]: For any chordal graph, there exists a clique separator C such that
  - $|A|, |B| \leq n/2$
  - C is a clique, i.e. $|C| \leq \omega(G)$
  - Computable in polynomial time



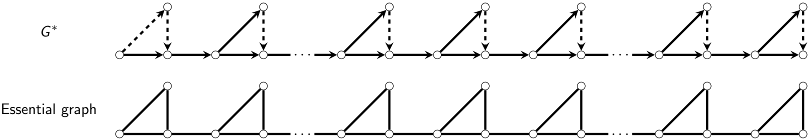
- Algorithm: Find clique separator in *each component*; Intervene on each node in all clique separators; Recurse
- Analysis:
  - $\mathcal{O}(\log n)$ rounds suffice $\leftarrow$ [GRE84]
  - $\mathcal{O}(\nu(G^*))$ per round $\leftarrow$ We prove new lower bound on $\nu(G^*)$

A stronger (but not computable) lower bound

Intuition [HB14]: In any interventional essential graph, interventions across different "connected components" *do not* help.

Claim: Fix an essential graph and some DAG G in it. Then,

$$\nu(G) \geq \sum_{\substack{\text{connected components} \\ H \in \text{ after removing oriented arcs}}} \left\lfloor \frac{\omega(H)}{2} \right\rfloor$$

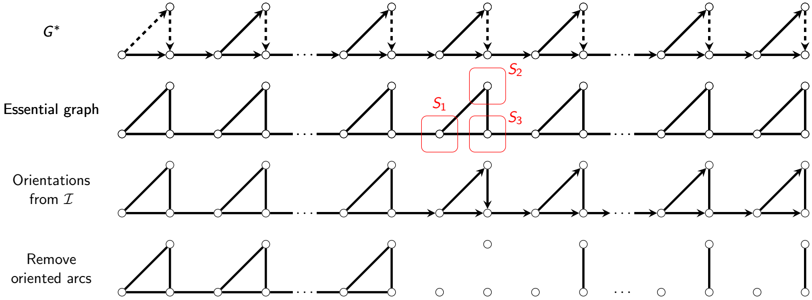


A stronger (but not computable) lower bound

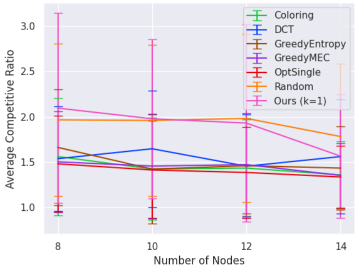
Intuition [HB14]: In any interventional essential graph, interventions across different "connected components" *do not* help.

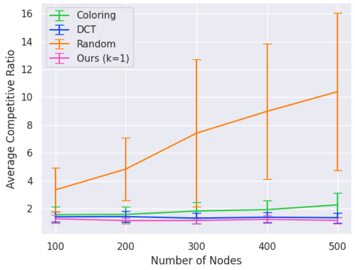
Claim: Fix an essential graph and some DAG G in it. Then,

$$\nu(G) \geq \max_{\substack{\text{atomic} \\ \text{interventions} \\ S_1,\ldots,S_t}} \sum_{\substack{\text{connected components} \\ H \in \text{ after removing oriented arcs} \\ \text{after interventions } S_1, \ldots, S_t}} \left\lfloor \frac{\omega(H)}{2} \right\rfloor$$



Experiments (Atomic search comparision)



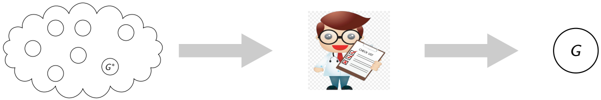


Qualitatively, our algorithm is competitive with the state-of-the-art search algorithms while being ∼10x faster in some experiments.

Implementation: https://github.com/cxjdavin/verification-and-search-algorithms-for-causal-DAGs

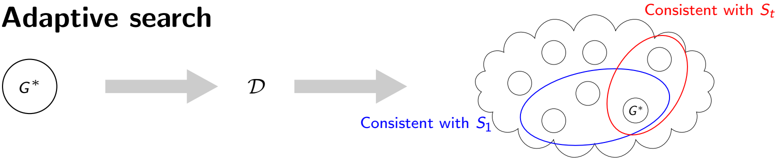
Summary

1. Verification



- Polynomial time *exact characterization* of $\nu(G)$
- $\nu(G) = \texttt{MVC}(\text{covered})$ to determine if $G \stackrel{?}{=} G^*$

2. Adaptive search



- Polynomial time adaptive search algorithm using interventions
- $\mathcal{O}(\log n \cdot \nu(G^*))$ suffice for *any general graph*
- $\Omega(\log n \cdot \nu(G^*))$ worst case necessary

Natural follow up questions

- In this work, we studied *verification* and *search* under an idealized setting with hard interventions and infinite samples.
- Soft interventions may be more realistic in certain real-life scenarios (e.g. effects from parental vertices are not completely removed but only altered); see [KJSB19]
- Sample complexities also play a crucial role when one has limited experimental budget; see [ABDK18]
- We also make standard assumptions such as the Markov assumption, the faithfulness assumption, and causal sufficiency [SGSH00]. Can we remove/weaken these assumptions?

References I

Jayadev Acharya, Arnab Bhattacharyya, Constantinos Daskalakis, and Saravanan Kandasamy.
Learning and Testing Causal Models with Interventions.
Advances in Neural Information Processing Systems, 31, 2018.

David Maxwell Chickering.
A Transformational Characterization of Equivalent Bayesian Network Structures.
In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, page 87–98, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.

David Maxwell Chickering.
Optimal Structure Identification with Greedy Search.
Journal of machine learning research, 3(Nov):507–554, 2002.

John R. Gilbert, Donald J. Rose, and Anders Edenbrandt.
A Separator Theorem for Chordal Graphs.
SIAM Journal on Algebraic Discrete Methods, 5(3):306–313, 1984.

Alain Hauser and Peter Bühlmann.
Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs.
The Journal of Machine Learning Research, 13(1):2409–2464, 2012.

Alain Hauser and Peter Bühlmann.
Two optimal strategies for active learning of causal models from interventional data.
International Journal of Approximate Reasoning, 55(4):926–939, 2014.

Murat Kocaoglu, Amin Jaber, Karthikeyan Shanmugam, and Elias Bareinboim.
Characterization and Learning of Causal Graphs with Latent Variables from Soft Interventions.
Advances in Neural Information Processing Systems, 32, 2019.

References II

Christopher Meek.
Causal Inference and Causal Explanation with Background Knowledge.
In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, page 403–410, San Francisco, CA, USA, 1995. Morgan Kaufmann Publishers Inc.

Vibhor Porwal, Piyush Srivastava, and Gaurav Sinha.
Almost Optimal Universal Lower Bound for Learning Causal DAGs with Atomic Interventions.
In *The 25th International Conference on Artificial Intelligence and Statistics*, 2022.

Peter Spirtes, Clark N. Glymour, Richard Scheines, and David Heckerman.
Causation, Prediction, and Search.
MIT press, 2000.

Chandler Squires, Sara Magliacane, Kristjan Greenewald, Dmitriy Katz, Murat Kocaoglu, and Karthikeyan Shanmugam.
Active Structure Learning of Causal DAGs via Directed Clique Trees.
Advances in Neural Information Processing Systems, 33:21500–21511, 2020.

Marcel Wienöbst, Max Bannach, and Maciej Liśkiewicz.
Extendability of causal graphical models: Algorithms and computational complexity.
In Cassio de Campos and Marloes H. Maathuis, editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 1248–1257. PMLR, 27–30 Jul 2021.