Interpreting and using CPDAGs with background knowledge

Emilija Perković, Markus Kalisch, and Marloes H. Maathuis

Seminar for Statistics, ETH Zurich

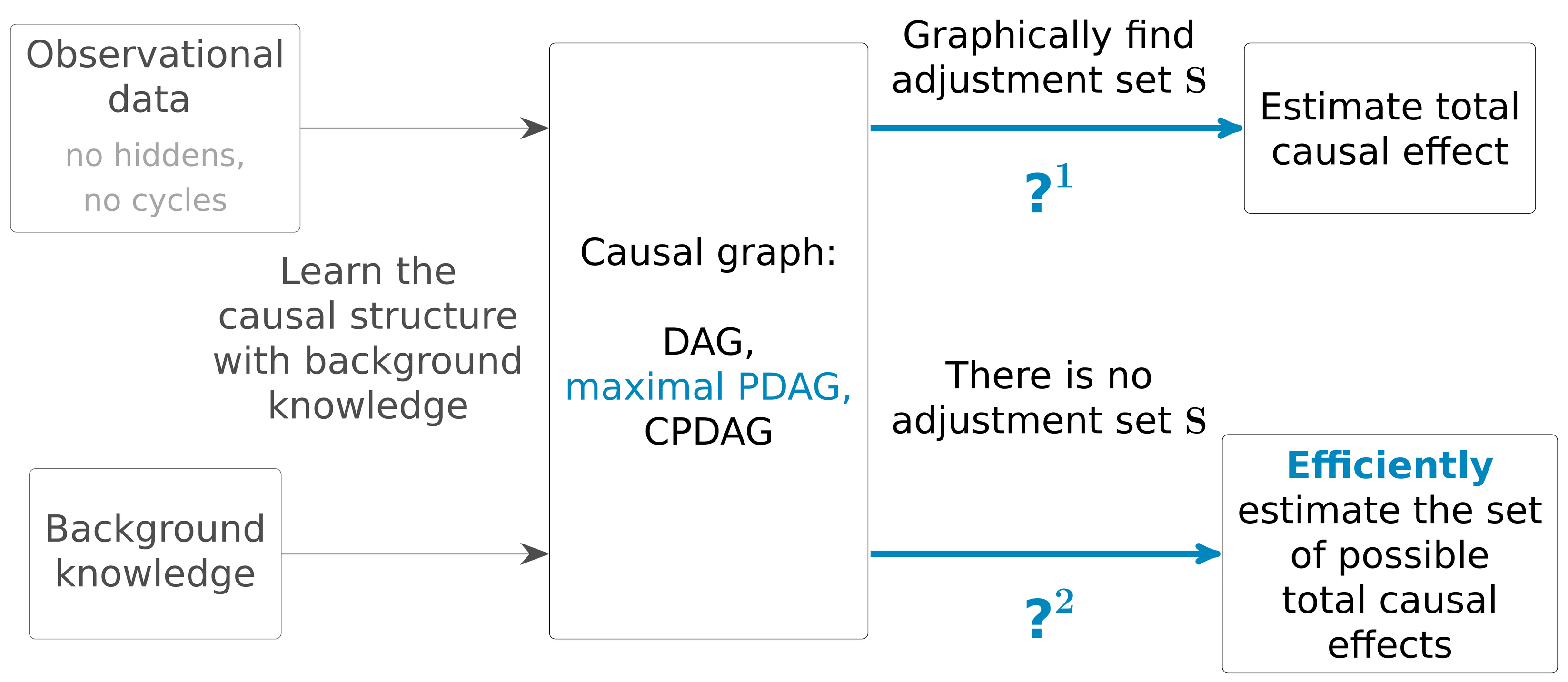
ETH zürich

Problem Total causal effects are often not identifiable from observational data.

Idea Use observational data and bg knowledge to identify more total causal effects.

Results On graphs obtained from observational data and background knowledge:
1. Adjustment criterion for estimating total causal effects.
2. Modified frameworks for estimating sets of possible total causal effects.
3. Implemented and modified algorithms in R package `pcalg` [5].

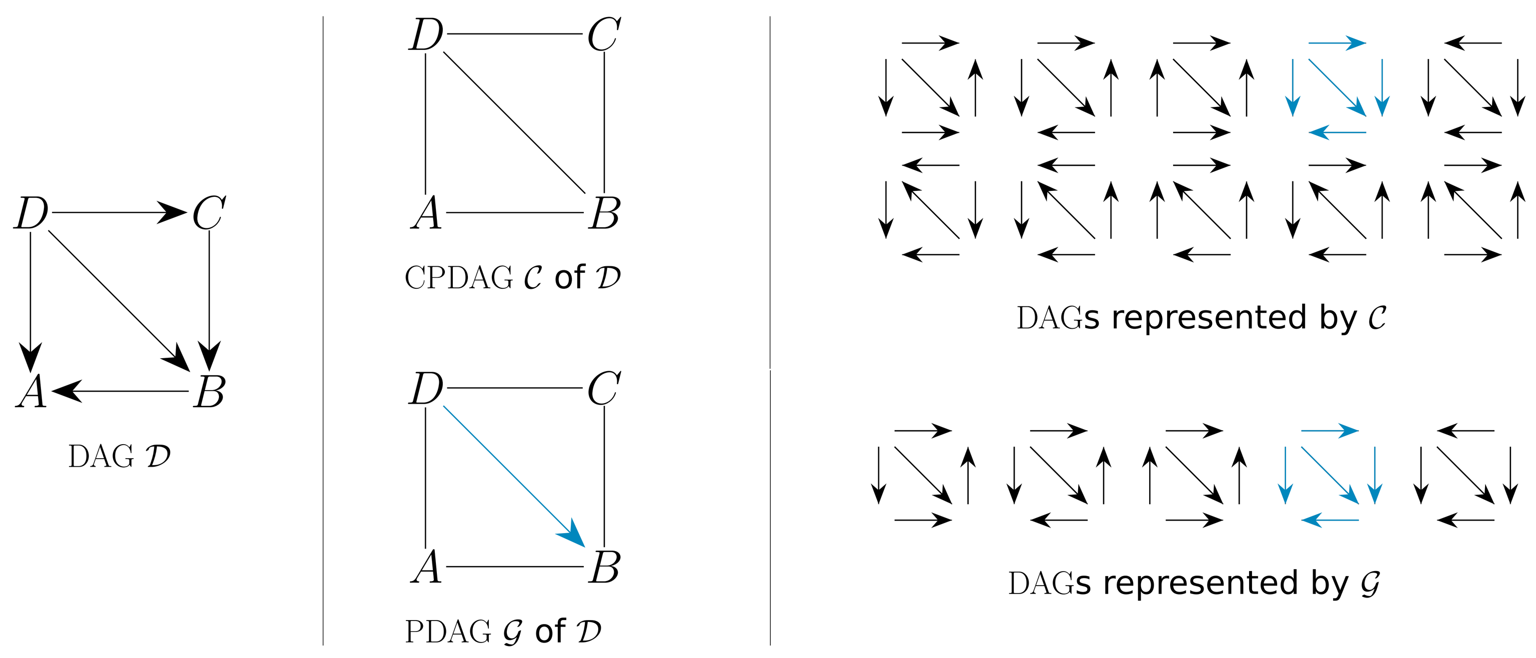
Framework



Sources of background knowledge:

- Applications - Expert knowledge, previous studies etc. [6, 8]
- Using a mix of observational and interventional data [3, 9]
- Model restrictions [4, 2, 1]

Adding background knowledge



Adjustment set

- A probability density f is compatible with the causal DAG $\mathcal{D}$ if:
  $f(\mathbf{v}) = \prod_{j=1}^{p} f(x_j|pa(x_j,\mathcal{D}))$ and $f(\mathbf{v}|do(\mathbf{x})) = \prod_{X_j \in \mathbf{V}\setminus\mathbf{X}} f(x_j|pa(x_j,\mathcal{D}))$.

- $\mathbf{S}$ is an adjustment set relative to $(\mathbf{X},\mathbf{Y})$ in causal $\mathcal{D}$ if for any f compatible with $\mathcal{D}$:

$$f(\mathbf{y}|do(\mathbf{x})) = \begin{cases} f(\mathbf{y}|\mathbf{x}) & \text{if } \mathbf{S} = \emptyset, \\ \int_{\mathbf{S}} f(\mathbf{y}|\mathbf{x},\mathbf{s}) f(\mathbf{s}) d\mathbf{s} = E_{\mathbf{S}}\{f(\mathbf{y}|\mathbf{x},\mathbf{s})\} & \text{otherwise.} \end{cases}$$

Result 1: Adjustment criterion

Theorem [7] $\mathbf{S}$ is an adjustment set relative to $(\mathbf{X},\mathbf{Y})$ and $\mathcal{G}$ iff

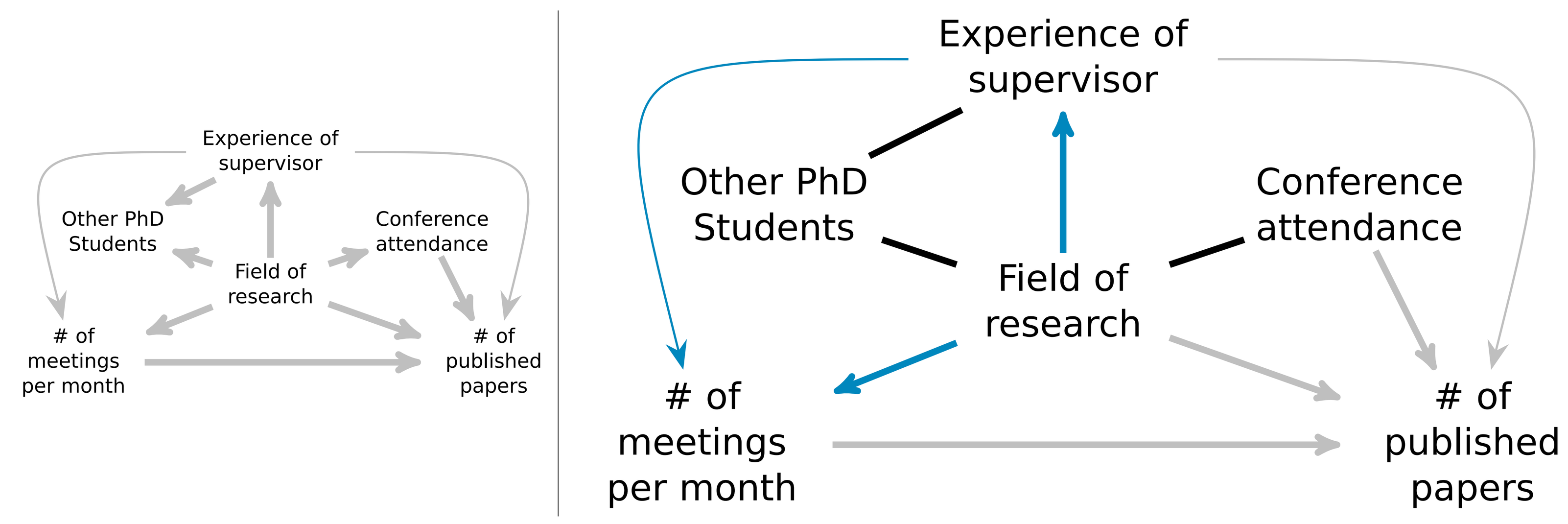
(Amenability) $\mathcal{G}$ is **b-amenable** relative to $(\mathbf{X},\mathbf{Y})$.

(Forbidden set) $\mathbf{S}$ does not contain nodes in **b-**$\mathrm{Forbidden}(\mathbf{X},\mathbf{Y},\mathcal{G})$.

(Blocking) $\mathbf{S}$ blocks all **proper b-non-causal definite status** paths from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$.

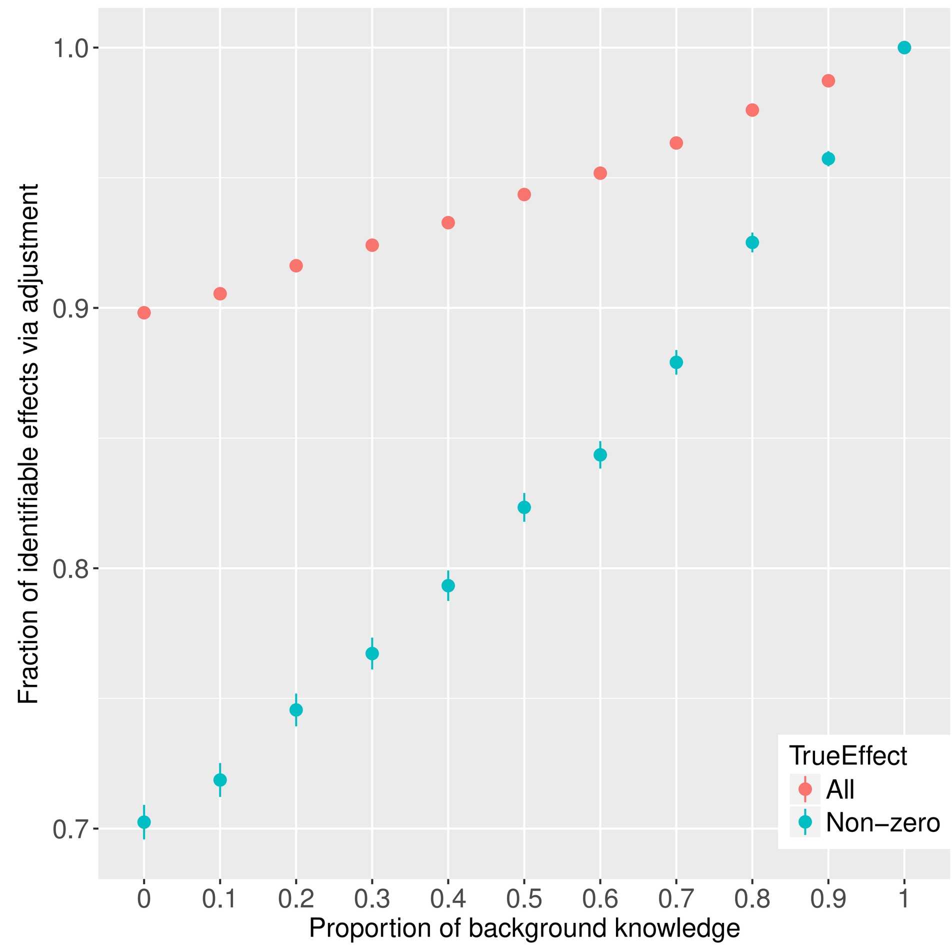
In a linear setting the total causal effect of X on Y is the linear regression coefficient of X in the regression $Y \sim X + \mathbf{S}$.

Example



- X = # of meetings per month, Y = # of published papers.
  $\mathrm{b\text{-}Forbidden}(X,Y,\mathcal{G})$ = {# of published papers}.
  Adjustment sets: {Experience of supervisor, Field of research}...
- X = Experience of supervisor, Y = # of published papers.
  $\mathrm{b\text{-}Forbidden}(X,Y,\mathcal{G})$ = {# of meetings per month, # of published papers}.
  Adjustment sets: {Field of research}, {Field of research, Other PhD students}...

Does an adjustment set always exist?



Randomly sampled 20 000 DAGs:
$p \in \{20, 30, \dots, 100\}$,
$E[N] \in \{3, 4, \dots, 10\}$.

X - randomly chosen,
Y - connected to X
and $Y \not\to X$.

Result 2: IDA and joint-IDA framework

Idea: Use parent sets of $\mathbf{X}$ to estimate all possible total causal effects of $\mathbf{X}$ on $\mathbf{Y}$.
Goal: Find all sets of parents of $\mathbf{X}$ in $\mathcal{G}$ *in an efficient way*.

For each set of possible parents $\mathbf{S}$ of $\mathbf{X}$ in graph $\mathcal{G}$:

IDA w/o background knowledge:

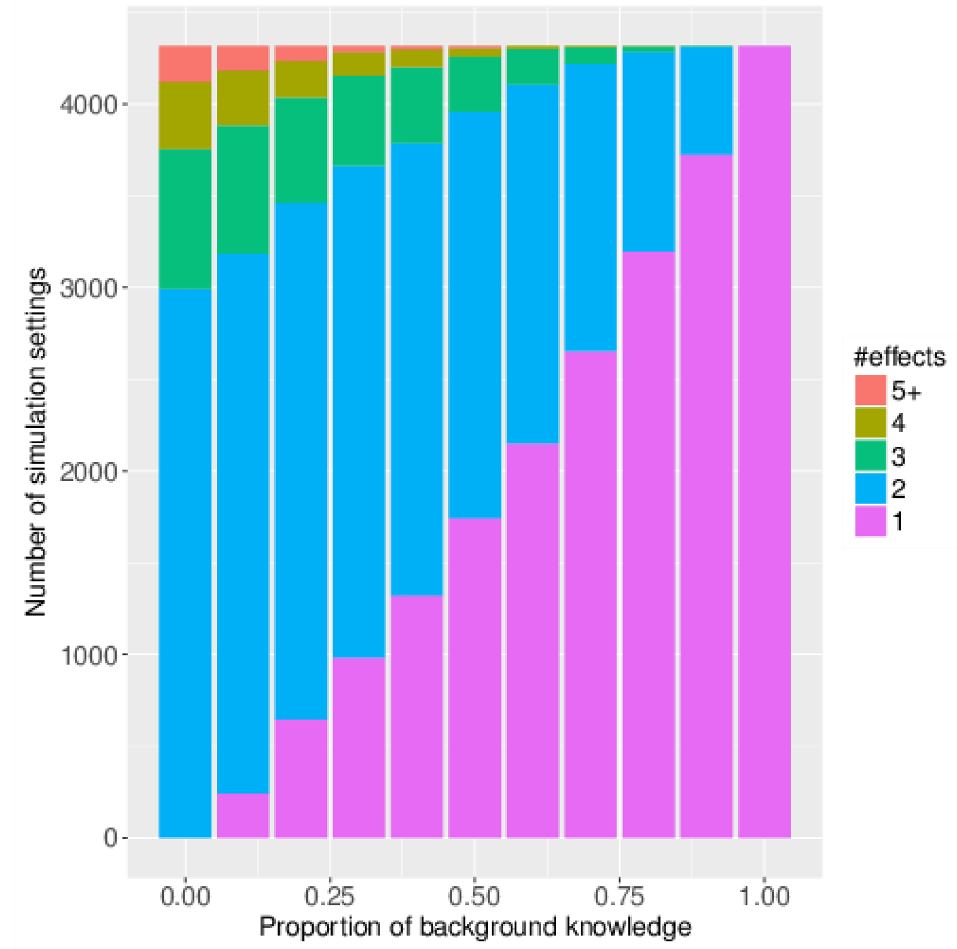
For each $S \in \mathbf{S}$, orient $S \to X$ in $\mathcal{G}$. Check for new v-structures.

IDA w background knowledge:

For each $S \in \mathbf{S}$, if possible, orient $S \to X$ and complete Meek (1995) rules, otherwise next set. For each $X - \bar{S}$ in $\mathcal{G}$ such that $\bar{S} \notin \mathbf{S}$, if possible, orient $X \to \bar{S}$ and complete Meek (1995) rules, otherwise next set.

Runtime:	Median	Mean	Max
IDA without bg	0.003	0.003	0.009
IDA with bg	0.003	0.016	4.881

Identifiability gain with bg knowledge



4315 out of 20 000 sampled DAGs:
$p \in \{20, 30, \dots, 100\}$,
$E[N] \in \{3, 4, \dots, 10\}$,
$n = 200$.

X - randomly chosen,
Y - connected to X
and $Y \not\to X$.

References

[1] M. Eigenmann, P. Nandy, and M. H. Maathuis. Structure learning of linear Gaussian structural equation models with weak edges. In *Proceedings of UAI*, 2017.

[2] J. Ernest, D. Rothenhäusler, and P. Bühlmann. Causal inference in partially linear structural equation models: identifiability and estimation. arXiv:1607.05980, 2016.

[3] A. Hauser and P. Bühlmann. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *J. Mach. Learn. Res.*, 2012.

[4] P. O. Hoyer, A. Hyvarinen, R. Scheines, P. L. Spirtes, J. Ramsey, G. Lacerda, and S. Shimizu. Causal discovery of linear acyclic models with arbitrary distributions. In *Proceedings of UAI*, 2008.

[5] M. Kalisch, M. Mächler, D. Colombo, M. H. Maathuis, and P. Bühlmann. Causal inference using graphical models with the R package pcalg. *J. Stat. Softw.*, 2012.

[6] C. Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of UAI 1995*, pages 403–410, 1995.

[7] E. Perković, M. Kalisch, and M. H. Maathuis. Interpreting and using CPDAGs with background knowledge. In *Proceedings of UAI*, 2017.

[8] R. Scheines, P. Spirtes, C. Glymour, C. Meek, and T. Richardson. The TETRAD project: constraint based aids to causal model specification. *Multivar. Behav. Res.*, 1998.

[9] Y. Wang, L. Solus, K. D. Yang, and C. Uhler. Permutation-based causal inference algorithms with interventions. arXiv:1705.10220, 2017.