Almost Optimal Universal Lower Bound for Learning Causal DAGs with Atomic Interventions

Vibhor Porwal
Adobe Research

Piyush Srivastava
Tata Institute of Fundamental Research

Gaurav Sinha
Adobe Research

Introduction

- Learning causal DAGs is an important problem in several fields including health science, molecular cell biology, and computational advertising.

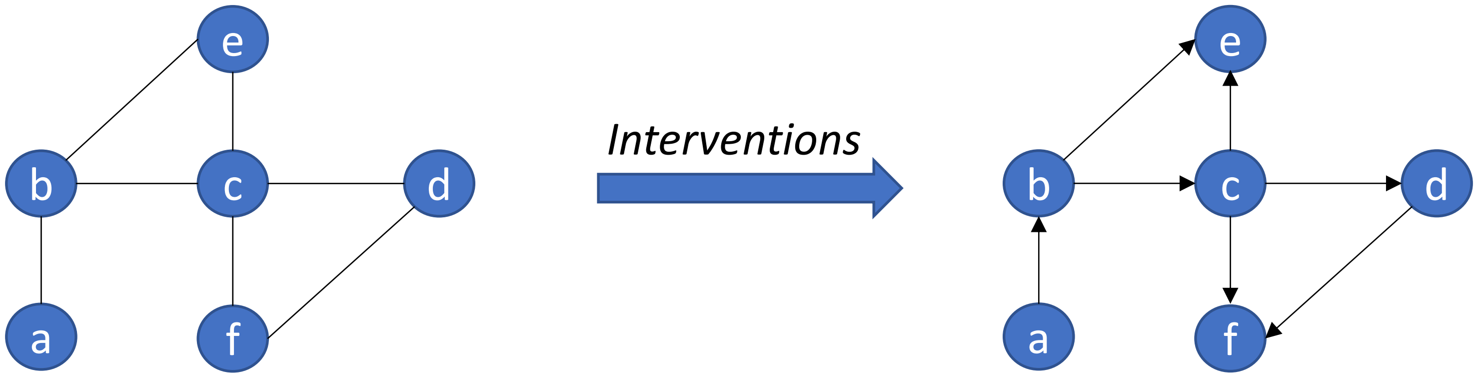
- Using observational data, it is possible to learn a causal DAG only up to its Markov Equivalence Class (MEC).

- Interventions (or Experiments) are required to orient the remaining edges.

- An intervention on v completely randomizes the distribution of v and reveals at least the direction of all edges incident on v.

Introduction

- Interventions are costly, and thus, a natural question is to find the *minimum number of interventions* required to learn all edges of a causal DAG.



- A *v-structure* in a DAG is an induced subgraph of the form $b \rightarrow a \leftarrow c$.

- Lower and upper bounds on the number of interventions for learning DAGs without v-structures generalize to arbitrary DAGs (Hauser and Bühlmann, 2014).

Universal Lower Bounds

- Squires et al. (2020) introduced the concept of universal lower bounds.

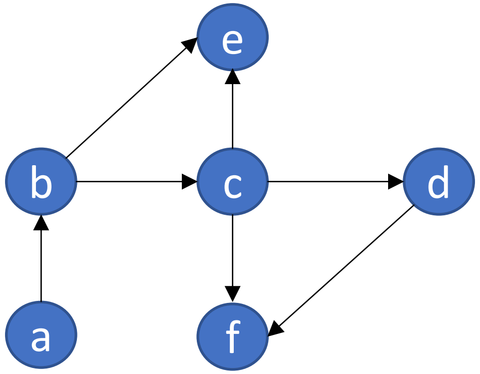
- A universal lower bound of L interventions for a MEC means that if a set of interventions I is of size less than L, then for any ground-truth DAG in the MEC, I will fail to fully orient the MEC.

- Universality properties:
  - The lower bound applies to all DAGs in the MEC.
  - The lower bound applies to every set of interventions regardless of the method by which the intervention set was obtained.

- We address the problem of obtaining *tight* universal lower bounds.

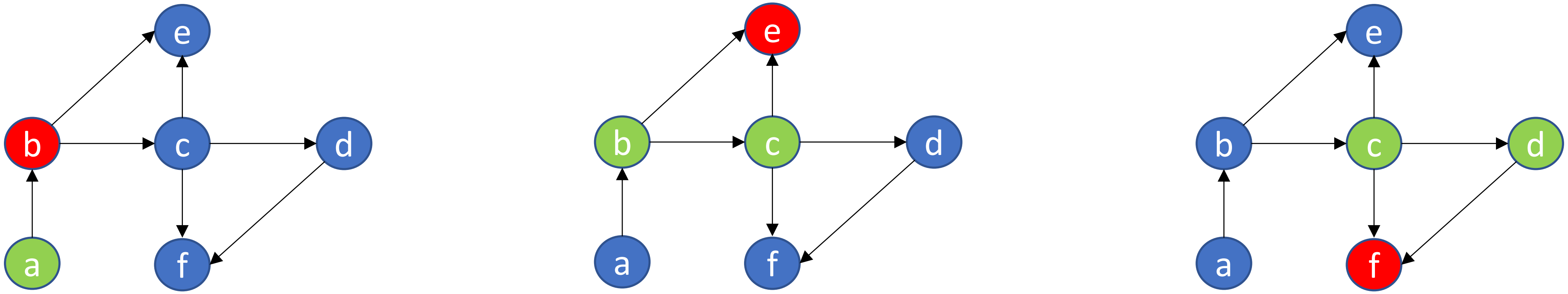
Maximal-Clique-Sink Nodes

- Let D be a DAG without v-structures having n nodes.



- Let r denote the number of maximal cliques in D.

- The node with no outgoing edges in $D[C]$ is called the *maximal-clique-sink* node corresponding to the maximal clique C.



Clique-Block Shared-Parents (CBSP) orderings

- Let σ be a topological ordering of D and let $s_1, s_2, \ldots, s_r$ be its maximal-clique-sink nodes indexed so that $\sigma(s_i) < \sigma(s_j)$ when $i < j$.

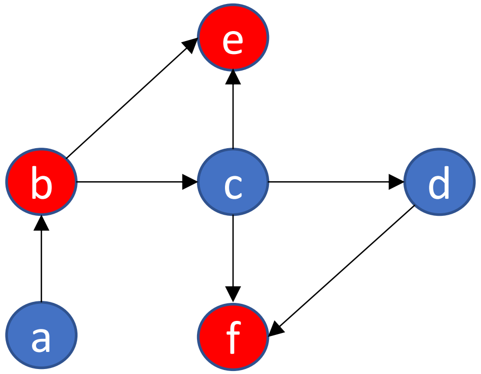
- Let $L_1(\sigma)$ be the set of nodes that occur before or at the same position as s_1 in σ.

- For $2 \leq i \leq r$, define $L_i(\sigma)$ to be the set of nodes that occur in σ before or at the same position as s_i, but strictly after s_{i-1}.

- σ is said to be a CBSP ordering if it satisfies the following two properties:
  - (P1) *Clique block property*: For each $i \in [r]$ the subgraph induced by $L_i(\sigma)$ in D is a clique.
  - (P2) *Shared parents property*: If a and b are consecutive in σ, and also lie in the same $L_i(\sigma)$ for some $i \in [r]$, then all parents of a are also parents of b in D.

CBSP orderings: Example

- Consider the following topological orderings.



$$\sigma' = a\ b\ c\ d\ e\ f$$

$$\sigma = a\ b\ c\ d\ f\ e$$

$$\tau = a\ b\ c\ e\ d\ f$$

- Here, σ' does not satisfy P1 or P2, σ satisfies only P1, τ satisfies both P1 and P2 and therefore, τ is a CBSP ordering.

Overview of our Results

- Lower bound: Let D be a DAG without v-structures. Then, any set of atomic interventions that learns all edges of D must be of size at least $\left\lceil \frac{n-r}{2} \right\rceil$.

- Upper bound: There exists an intervention set I of size $n - r$ that can learn all edges of D, showing that our lower bound is tight up to a factor of 2.

- We show that $n - r \geq \omega - 1$ in chordal graphs, proving that our lower bound is at least as good as the one by Squires et al. (2020).

- Through simulations and by constructing examples of graphs, we show that our lower bound is often significantly better.

Proof Sketch for our Lower Bound

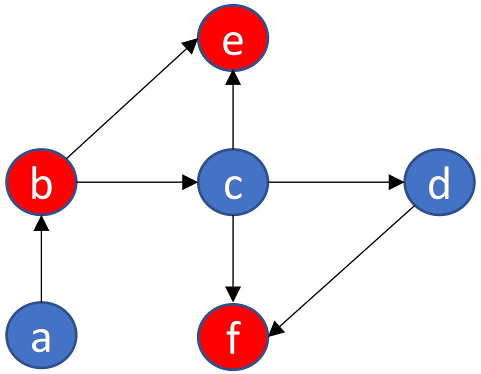
- Step 1: Prove that any DAG without v-structures has a CBSP ordering.
- Step 2: Take a CBSP ordering σ of D. Let a, b be two consecutive nodes of σ and $a, b \in L_i(\sigma)$ for some $i \in [r]$.
- Step 3: Show that if I is such that $I \cap \{a, b\} = \emptyset$, then I cannot orient the edge $a - b$.
- Step 4: Thus, if I learns all edges of D then,

$$|I| \geq \sum_{i=1}^{r} \left\lceil \frac{|L_i(\sigma)| - 1}{2} \right\rceil \geq \left\lceil \sum_{i=1}^{r} \frac{|L_i(\sigma)| - 1}{2} \right\rceil$$

$$= \left\lceil \frac{\sum_{i=1}^{r} |L_i(\sigma)|}{2} - \frac{r}{2} \right\rceil = \left\lceil \frac{n - r}{2} \right\rceil$$

Example

- Let D be the following DAG and τ be a CBSP ordering of D.



$$\tau = a\ b\ c\ e\ d\ f$$

- Let I be an intervention set that can learn all edges of D.

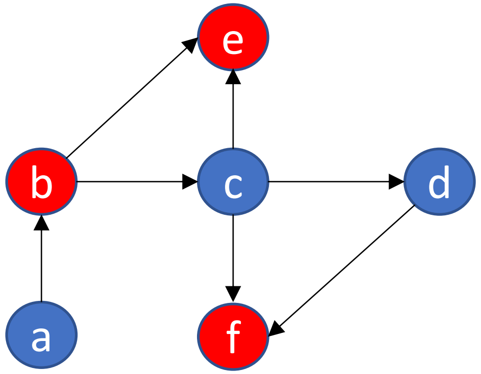
- I must of size at least $\left\lceil \frac{|\{a,b\}|-1}{2} \right\rceil + \left\lceil \frac{|\{c,e\}|-1}{2} \right\rceil + \left\lceil \frac{|\{d,f\}|-1}{2} \right\rceil = 3.$

Tightness of Universal Lower Bound

- Let the set of maximal-clique-sink nodes of D be $S = \{s_1, s_2, \ldots, s_r\}$.

- We can show that $I^* = V \setminus S$ is a set of interventions that can orient the MEC of D if the underlying DAG was D.

- Note that, $|I^*| = n - r$, showing that our universal lower bound is tight up to a factor of 2.



$$I^* = \{a, c, d\}$$

$$\left\lceil \frac{n-r}{2} \right\rceil = 2$$

References

- Hauser, A. and Bühlmann, P. (2014). Two Optimal Strategies for Active Learning of Causal Models from Interventional Data. *International Journal of Approximate Reasoning*, 55(4):926-939.

- Squires, C., Magliacane, S., Greenewald, K. H., Katz, D., Kocaoglu, M., and Shanmugam, K. (2020). Active Structure Learning of Causal DAGs via Directed Clique Trees. *Proceedings of 34th Annual Conference on Neural Information Processing Systems (NeurIPS 2020)*. arXiv:2011.00641.

Thanks!