Verification and search algorithms for causal DAGs

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

[1]National University of Singapore

[2]Stanford University



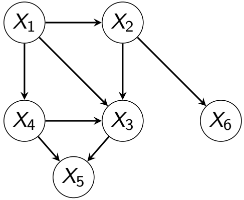
* Equal contribution

Motivation

Underlying data
generation process
(modelled as a DAG)

Observational data D

G



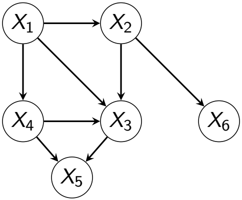
	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.3	0.4	0.1	-0.5	0.2	-0.3
Sample 2	0.1	1.2	0.6	-0.2	-0.1	-0.4
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

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

specific to node X_4

Motivation

G

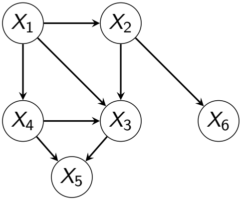


D

	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
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Motivation

G

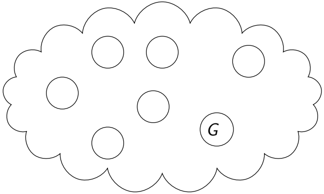


D

	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])

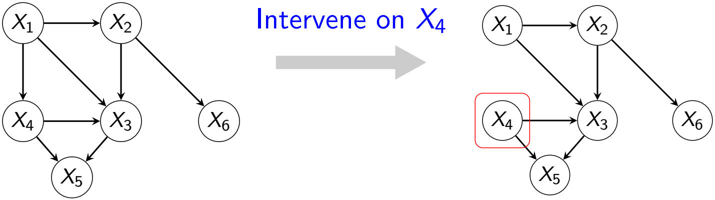
Equivalence
class of DAGs



Which is G ?

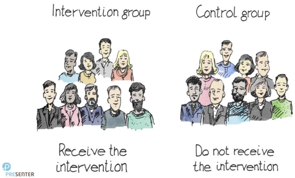
Two ways forward

1. Make model assumptions about the functional dependencies
   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

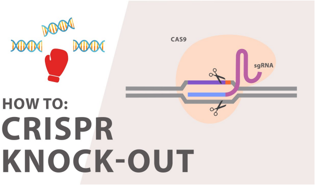


Interventions in real-life

Randomized controlled trials



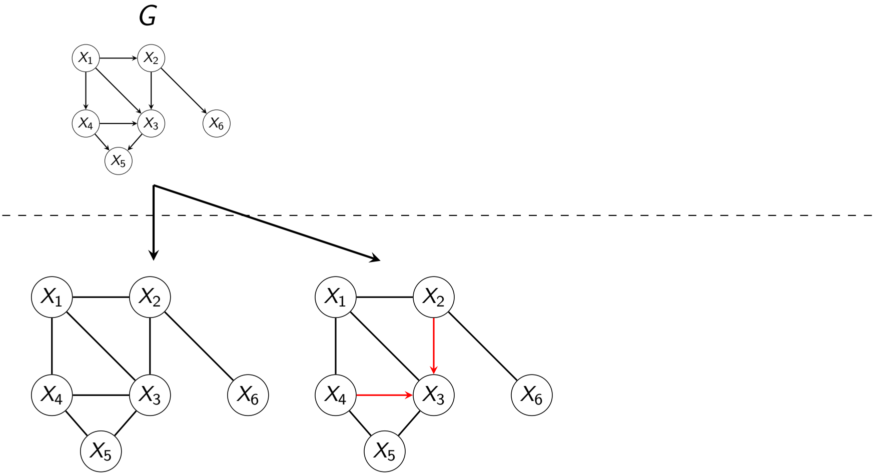
Gene knockout experiments



Can be expensive to perform) Minimize number of interventions!

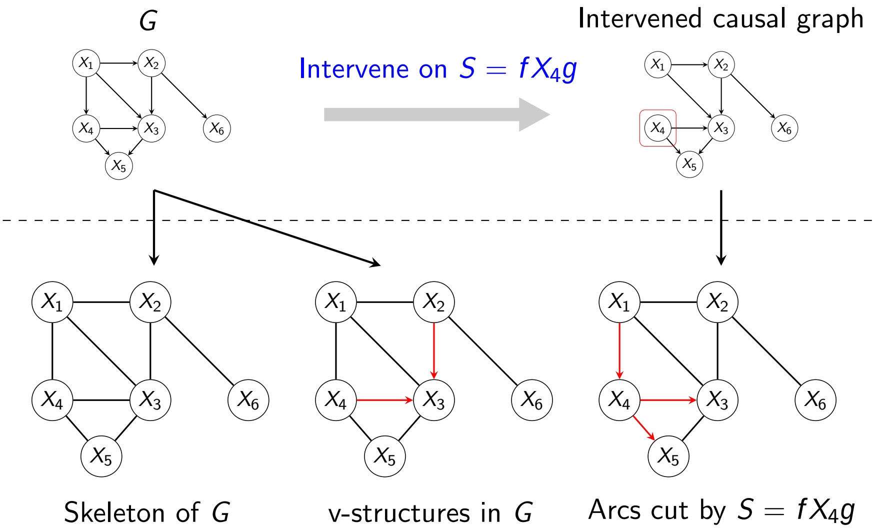
Image credit: https://i.ytimg.com/vi/vyVyg4Bf2JY/maxresdefault.jpg
https://i.ytimg.com/vi/1BSiFfyObTY/maxresdefault.jpg

What can we learn about G from D and interventions?



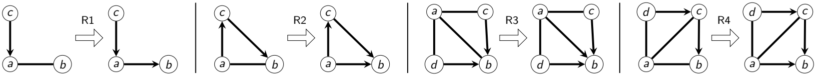
Skeleton of G v-structures in G

What can we learn about G from D and interventions?



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 a,
then v-structure

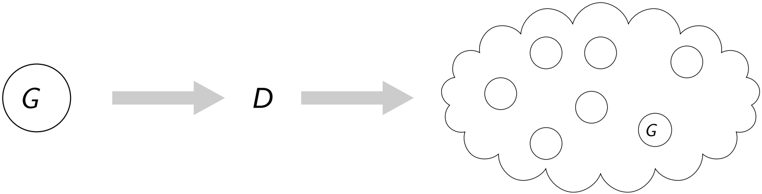
If b a,
then cycle

If b a, then the unoriented arcs would have been *oriented in the same way in all DAGs within the equivalence class* (via R2)

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

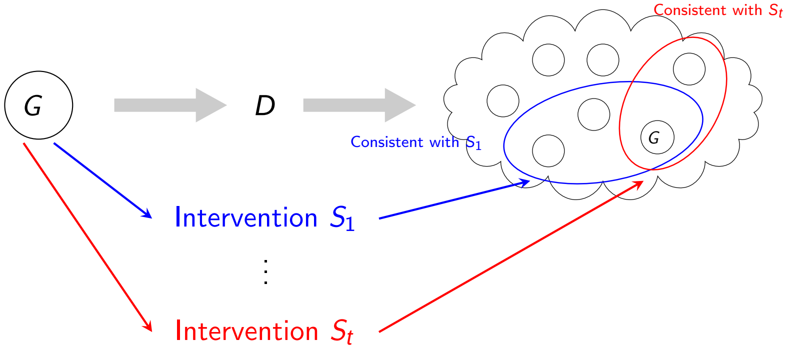
Problem setup

Identify G



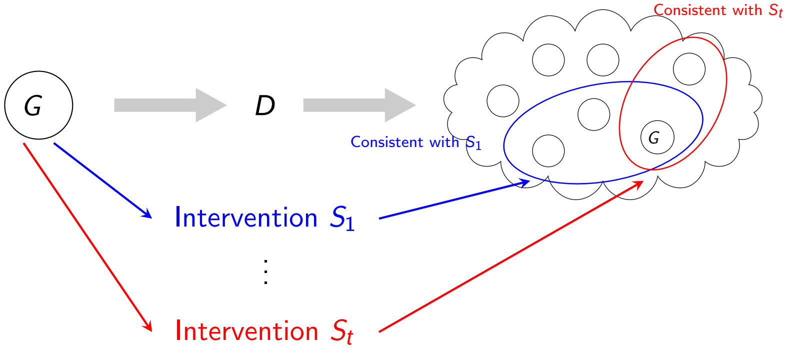
Problem setup (using interventions)

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



Problem setup (using **atomic** interventions)

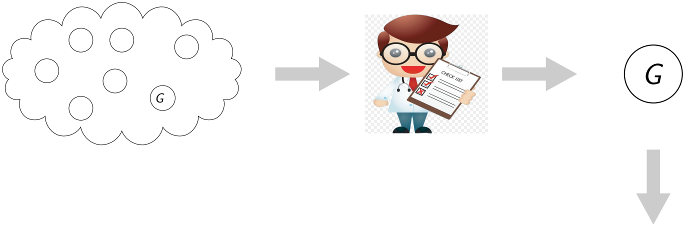
Identify G using as few interventions as possible (minimize t)



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!

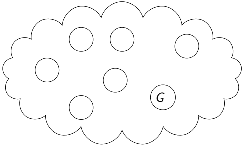


Do stuff with
discovered causal graph G

Wait a minute... we have domain experts!

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

~~Problem solved with zero interventions!~~

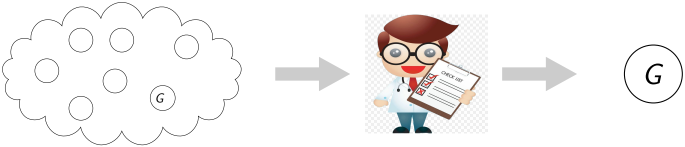


Experts can be wrong...

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. $\nu(G)$ MVC(unoriented)
(Can be a very bad upper bound!)

Verification: A complete characterization via covered edges

Meek rules $\Rightarrow$ Outperform `MVC`(unoriented)

Surprisingly, enough to compute `MVC` on a *subset of edges*

Covered edges [Chi95]:
$u \to v$ is covered edge $\iff$ $\mathrm{Pa}(u) \setminus \{v\} = \mathrm{Pa}(v) \setminus \{u\}$

Claim: Necessary and sufficient to intervene on `MVC`(covered)
Proof: Simple (but subtle) using the notion of covered edges

Claim: Covered edges form a forest.
Implication: `MVC`(covered) can be computed *exactly* in *linear time*.

Easy re-interpretation of known facts via covered edges

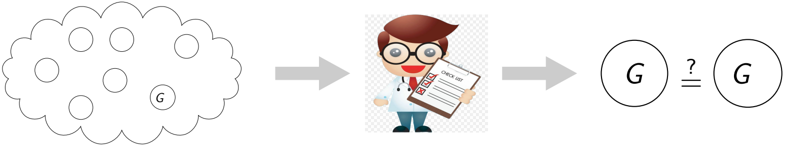
- Covered edges of clique K_n: $v_1 \to v_2, \ldots, v_{n-1} \to v_n$
- Covered edges of a tree: incident edges to root vertex
- Necessity of separating system for non-adaptive interventions
  - [Chi95]: Two graphs are equivalent $\iff$ there is a sequence of covered edge reversals to transform between 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 $G \overset{?}{=} G$

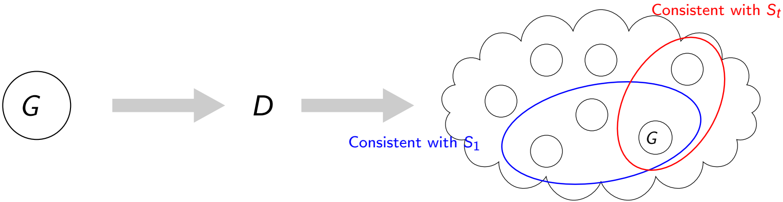
- Using $\nu(G) = \texttt{MVC}(\text{covered})$ interventions
- Computable in polynomial time



What about actually searching for G without the expert?

The adaptive search problem

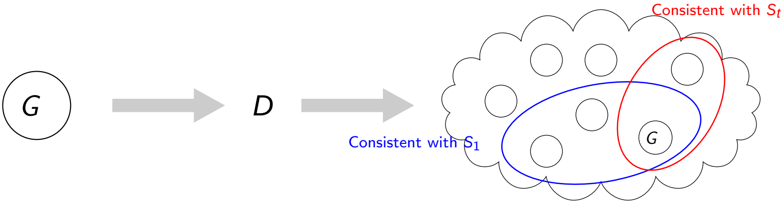
Goal: Identify G using as few interventions as possible



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

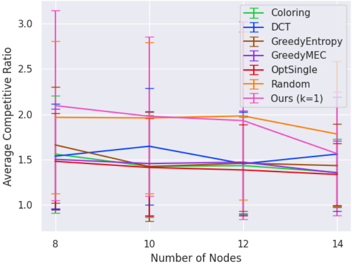
The adaptive search problem

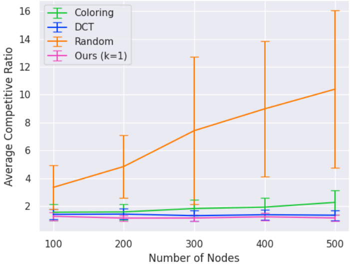
Goal: Identify G using as few interventions as possible



- We know that at least $\nu(G)$ interventions is *necessary*
- Punchline: $O(\log n \quad \nu(G))$ interventions suffice
  - Algorithm: Use chordal graph separators; recurse on subgraphs
  - Analysis: We prove stronger lower bound on $\nu(G)$
- Prior works only have theoretical guarantees on special classes of graphs; The guarantee that we have holds for *any* graph.

Experiments (Atomic search comparision)



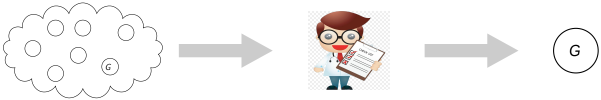


Qualitatively, our algorithm is competitive with the state-of-the-art search algorithms while being 10x faster in some experiments.

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

Summary

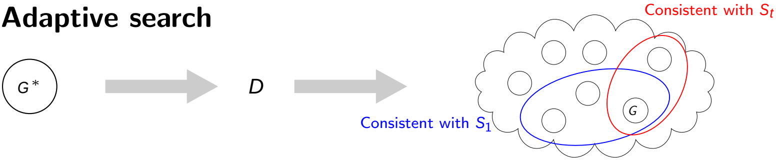
1. **Verification**



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

$O(\log n \quad \nu(G \quad))$ suffice for *any general graph*

$\Omega(\log n \quad \nu(G \quad))$ 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?

Want to learn more?

Read our paper and/or see our longer talk here:

https://github.com/cxjdavin/verification-and-search-algorithms-for-causal-DAGs/tree/main/talk

More examples to facilitate understanding and explanation of intuition behind some of our techniques, including:

- Why is identifying a set of interventions to fully orient G is equivalent to answering $G \stackrel{?}{=} G$
- A simple concrete example showing why the prior known bounds on $\nu(G)$ is loose.
- Why is $\Omega(\log n \quad \nu(G \quad))$ necessary for search?
- What is our stronger lower bound? How does it work?

Thank you for your kind attention!

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 Arti cial 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.

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.

Christopher Meek.
Causal Inference and Causal Explanation with Background Knowledge.
In *Proceedings of the Eleventh Conference on Uncertainty in Arti cial 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 Arti cial Intelligence and Statistics*, 2022.

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

References II

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 Arti cial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 1248–1257. PMLR, 27–30 Jul 2021.