A Transformational Characterization of Unconditionally Equivalent Bayesian Networks

Alex Markham, Danai Deligeorgaki, Pratik Misra and Liam Solus

Math of Data and AI Group, KTH Royal Institute of Technology, Sweden

Oct 5, 2022
11th International Conference on Probabilistic Graphical Models

Outline

Introduction

Motivation

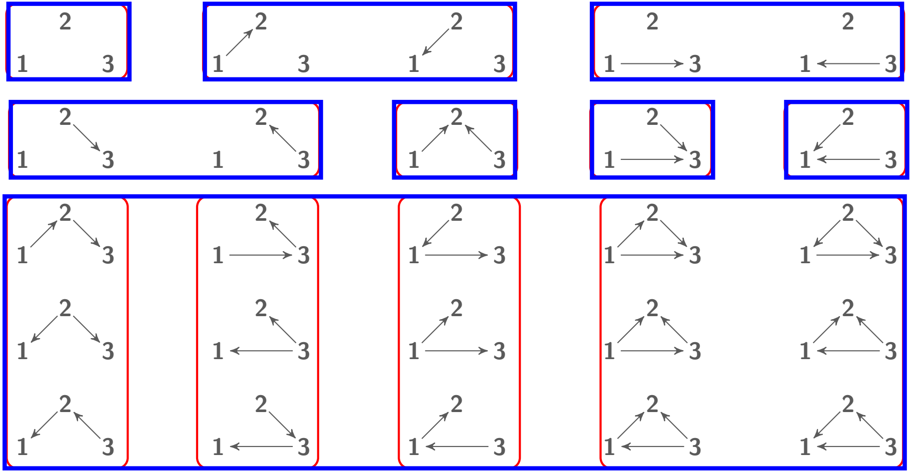
- Two DAGs $\mathcal{D}, \mathcal{D}'$ that share the same set of *d*-separation statements of the form $A \perp B \mid \emptyset$ are called *unconditionally equivalent*.
- Corresponding *unconditional equivalence classes (UECs)* of DAGs:
  - are easily estimated from data (e.g., nonzero entries of a covariance matrix),
  - are represented by simple undirected graphs,
  - but still contain a surprising amount of causal information.
- Hence, understanding UECs can help with:
  - (partial) causal discovery for otherwise prohibitively large structures
  - and speeding up existing causal discovery methods (e.g., GES).
- Previous work[1] focused on enumerating DAGs faithful to a given UEC.
- In contrast, we give novel characterizations and properties that facilitate exploring the space of UECs.

---

[1] J. Textor, A. Idelberger, and M. Liśkiewicz. "Learning from pairwise marginal independencies". In: *31st Conference on Uncertainty in Artificial Intelligence (UAI 2015)* (2015). arXiv:1508.00280 [cs.AI]

Introduction

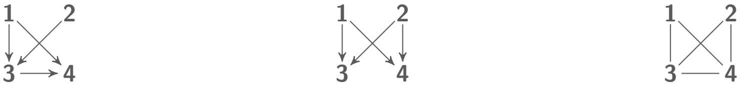
DAGs, MECs, and UECs example



Introduction

Unconditional dependence graphs (UDGs)

The *unconditional dependence graph (UDG)* of a DAG $\mathcal{D} = (V, E)$ is the undirected graph $\mathcal{U} = (V, \{\{v, w\} : v \not\perp_{\mathcal{D}} w;\ v, w \in V\})$.



(a) DAG $\mathcal{D}$ with d-separations $\mathcal{I} = \{1 \perp 2 \mid \emptyset,\ 2 \perp 4 \mid \{1, 3\}\}$

(b) DAG $\mathcal{D}'$ with d-separations $\mathcal{I} = \{1 \perp 2 \mid \emptyset,\ 3 \perp 4 \mid \{1, 2\}\}$

(c) UDG representing $\mathcal{I} = \{1 \perp 2\}$

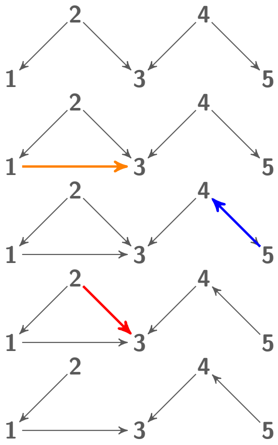
Figure: Unconditionally equivalent DAGs $\mathcal{D}, \mathcal{D}'$ with common unconditional d-separation statement $1 \perp 2$ and corresponding UDG missing edge $\{1, 2\}$

Transformational characterization for DAGs

Theorem (Transformational Characterization)

Let $\mathcal{D}$ and $\mathcal{D}'$ be two unconditionally equivalent DAGs. There exists a sequence of $|E^{D'} \setminus E^{D}|$ edge insertions, followed by $|\Delta(\mathcal{R}^{\mathcal{D},\mathcal{D}'}, \mathcal{R}^{\mathcal{D}',\mathcal{D}})|$ edge reversals, followed by $|E^{D} \setminus E^{D'}|$ edge deletions that transforms $\mathcal{D}$ into $\mathcal{D}'$ with the following properties:

1. *Each edge inserted or deleted in $\mathcal{D}$ is partially weakly covered or implied by transitivity.*
2. *Each edge reversed in $\mathcal{D}$ is weakly covered.*
3. *After each operation, the resulting $\mathcal{D}$ is a DAG and $\mathcal{U}^{\mathcal{D}} = \mathcal{U}^{\mathcal{D}'}$.*
4. *After all operations, $\mathcal{D} = \mathcal{D}'$.*



Transforming essential graphs (CPDAGs)

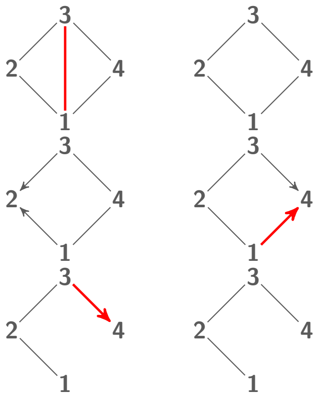
Theorem sketch and example

Theorem (11)

The edge $v \multimap w$ of an essential graph G is removable if and only if $\mathrm{mt}_{\mathcal{G}_{-v \multimap w}}(v) \subseteq \mathrm{mt}_{\mathcal{G}_{-v \multimap w}}(w)$.

Theorem (13)

Given an essential graph $\mathcal{G}$ and removable edge $v \multimap w$, define $T := \mathrm{ne}_{\mathcal{G}}(v) \cap \mathrm{ne}_{\mathcal{G}_{\mathrm{cc}(w)}}(w)$. Based on $|T|$ and whether $v \to w$ or $v - w$, we can determine if the PDAG $\mathcal{G}_{-v \multimap w}$ is complete and can enumerate all unconditionally equivalent completions otherwise.



Conclusion and future work

Conclusion: We derive novel characterizations of *unconditional equivalence classes (UECs)*, resulting in a new, more efficient way of moving around the space of DAGs.

Future work:

- Markov Chain Monte Carlo (MCMC) methods for learning UECs[2]
  - learn BIC-optimal UEC, MAP estimate, and full posterior $\pi(\mathcal{U} \mid \texttt{Data})$
- GES algorithms (including extension to MAGs)
  - first, take "big" steps between UECs in an uphill/forward phase
  - then, take "small" steps between CPDAGs (or PAGs) within the vastly reduced space of a single UEC in a downhill/backward phase
- generalizing to small instead of empty conditioning sets
  - more causal info than UECs; less computationally expensive than arbitrary conditioning sets

[2] D. Deligeorgaki, A. Markham, P. Misra, and L. Solus. "Combinatorial and algebraic perspectives on the marginal independence structure of Bayesian networks". In: (2022). arXiv:2210.00822 [stat.ME]