Local Discovery by Partitioning

Polynomial-Time Causal Discovery Around Exposure-Outcome Pairs

Jacqueline Maasch

maasch@cs.cornell.edu | arXiv:2310.17816

Joint work by: J Maasch[1], W Pan[2], S Gupta[3], V Kuleshov[1], K Gan[4], F Wang[2]

[1]Computer Science, Cornell Tech
[2]Population Health Sciences, Weill Cornell Medicine
[3]Machine Learning, Carnegie Mellon University
[4]Operations Research and Information Engineering, Cornell Tech

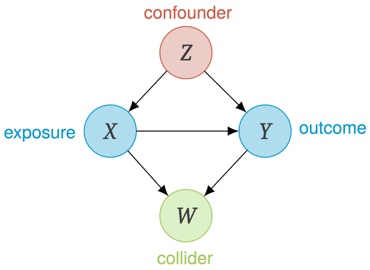
POMS Annual Conference | Minneapolis, MN | 27 April 2024

Causal inference with observational data

1. **Identify the causal quantity of interest.**
   - *Example:* Average treatment effect (ATE) of a drug on a disease state.
   - A graphical model of the data generating process (DGP) enables identifiability.
   - We can learn this model with data-driven methods.

2. **Perform inference to estimate this quantity.**
   - Express the parameter as a function of the DGP.
   - Apply estimation methods (e.g., TMLE, doubly robust ML, etc.).

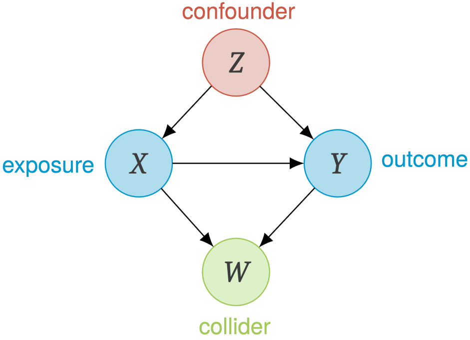
Graphical models for causal effect identification



Blocking all **backdoor paths** for $\{X, Y\}$ by adjusting for confounder Z allows for *unconfoundedness* or *conditional exchangeability*: $Y(1), Y(0) \perp\!\!\!\perp X \mid Z$.

This removes *noncausal association* for unbiased ATE estimation.

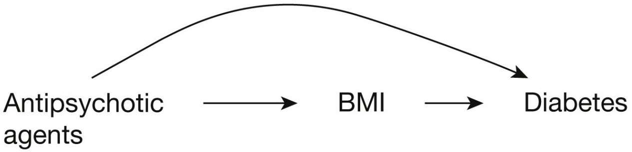
Graphical models for causal effect identification



The correct directed acyclic graph (DAG) enables unique identification of the true ATE:

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_Z\left[\mathbb{E}[Y \mid X = 1, Z] - \mathbb{E}[Y \mid X = 0, Z]\right]$$

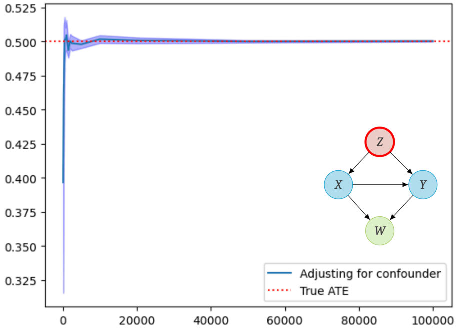
Real-world example: diabetes risk

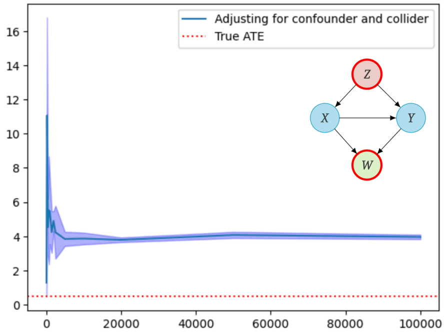


Whether the patient takes certain antipsychotics is a confounder for BMI and risk of developing diabetes [ECM20].

Effect estimation with a misspecified model

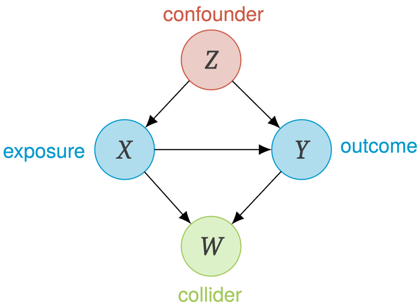
1 Background





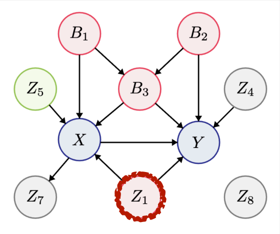
ATE estimates converge to the true value when controlling for Z only (left), but remain biased when controlling for $\{W, Z\}$ (right).

Causal discovery: learning structure from data

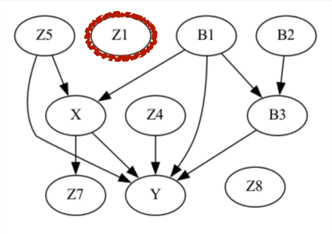


- **Data-driven:** Learn the underlying graphical model, with or without prior knowledge.
- **Global discovery:** Learn the entire DAG from data.
- **Local discovery:** Learn only the relevant substructures (e.g., role of Z only).

Failure modes of global discovery



Ground truth



PC $(n = 10k)$

- **Constraint-based methods PC and FCI** [SGS00] use conditional independence tests to identify the undirected skeleton of the graph and orient edges.
- **Drawbacks:** Exponential time complexity, high sample complexity, order dependence.

Local Discovery by Partitioning (LDP)

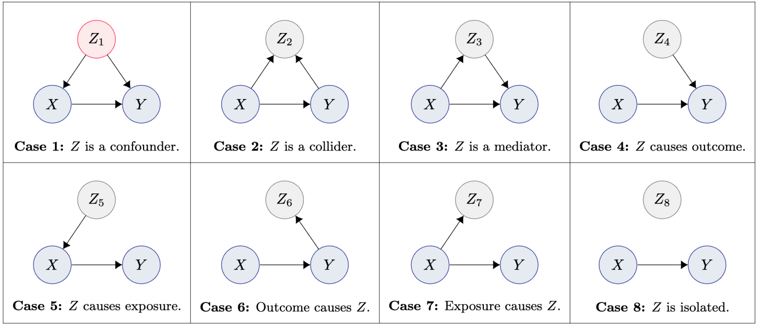
To address these failure modes for the setting of downstream causal effect estimation:

1. We prove the existence of an exhaustive **causal partition taxonomy** defining any arbitrary DAG w.r.t. the exposure and outcome.
2. We propose a **local discovery procedure** that learns causal partitions directly.
3. LDP is asymptotically **guaranteed to return a confounder set** for unbiased ATE estimation.

Local causal partition learning

2 Local Discovery by Partitioning (LDP)

For downstream inference, we only care about the **local structure** relative to $\{X, Y\}$.

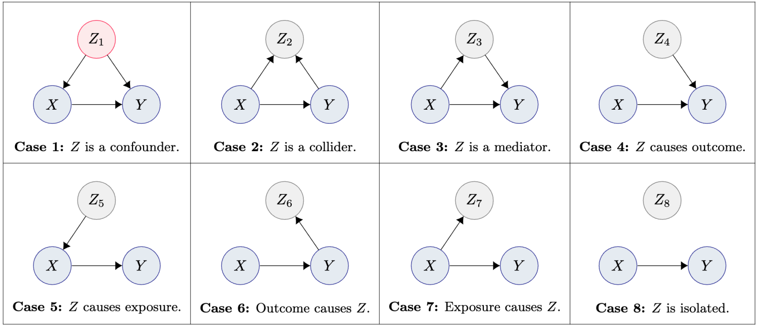


Case 1: Z is a confounder.
Case 2: Z is a collider.
Case 3: Z is a mediator.
Case 4: Z causes outcome.
Case 5: Z causes exposure.
Case 6: Outcome causes Z.
Case 7: Exposure causes Z.
Case 8: Z is isolated.

Local causal partition learning

2 Local Discovery by Partitioning (LDP)

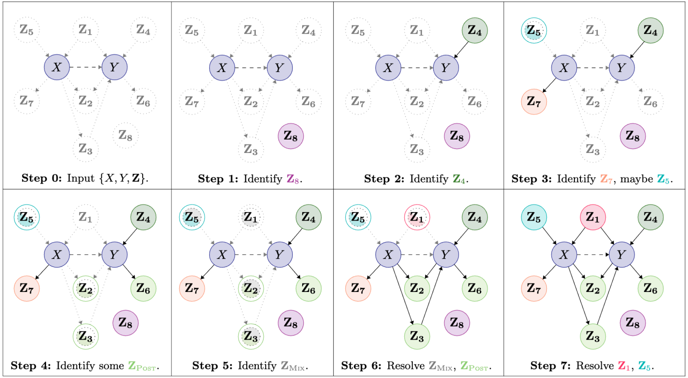
Universal property of DAGs: There exists a unique partitioning of the variables into eight exhaustive, mutually exclusive subsets defined by their relation to $\{X, Y\}$.



Case 1: Z is a confounder.
Case 2: Z is a collider.
Case 3: Z is a mediator.
Case 4: Z causes outcome.
Case 5: Z causes exposure.
Case 6: Outcome causes Z.
Case 7: Exposure causes Z.
Case 8: Z is isolated.

LDP learns causal partitions directly

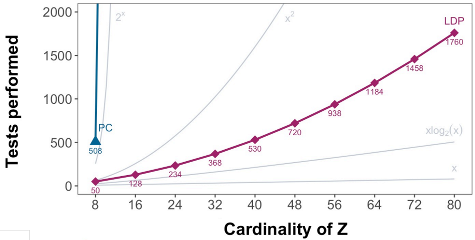
2 Local Discovery by Partitioning (LDP)

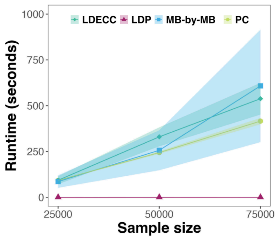


Partition labels can be obtained with nonparametric or parametric independence tests.

Fewer tests and faster runtimes

2 Local Discovery by Partitioning (LDP)



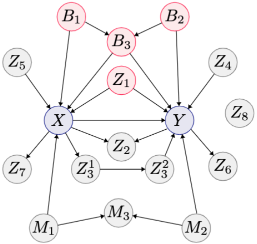


- **Polynomial-time:** Worst-case quadratic number of CI tests w.r.t. cardinality.
  - *Left:* Local and global constraint-based baselines are worst-case exponential.
  - *Right:* On a bnlearn benchmark (33 nodes), LDP ran **1400× to 2500× faster** than PC.

LDP for confounder discovery

2 Local Discovery by Partitioning (LDP)

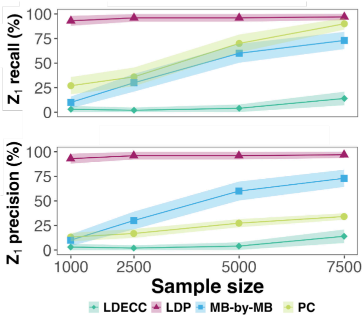
Asymptotically guaranteed to return a **valid adjustment set** (VAS) under **latent confounding** and mild graphical conditions.



$\{Z_1, B_1, B_2, B_3\}$ is a VAS of confounders for $\{X, Y\}$:
1) Blocks all backdoor paths and 2) contains no descendants of X [PJS17].

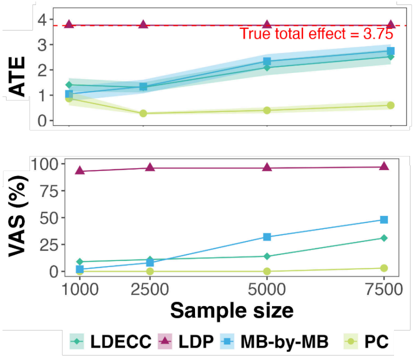
LDP for confounder discovery

2 Local Discovery by Partitioning (LDP)



- **Sample efficient:** Most conditioning sets of size one or two.
  - Local and global baselines use larger conditioning set sizes, on average.
  - LDP is more performant on small finite samples.

LDP for precise and unbiased ATE estimation



Results on a 10-node linear-Gaussian DAG.

Thank you! Any questions?

maasch@cs.cornell.edu

arXiv:2310.17816

jmaasch.github.io

References

[ECM20] M. Etminan et al. "Using causal diagrams to improve the design and interpretation of medical research". In: *Chest* 158.1 (2020), S21–S28.

[SGS00] P. Spirtes et al. *Causation, Prediction, and Search*. en. Ed. by J. Berger et al. Vol. 81. Lecture Notes in Statistics. New York, NY: Springer New York, 2000. DOI: 10.1007/978-1-4612-2748-9.

[PJS17] J. Peters et al. *Elements of causal inference: foundations and learning algorithms*. Cambridge, Massachusetts: The MIT Press, 2017.

[Lu+21] H. Lu et al. "Revisiting Overadjustment Bias". en. In: *Epidemiology* 32.5 (2021), e22–e23. DOI: 10.1097/EDE.0000000000001377.

[SCP09] E. F. Schisterman et al. "Overadjustment Bias and Unnecessary Adjustment in Epidemiologic Studies". en. In: *Epidemiology* 20.4 (2009), pp. 488–495. DOI: 10.1097/EDE.0b013e3181a819a1.

[HHR04] M. A. Hernán et al. "A Structural Approach to Selection Bias". en. In: *Epidemiology* 15.5 (2004), pp. 615–625. DOI: 10.1097/01.ede.0000135174.63482.43.

[EW14] F. Elwert et al. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable". In: *Annual Review of Sociology* 40.1 (2014), pp. 31–53. DOI: 10.1146/annurev-soc-071913-043455.

[HA22] M. J. Holmberg et al. "Collider bias". In: *JAMA Guide to Statistics and Methods* 327.13 (2022).

[Pea01] J. Pearl. "Direct and Indirect Effects". In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. 2001.

[Pea12] J. Pearl. "On a Class of Bias-Amplifying Variables that Endanger Effect Estimates". en. In: (2012).

[SLG16] M. E. Schnitzer et al. "Variable Selection for Confounder Control, Flexible Modeling and Collaborative Targeted Minimum Loss-Based Estimation in Causal Inference". In: *The International Journal of Biostatistics* 12.1 (2016), pp. 97–115. DOI: 10.1515/ijb-2015-0017.

[Lee+22] J. J. Lee et al. "Causal determinants of postoperative length of stay in cardiac surgery using causal graphical learning". In: *The Journal of Thoracic and Cardiovascular Surgery* (2022), S002252232200900X. DOI: 10.1016/j.jtcvs.2022.08.012.

Causal Markov and faithfulness

Faithfulness Assumption

Recall the Markov assumption: $X \perp\!\!\!\perp_G Y \mid Z \implies X \perp\!\!\!\perp_P Y \mid Z$

Causal graph $\longrightarrow$ Data

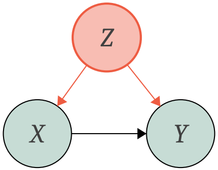
Causal graph $\longleftarrow$ Data

Faithfulness: $X \perp\!\!\!\perp_G Y \mid Z \impliedby X \perp\!\!\!\perp_P Y \mid Z$

https://www.bradyneal.com/causal-inference-course

Preliminaries: Non-causal associations

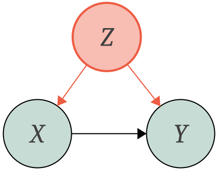
Definition 2.3 (Backdoor path, Pearl 2009)**.** Any non-causal path between exposure X and outcome Y with an edge pointing into X $(\cdots \rightarrow X)$.



Valid adjustment under the backdoor criterion

Definition 2.4 (Valid adjustment under the backdoor criterion, Peters et al. 2017)**.** Let $\mathbf{A}_{XY}$ be an adjustment set for $\{X, Y\}$ that does not contain $\{X, Y\}$. $\mathbf{A}_{XY}$ is valid if

1. $\mathbf{A}_{XY}$ contains no descendants of X and
2. $\mathbf{A}_{XY}$ blocks all backdoor paths from X to Y.



Why not adjust for *everything*?

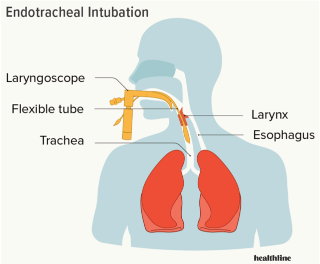
- **Bias:** Multiple variable types can induce bias when retained for adjustment [Lu+21; SCP09].
  1. Colliders induce selection bias [HHR04; EW14; HA22].
  2. Mediators bias total effects by controlling for indirect effects [Pea01].
  3. Instruments can amplify existing bias or introduce new bias in some settings [Pea12].

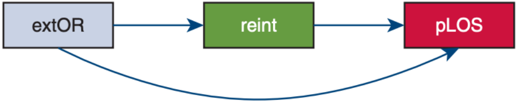
- **Variance:** Unnecessary adjustment can inflate the variance of effect estimates [SCP09].

- **Curse of dimensionality:** Unnecessary adjustment can undermine model fitting [SLG16].

Graphical models for causal effect identification

Real-world example: causal determinants of postoperative length of stay





Extubation in the operating room (**extOR**) is a confounder for the effect of reintubation (**reint**) on postoperative length of stay (**pLOS**) after cardiac surgery [Lee+22].

Sufficient conditions for identifiability

Sufficient conditions for VAS discovery are more relaxed than for correct partitioning.

Sufficient Conditions for Correct Partitioning Given an independence oracle, we define *sufficient* (but not necessary) conditions for asymptotically correct partition labeling:

- C1 The absence of inter-partition active paths (Def. 3.2).
- C2 The existence of at least one Z_4.
- C3 The existence of at least one Z_5. Further, all $\mathbf{Z}_1$ are marginally independent of at least one observed Z_5.
- C4 Causal sufficiency in $\mathcal{G}_{XY\mathbf{Z}}$.

Sufficient Conditions for VAS Identification Per Definiton 2.4, a VAS 1) contains no descendants of X and 2) blocks all backdoor paths from X to Y. With Theorem 4.5, we show that the VAS returned by LDP (Partition $\mathbf{Z}_1$) meets both criteria (Lemmas 4.2, 4.4) in the presence of causal insufficiency and arbitrary inter-partition active paths, given Condition C2, Condition C3, and a non-empty $\mathbf{Z}_{5 \in adj(X)}$.

LDP learns partitions directly

Algorithm 1 *Local Discovery by Partitioning (LDP)*

input $X, Y, \mathbf{Z}$, independence test, significance level α.
output Partitions of $\mathbf{Z}$: $\mathbf{Z}_1, \mathbf{Z}_4, \mathbf{Z}_5, \mathbf{Z}_7, \mathbf{Z}_8, \mathbf{Z}_{\text{POST}}$.

Copy $\mathbf{Z}' \leftarrow \mathbf{Z}$
for all $Z \in \mathbf{Z}'$ **do**
  ▷ STEP 1: TEST FOR $\mathbf{Z}_8$
  if $X \perp\!\!\!\perp Z$ and $Y \perp\!\!\!\perp Z$ **then** $Z \in \mathbf{Z}_8$
  ▷ STEP 2: TEST FOR $\mathbf{Z}_4$
  else if $X \perp\!\!\!\perp Z$ and $X \not\perp\!\!\!\perp Z|Y$ **then** $Z \in \mathbf{Z}_4$
  ▷ STEP 3: TEST FOR $\mathbf{Z}_{5,7}$
  else if $Y \not\perp\!\!\!\perp Z$ and $Y \perp\!\!\!\perp Z|X$ **then** $Z \in \mathbf{Z}_{5,7}$
$\mathbf{Z}' \leftarrow \mathbf{Z}' \setminus \mathbf{Z}_4 \cup \mathbf{Z}_{5,7} \cup \mathbf{Z}_8$
  ▷ STEP 4: TEST FOR $\mathbf{Z}_{\text{POST}}$
if $|\mathbf{Z}_4| > 0$ **then**
  for all $Z \in \mathbf{Z}'$ **do**
    if $\exists\, Z_4$: $Z \not\perp\!\!\!\perp Z_4$ or $Z \perp\!\!\!\perp Z_4 | X \cup Y$ **then** $Z \in \mathbf{Z}_{\text{POST}}$
$\mathbf{Z}' \leftarrow \mathbf{Z}' \setminus \mathbf{Z}_{\text{POST}}$
  ▷ STEP 5: TEST FOR $\mathbf{Z}_{\text{MIX}}$
for all $Z \in \mathbf{Z}'$ **do**
  if $Y \not\perp\!\!\!\perp Z$ and $Y \perp\!\!\!\perp Z|X \cup \mathbf{Z}' \setminus Z$ **then**
    $Z \in \mathbf{Z}_{1,2,3,5} \in \mathbf{Z}_{\text{MIX}}$
$\mathbf{Z}' \leftarrow \mathbf{Z}' \setminus \mathbf{Z}_{\text{MIX}}$
  ▷ STEP 6: SPLIT $\mathbf{Z}_{\text{MIX}}$ BETWEEN $\mathbf{Z}_{1,5}$, $\mathbf{Z}_7$, $\mathbf{Z}_{\text{POST}}$
$\mathbf{Z}_{\text{MIX}} \leftarrow \mathbf{Z}_{\text{MIX}} \cup \mathbf{Z}_{5,7}$
if $|\mathbf{Z}_{\text{MIX}}| > 0$ and $|\mathbf{Z}'| > 0$ **then**
  for all $Z \in \mathbf{Z}'$ **do**
    if $\exists\, Z_{\text{MIX}} \in \mathbf{Z}_{\text{MIX}}$: $Z_{\text{MIX}} \perp\!\!\!\perp Z$ and $Z_{\text{MIX}} \not\perp\!\!\!\perp Z|X$ **then**
      $Z \in \mathbf{Z}_1$, $Z_{\text{MIX}} \in \mathbf{Z}_{1,5} \notin \mathbf{Z}_{\text{MIX}}$
    else $Z \in \mathbf{Z}_{\text{POST}}$
  for all $Z_{\text{MIX}} \in \mathbf{Z}_{\text{MIX}}$ **do**
    if $\exists\, Z_{1,5}$: $Z_{1,5} \perp\!\!\!\perp Z_{\text{MIX}}$ **then** $Z_{\text{MIX}} \in \mathbf{Z}_1$
    else $Z_{\text{MIX}} \in \mathbf{Z}_{\text{POST}}$
if $|\mathbf{Z}_{1,5}| > 0$ **then** $\mathbf{Z}_7 \leftarrow \mathbf{Z}_{5,7}$
  ▷ STEP 7: FINALIZE $\mathbf{Z}_1$ AND $\mathbf{Z}_5$
if $|\mathbf{Z}_{1,5}| > 0$ and $|\mathbf{Z}_1| > 0$ **then**
  for all $Z_{1,5} \in \mathbf{Z}_{1,5}$ **do**
    if $\exists\, Z_1 \in \mathbf{Z}_1$: $Z_{1,5} \not\perp\!\!\!\perp Z_1$ **then** $Z_{1,5} \in \mathbf{Z}_1$
    else $Z_{1,5} \in \mathbf{Z}_5$
if $|\mathbf{Z}_5| > 0$ **then**
  for $Z_5 \in \mathbf{Z}_5$ **do**
    if $Z_5 \not\perp\!\!\!\perp X | \mathbf{Z}_5 \cup \mathbf{Z}_{\text{POST}} \setminus Z_5$ **then**
      $Z_5 \in \mathbf{Z}_{5 \in adj(X)}$ and $\mathbf{Z}_1$ is a VAS
$\{\text{not identifiable}\} \leftarrow \mathbf{Z} \notin \mathbf{Z}_1, \mathbf{Z}_4, \mathbf{Z}_5, \mathbf{Z}_7, \mathbf{Z}_8, \mathbf{Z}_{\text{POST}}$
return Partitions of $\mathbf{Z}$ and $\{\text{not identifiable}\}$.

LDP learns partitions directly

High-Level Overview Here, we describe the basic logic of Algorithm 1 in plain English.

Step 1 $\mathbf{Z}_8$ discovered with knowledge of $\{X, Y, Z\}$ only.

Step 2 $\mathbf{Z}_4$ discovered with knowledge of $\{X, Y, Z\}$ only.

Step 3 $\mathbf{Z}_7$ discovered with knowledge of $\{X, Y, Z\}$ only. $\mathbf{Z}_5$ might also be discovered for some graphical structures (e.g., when $|\mathbf{Z}_1| = 0$).

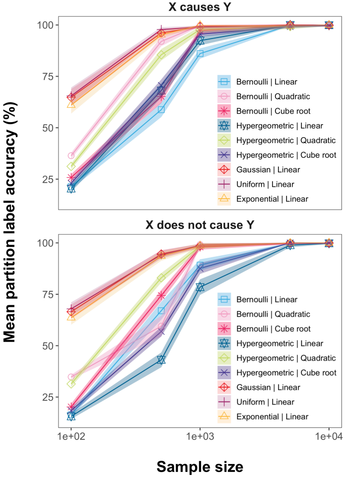
Step 4 A fraction of $\mathbf{Z}_{\text{POST}}$ is discovered, providing complete knowledge of $\mathbf{Z}_6$ and partial knowledge of $\mathbf{Z}_2$ and $\mathbf{Z}_3$. This step leverages prior knowledge of $\mathbf{Z}_4$ that was obtained programmatically at Step 2.

Step 5 $\mathbf{Z}_{\text{MIX}}$ is temporarily aggregated, providing partial knowledge of $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{Z}_3$, and $\mathbf{Z}_5$. $\mathbf{Z}_{\text{MIX}}$ is a transient superset that is used to differentiate $\mathbf{Z}_1$ and $\mathbf{Z}_5$ from $\mathbf{Z}_{\text{POST}}$ in Step 6.

Step 6 Knowledge of $\mathbf{Z}_{\text{POST}}$ is complete. $\mathbf{Z}_{\text{MIX}}$ is fully disaggregated, providing final partition labels for some members and moving others to superset $\mathbf{Z}_{1,5}$. At this stage, we also finalize our knowledge of $\mathbf{Z}_7$. By Line 19, all members of $\mathbf{Z}_5$ have been placed in $\mathbf{Z}_{1,5}$. By Line 22, $\mathbf{Z}_1$ that are adjacent to Y have been uniquely identified.

Step 7 $\mathbf{Z}_1$ and $\mathbf{Z}_5$ are fully differentiated from each other. This step tests whether a member of superset $\mathbf{Z}_{1,5}$ is marginally dependent on known members of $\mathbf{Z}_1$. All previously known members of $\mathbf{Z}_1$ are adjacent to Y. $\mathbf{Z}_1$ that are left to be discovered are those with indirect active paths to Y. Even when C1 is violated, no Z_5 will ever be dependent on a Z_1 that is directly adjacent to Y. However, all members of $\mathbf{Z}_1$ are marginally dependent on at least one Z_1 adjacent to Y. This step concludes by testing the $\mathbf{Z}_5$ criterion, which raises a warning when failed.

LDP partition correctness



LDP partition correctness

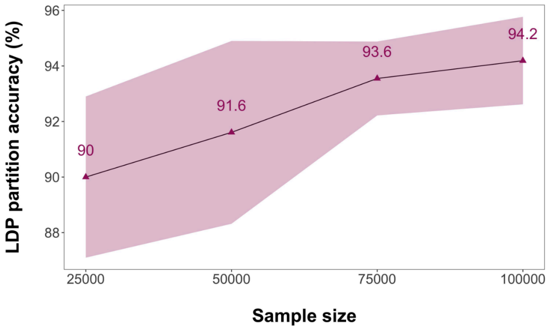
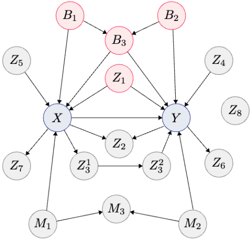


Figure G.1: LDP partition accuracy on the MILDEW benchmark. Mean accuracy was computed for 10 replicate samples from the ground truth DAG using `bnlearn` [Scutari, 2010]. We measure partition accuracy as the percent of partition labels that are consistent with ground truth. Independence was determined by chi-square tests ($\alpha = 0.005$). Shaded regions represent the 95% confidence interval. All experiments were run on a 2017 MacBook with 2.9 GHz Quad-Core Intel Core i7.

LDP partition correctness



LDP correctly partitions $98.7\%[97.6, 99.9]$ of linear-Bernoulli instantiations and $98.7\%[98.0, 99.4]$ of quadratic hypergeometric instantiations of this DAG (100 replicates each, $n = 20k$).

LDP partition correctness

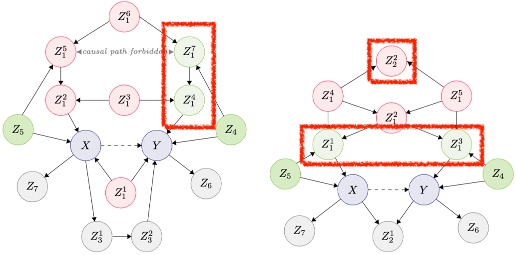
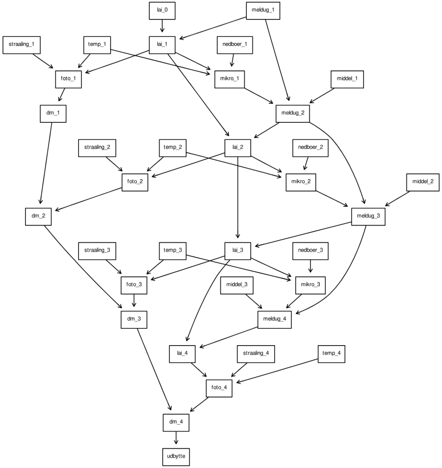


Figure D.1: Two DAGs that exemplify the behavior of LDP for valid adjustment set detection in the presence of inter-partition active paths. All red nodes will be placed in $\mathbf{Z}_1$ by LDP. All confounders for $\{X, Y\}$ that are colored green will be mislabeled due to their marginal dependence on Z_4 or Z_5.
Left: Per Lemma D.20, Z_1^1, Z_1^3 and Z_1^6 will be placed in $\mathbf{Z}_1$. Despite their marginal dependence on the only Z_5 in this structure, Z_1^2 and Z_1^5 will never be placed in $\mathbf{Z}_{\text{POST}}$ due to the presence of Z_1^1, as $Z_1^2 \perp\!\!\!\perp Z_1^1$ and $Z_1^5 \perp\!\!\!\perp Z_1^1$. Together, the confounders highlighted in red ($\{Z_1^1, Z_1^2, Z_1^3, Z_1^5, Z_1^6\}$) constitute a valid adjustment set that blocks all backdoor paths and contains no descendents of X. No causal path of either directionality is permissible between Z_1^5 and Z_1^7 per Proposition D.18. If this path were to contain a confounder analogous to Z_1^3, this would be permissible and this node would be placed in $\mathbf{Z}_1$ by LDP.
Right: This DAG contains a modified butterfly structure, which will be partially retained in $\mathbf{Z}_1$ ($\{Z_1^2, Z_1^4, Z_1^5\}$) while still blocking all backdoor paths. As there is only one Z_5 in this structure and no backdoor path whose members are marginally independent of Z_1^1, this confounder will be mislabeled as $\mathbf{Z}_{\text{POST}}$ at Step 6. This DAG also illustrates a case where a member of $\mathbf{Z}_2$ (Z_2^2) is placed in $\mathbf{Z}_1$. Inclusion of Z_2^2 does not violate the validity of the adjustment set returned by LDP, as this node is not a descendent of X and adjusting for $\{Z_1^2, Z_1^4, Z_1^5\}$ prevents collider bias.

The MILDEW benchmark



VAS discovery

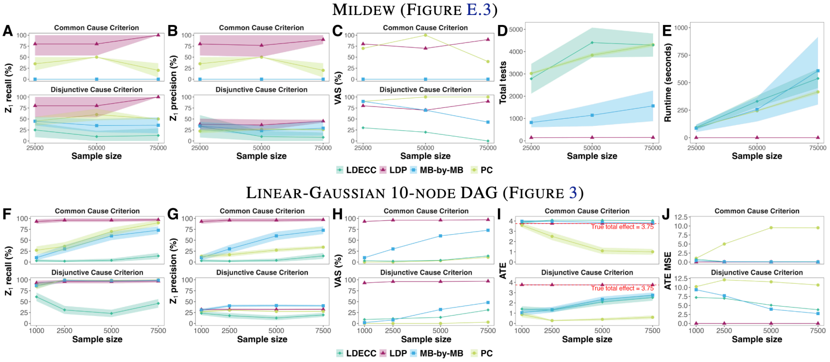


Figure 7: Baselines on MILDEW ($|\mathbf{Z}| = 31$) and a linear-Gaussian DAG ($|\mathbf{Z}| = 8$) (Tables G.8, G.9). Independence was determined with chi-square tests for MILDEW ($\alpha = 0.001$) and Fisher-z tests for the linear-Gaussian DAG ($\alpha = 0.01$). Results were averaged over 10 and 100 replicates per sample size for MILDEW and the linear-Gaussian DAG, respectively (95% confidence intervals in shaded regions). Precision and recall for $\mathbf{Z}_1$ identification were computed per adjustment set.

VAS with latent variables

Latent	VAS Exists	Z_5 Crit	% Valid
B_1	✓	✓	100
B_2	✓	✓	99
Z_{4a}	✓	✓	99
M_2	✓	✓	100
Z_{5a}	✓	✓	99
M_1	✓	✓	100
Z_1	✗	✗	0
B_3	✗	✗	0