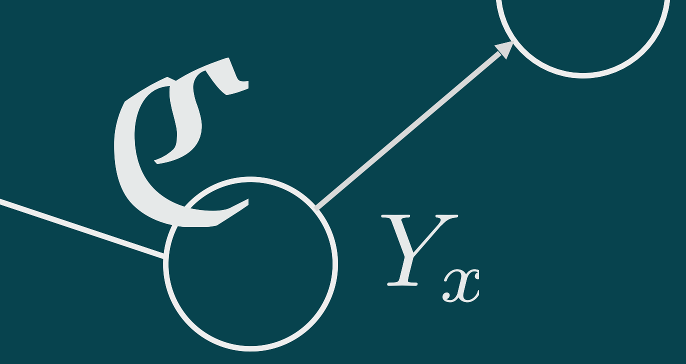
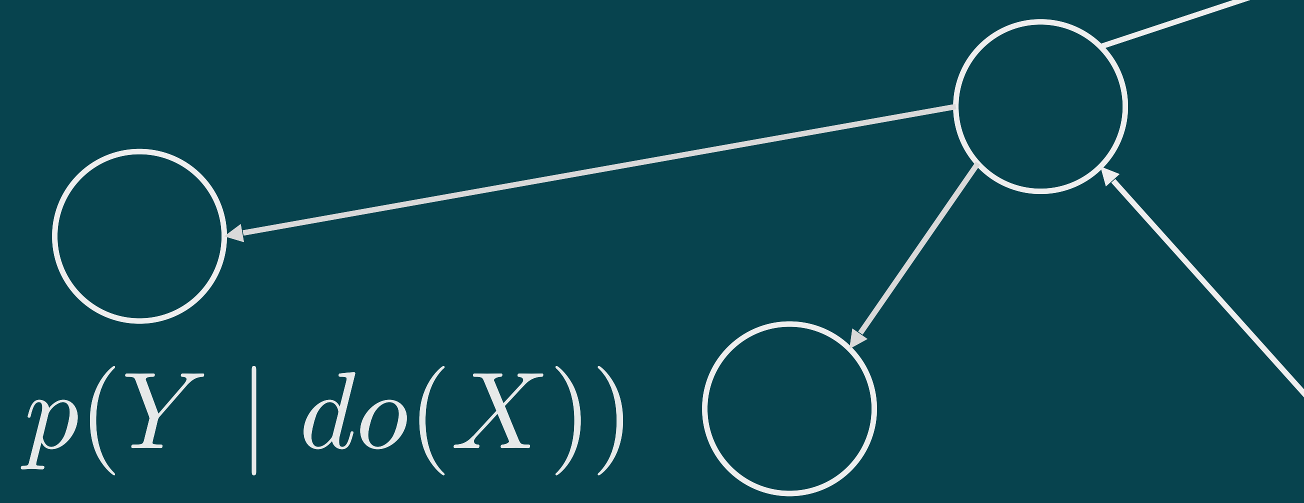
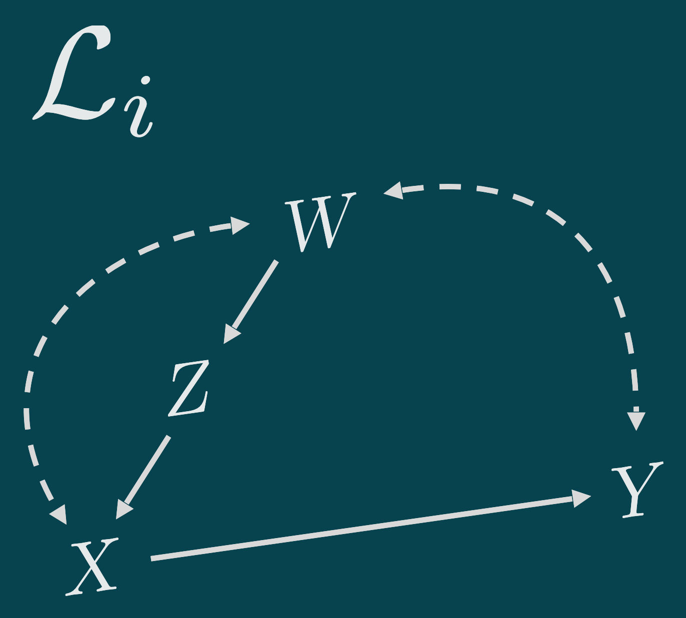




Causal Discovery

Current State of the Art



ESSAI & ACAI 2023
LJUBLJANA, SLOVENIA

Machines Climbing Pearl's Ladder of Causation

Causal Discovery - Current State of Art

- Learning from a combination of observational and experimental data
- Learning dynamic causal models from time-series data
- Extension to cyclic causal models

Learning from Observational and Experimental Data

Under causal sufficiency:

- Gonçalo Rui Alves Faria, Andre Martins, Mario A. T. Figueiredo. Differentiable Causal Discovery Under Latent Interventions. Proceedings of the First Conference on Causal Learning and Reasoning, PMLR 177:253-274, 2022.

Allowing unmeasured confounders:

- Jaber, A., Kocaoglu, M., Shanmugam, K., Bareinboim, E. Causal Discovery from Soft Interventions with Unknown Targets: Characterization & Learning. In Advances in Neural Information Processing Systems 2020 (Link).

- Kocaoglu, M., Jaber, A., Shanmugam, K., Bareinboim, E. Characterization and Learning of Causal Graphs with Latent Variables from Soft Interventions. In Proceedings of the 33rd Annual Conference on Neural Information Processing Systems. 2019 (Link).

Learning Dynamic Causal Models from Time-Series Data

SVAR-FCI — Malinsky, D. and Spirtes, P. (2018).

- Assume the data-generating process is a structural vector autoregression (SVAR) with latent components.
- Allow for both "contemporaneous" causal relations and arbitrary unmeasured ("latent") processes influencing observed variables
- Algorithm is a modification of the FCI algorithm to search for an equivalence class of Dynamic Ancestral Graphs.
    1. respects the time order of the variables by restricting possible conditioning sets to variables in the "present" or "past" time slices and prohibiting orientations backwards in time.
    2. enforces the repeating structure of the underlying dynamic DAG in both determining adjacencies and orientations.

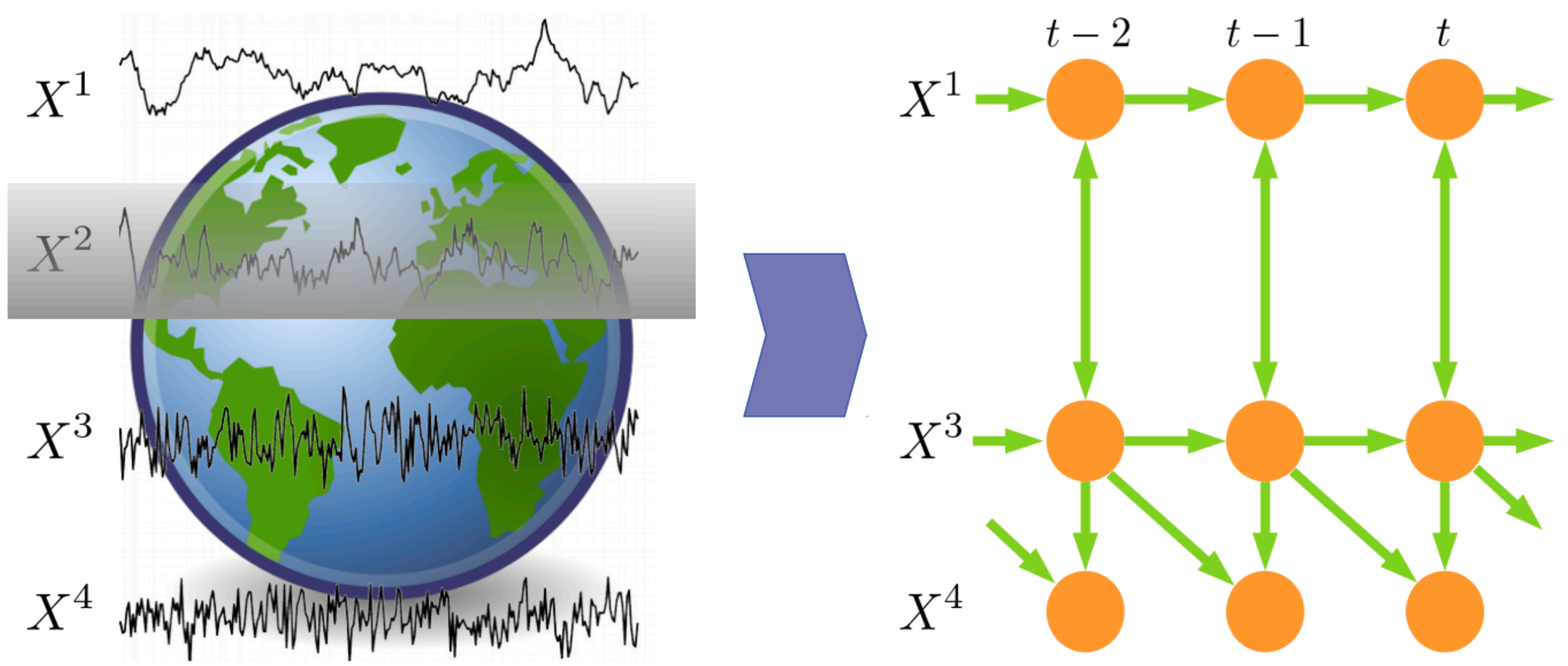
Malinsky, D. and Spirtes, P. (2018). *Causal Structure Learning from Multivariate Time Series in Settings with Unmeasured Confounding.* Proceedings of 2018 ACM SIGKDD Workshop on Causal Discovery, volume 92 of Proceedings of Machine Learning Research, pages 23–47, London, UK. PMLR.

Learning Dynamic Causal Models from Time-Series Data

Nice Overview: Runge, J., Gerhardus, A., Varando, G. et al. Causal inference for time series. Nat Rev Earth Environ (2023). (Link)

Under causal sufficiency:

- **PCMCI:** J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, D. Sejdinovic, Detecting and quantifying causal associations in large nonlinear time series datasets. Sci. Adv. 5, eaau4996 (2019). (Link)



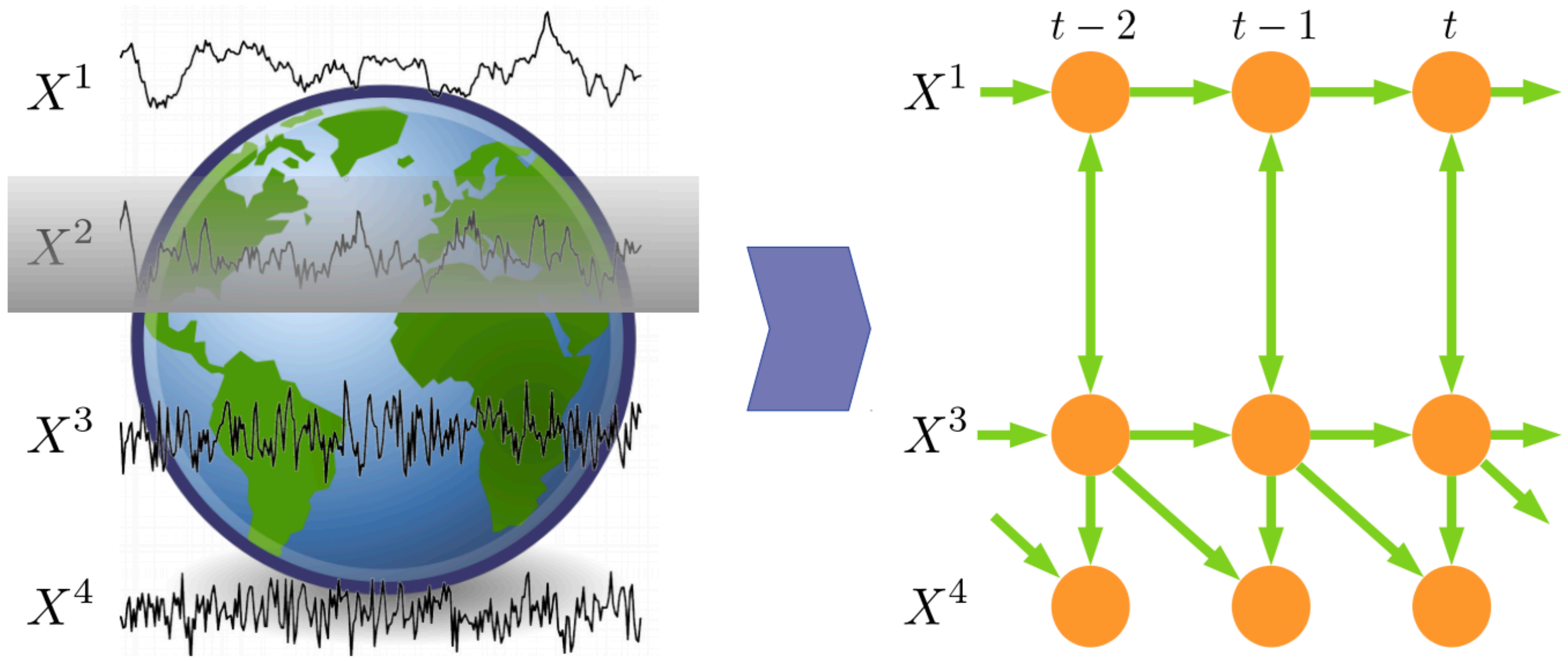
- **PCMCI+:** J. Runge (2020): Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence, UAI 2020,Toronto, Canada, 2019, AUAI Press, 2020. (Link)

Learning Dynamic Causal Models from Time-Series Data

Nice Overview: Runge, J., Gerhardus, A., Varando, G. et al. Causal inference for time series. Nat Rev Earth Environ (2023). (Link)

Allowing unmeasured confounders:

- **Latent PCMCI (L-PCMCI):** Gerhardus, A. & Runge, J. High-recall causal discovery for autocorrelated time series with latent confounders Advances in Neural Information Processing Systems, 2020, 33. (Link)



Learning Dynamic Causal Models from Time-Series Data

Method	Assumptions	Output
	(in addition to Causal Markov Condition and Faithfulness)	
PCMCI	Causal stationarity, no contemporaneous causal links, no hidden variables	Directed lagged links, undirected contemporaneous links (for tau_min=0)
PCMCIplus	Causal stationarity, no hidden variables	Directed lagged links, directed and undirected contemp. links (Time series CPDAG)
LPCMCI	Causal stationarity	Time series PAG

https://github.com/jakobrunge/tigramite

Extension to Cyclic Causal Models

Surprisingly, the output of the Fast Causal Inference (FCI) algorithm is correct if it is applied to observational data generated by a system that involves feedback.

Joris M. Mooij, Tom Claassen. *Constraint-Based Causal Discovery using Partial Ancestral Graphs in the presence of Cycles.*

Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), PMLR 124:1159-1168, 2020.

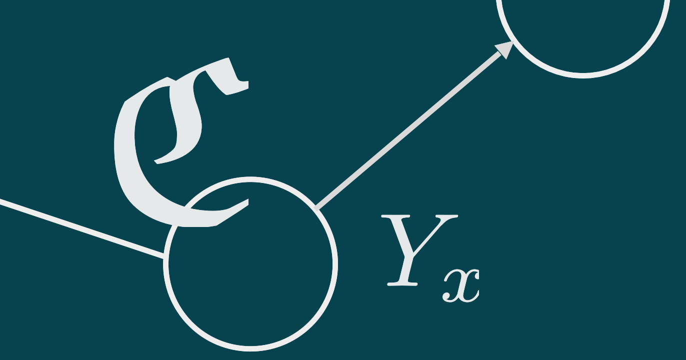
Extension to Cyclic Causal Models

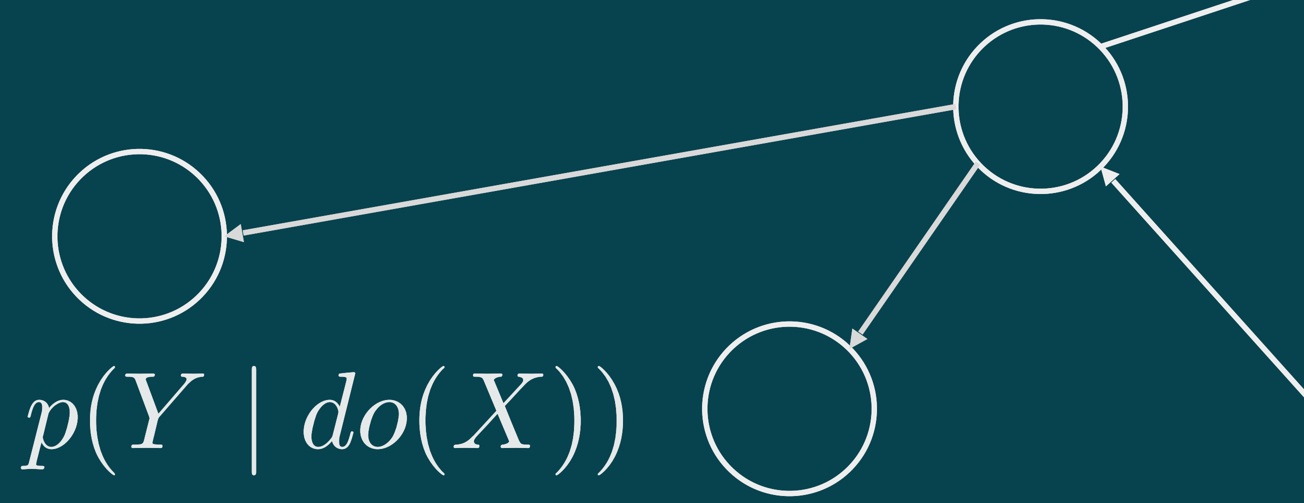
- Surprisingly, the output of the **Fast Causal Inference (FCI) algorithm** is correct if it is applied to observational data generated by a system that involves feedback.

- In other words, it learns a more general class of directed mixed graphs (DMGs), where feedback loops may be present.

- The work focuses on **Simple SCMs**, which admit (sufficiently weak) cyclic interactions but retain many of the convenient properties of acyclic SCMs. In particular, they satisfy the σ-separation Markov property and there is not selection bias.

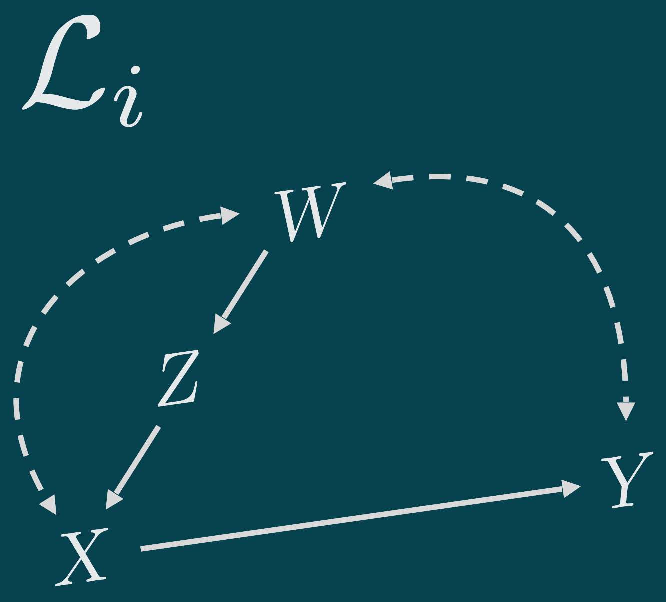
Joris M. Mooij, Tom Claassen. *Constraint-Based Causal Discovery using Partial Ancestral Graphs in the presence of Cycles.* Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI), PMLR 124:1159-1168, 2020.





Causal Identification

Current State of the Art



Causal Identification - Current State of Art

- Identification from observational and experimental data
- Identification in partially understood domains (C-DAGs)
- Partial identification of (bounds on) causal effects
- Identification in Markov Equivalence Classes (PAGs)

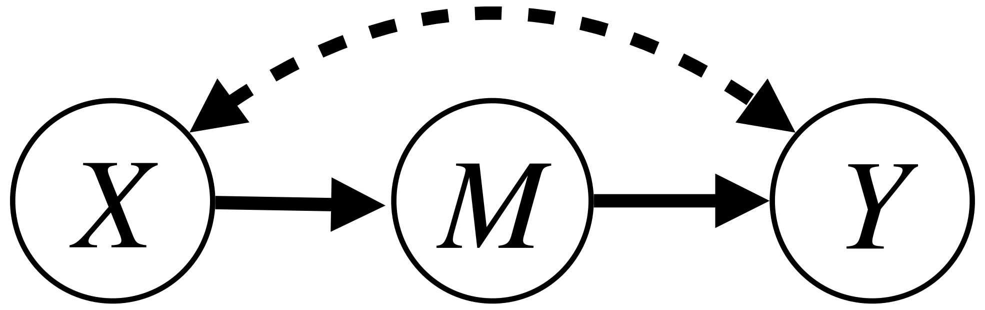
Classical Causal Effect Identification

Identification from observational data

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 **Data**

$P(x, m, y)$ **Observational Distribution**

Inference Engine

Solution

yes / no

$$P(y \mid do(x)) = \sum_m P(m \mid x) \sum_{x'} P(y \mid m, x') P(x')$$

Interventional Distribution ← Available Distributions

- Tian, J. and Pearl, J. A General Identification Condition for Causal Effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002), pp. 567–573, Menlo Park, CA, 2002. AAAI Press/MIT Press.

Causal Inference and Data-Fusion

Causal inference and the data-fusion problem

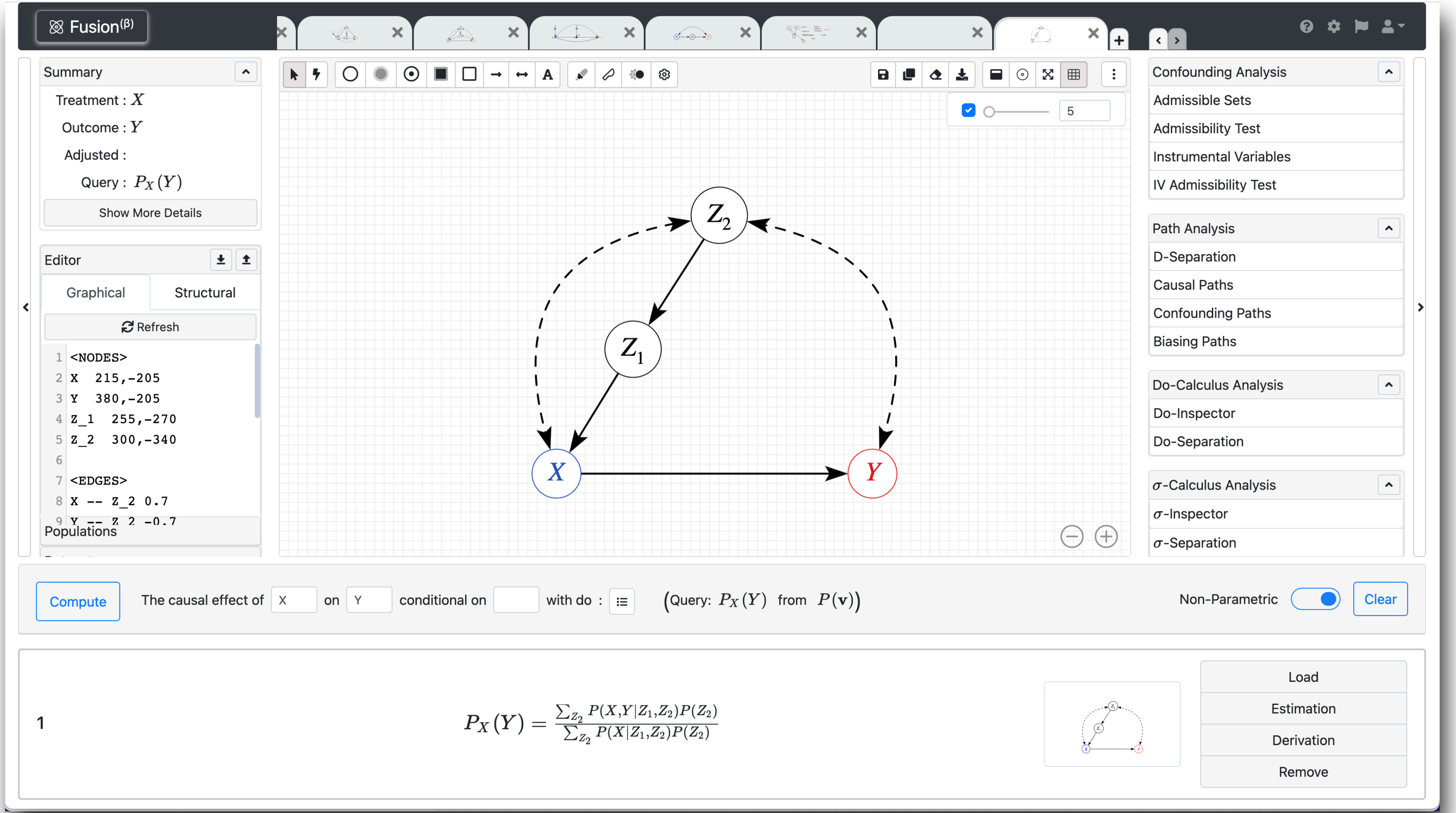
Elias Bareinboim[a,b,1] and Judea Pearl[a]

[a]Department of Computer Science, University of California, Los Angeles, CA 90095; and [b]Department of Computer Science, Purdue University, West Lafayette, IN 47907

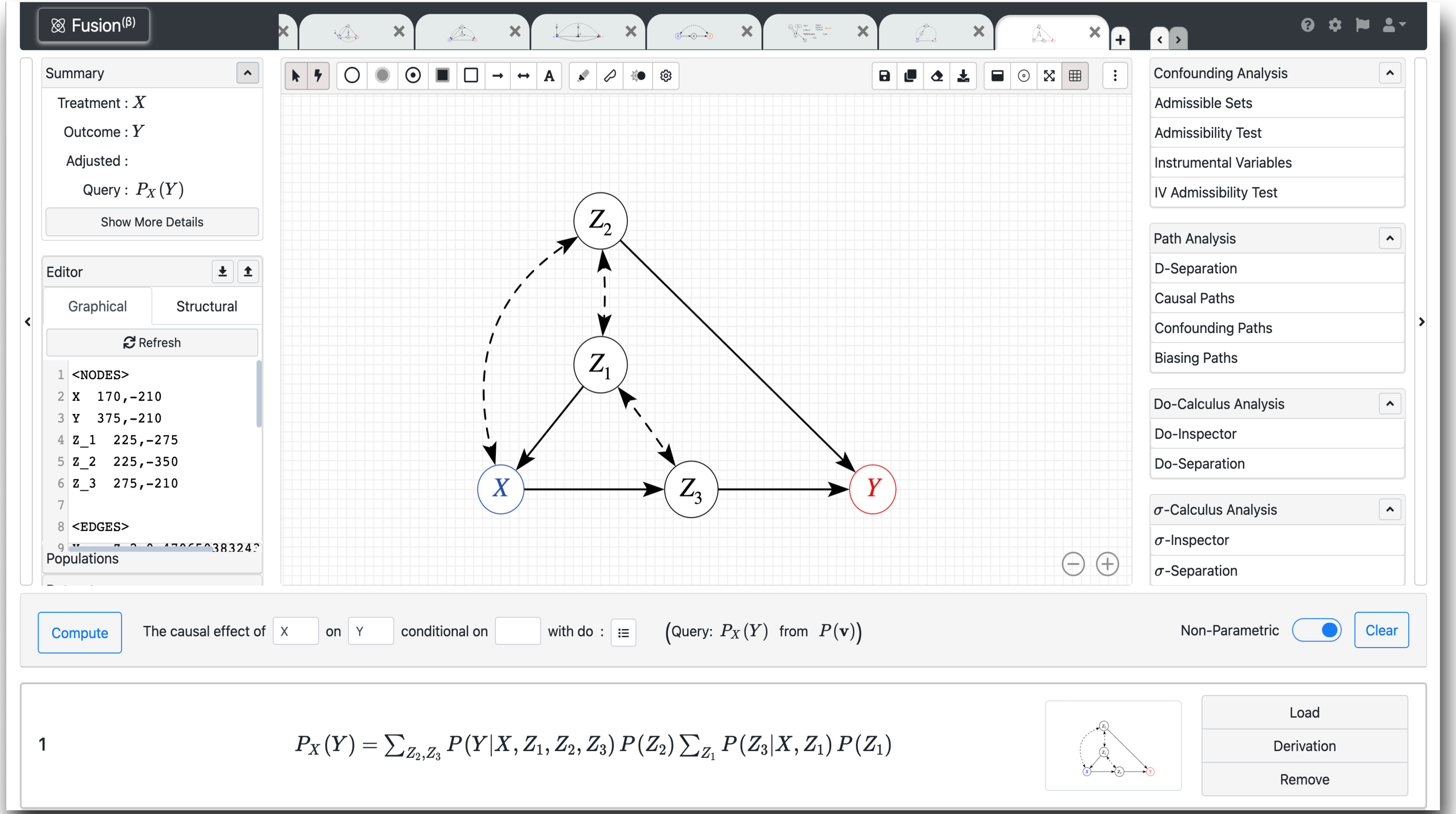
Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved March 15, 2016 (received for review June 29, 2015)

http://causalfusion.net

http://causalfusion.net



http://causalfusion.net



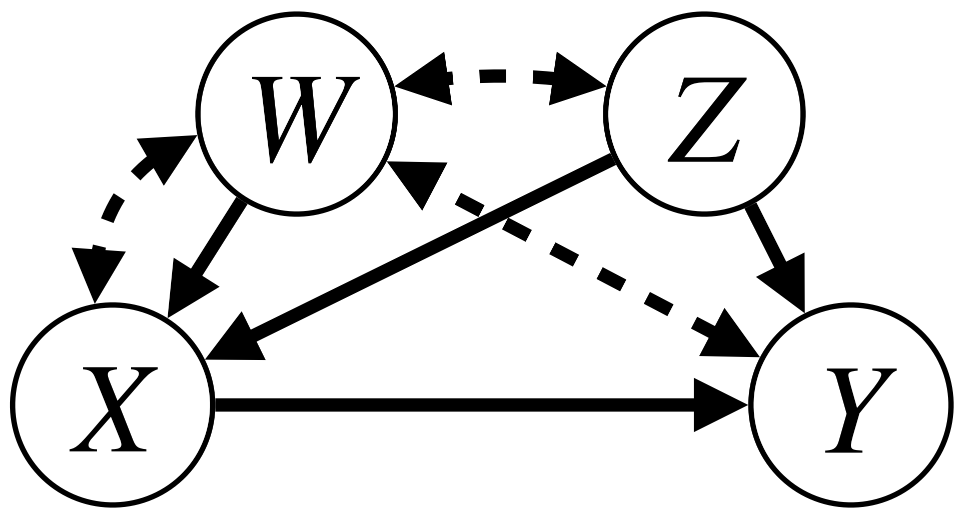
Classical Causal Effect Identification

Identification from observational data

1. **Query**

$P(y \mid do(x))$

2. **Causal Contraints**



3. **Data**

$P(x, y, z, w)$

Observational Distribution

Inference Engine

Solution

yes / no

Interventional Distribution ← Available Distributions 👎

- Tian, J. and Pearl, J. (**2002**) A General Identification Condition for Causal Effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI), pp. 567–573, Menlo Park, CA. AAAI Press/MIT Press.

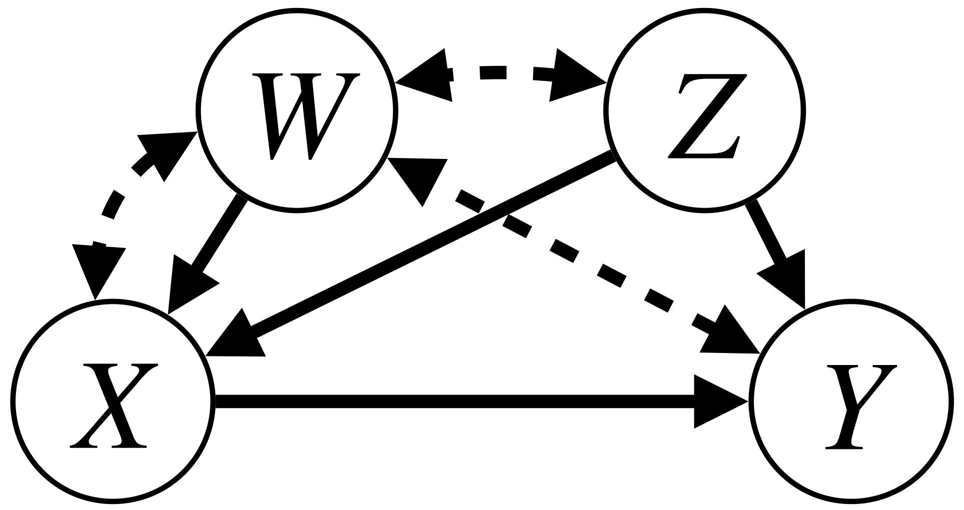
General Causal Effect Identification

Identification from observational and experimental data

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 **Data**

$P(x, y, z, w)$ **Observational Distribution**

$\boxed{P(x, y, z \mid do(w))}$ **Experimental Distribution**

Inference Engine

Solution

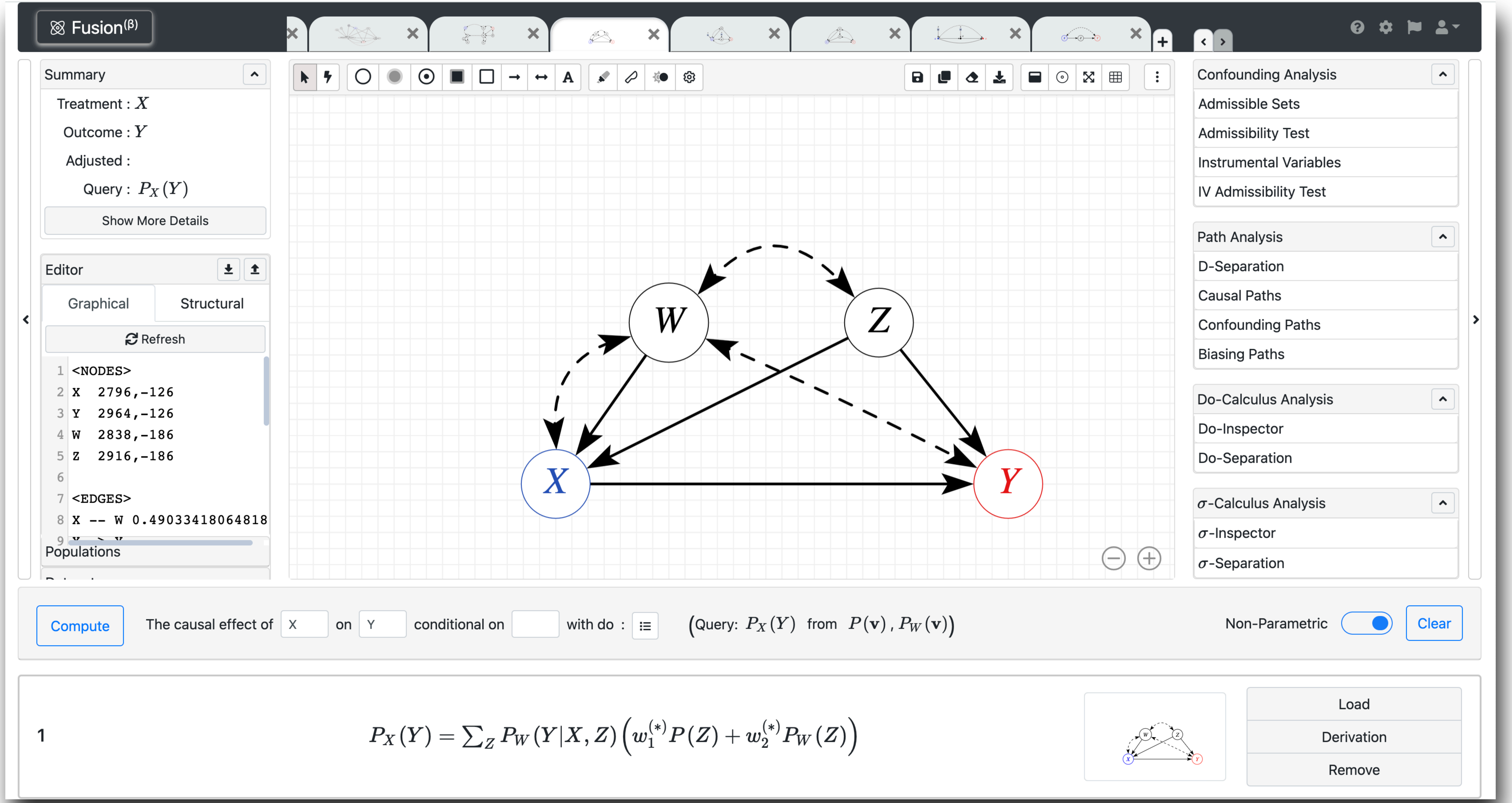
yes / no

$$P(y \mid do(x)) = \sum_z P(y \mid x, z, do(w))\big(\alpha_1 P(z) + \alpha_2 P(z \mid do(w))\big)$$

Interventional Distribution ← Available Distributions

- Lee, S., Correa, J., and Bareinboim, E. (**2019**). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 35, Tel Aviv, Israel. AUAI Press. [Link]

http://causalfusion.net



Identification from Observational and Experimental Data

Identification from observational and experimental data:

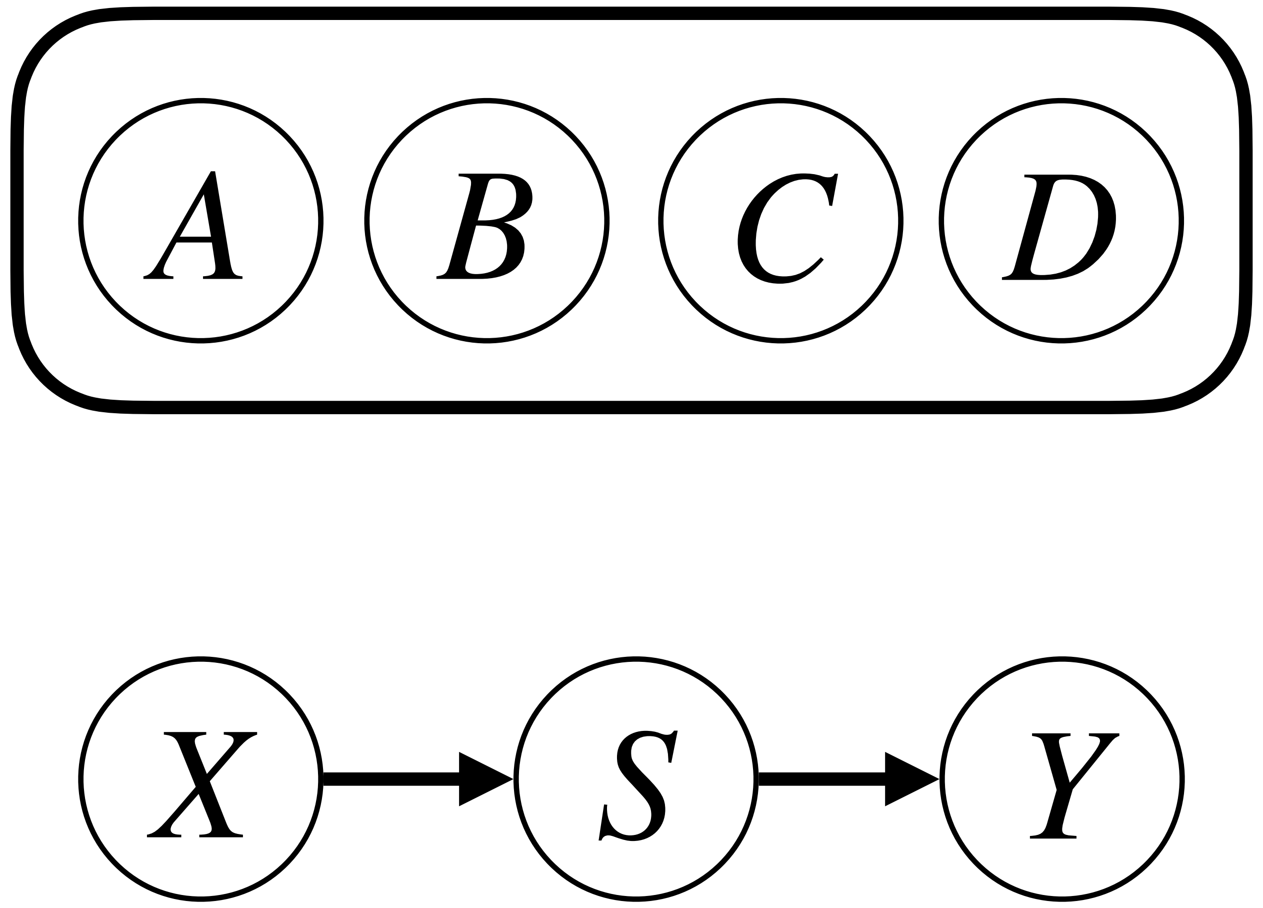
Lee, S., Correa, J., and Bareinboim, E. (**2019**). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 35, Tel Aviv, Israel. AUAI Press. Link

Identification of stochastic/soft (and possibly imperfect) interventions:

Correa, J. and Bareinboim, E. (2020). A calculus for stochastic interventions: Causal effect identification and surrogate experiments. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY. AAAI Press. Link

Identification in Partially Understood Systems

(A) Age
(B) Blood pressure
(C) Comorbidities
(D) Medication history
(X) Lisinopril
(S) Sleep Quality
(Y) Stroke

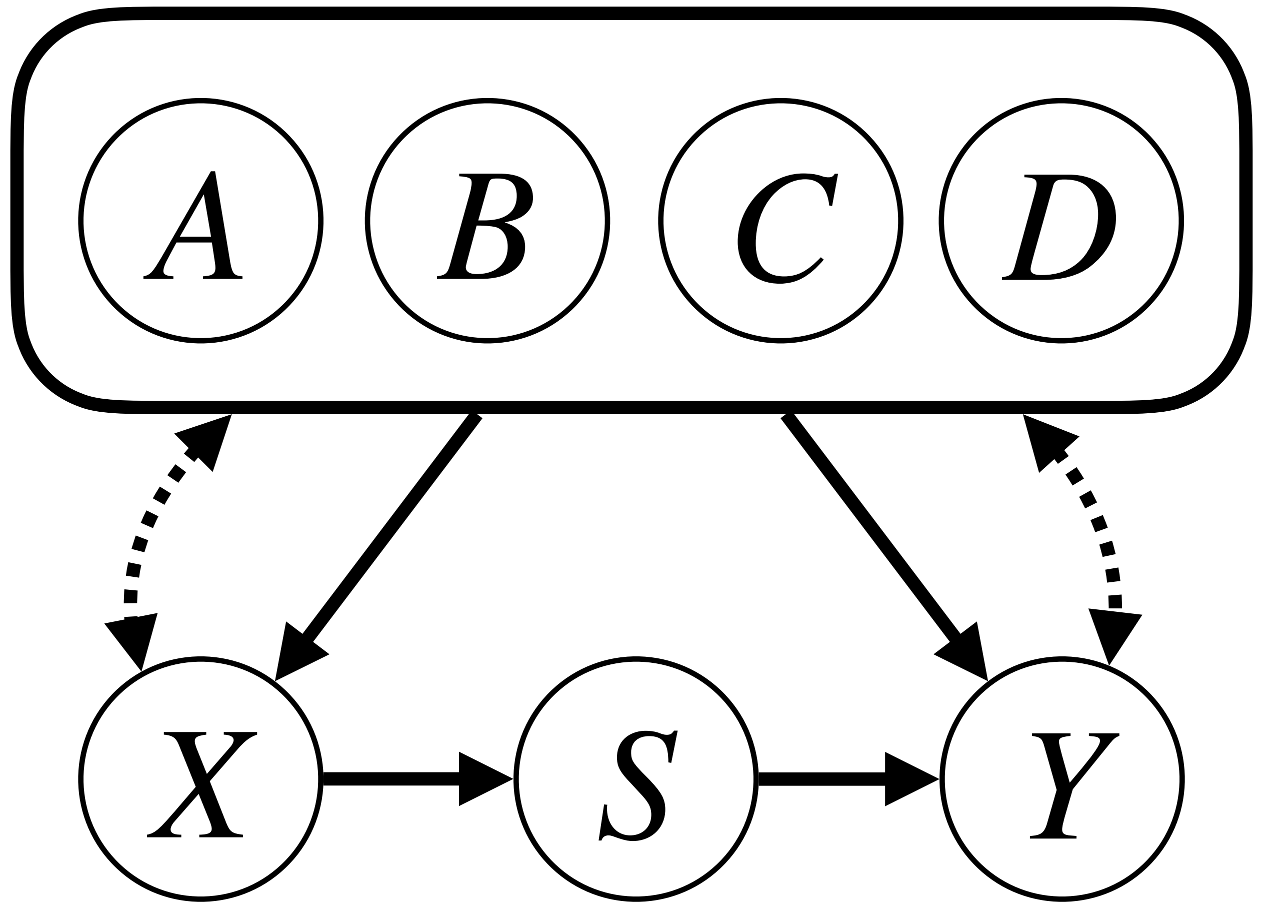


A causal diagram cannot be specified given the existing knowledge!

How can we identify $P(y \mid do(x))$ in this case?

Cluster Causal Diagrams (C-DAGs)

(A) Age
(B) Blood pressure
(C) Comorbidities
(D) Medication history
(X) Lisinopril
(S) Sleep Quality
(Y) Stroke



$\{\{X\}, \{S\}, \{Y\}, \{A, B, C, D\}\}$

A *cluster causal diagram* $G_{\mathbf{C}}$ over a given partition $\mathbf{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_k\}$ of $\mathbf{V}$ is compatible with a causal diagram G over $\mathbf{V}$ if **for every $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{C}$**:

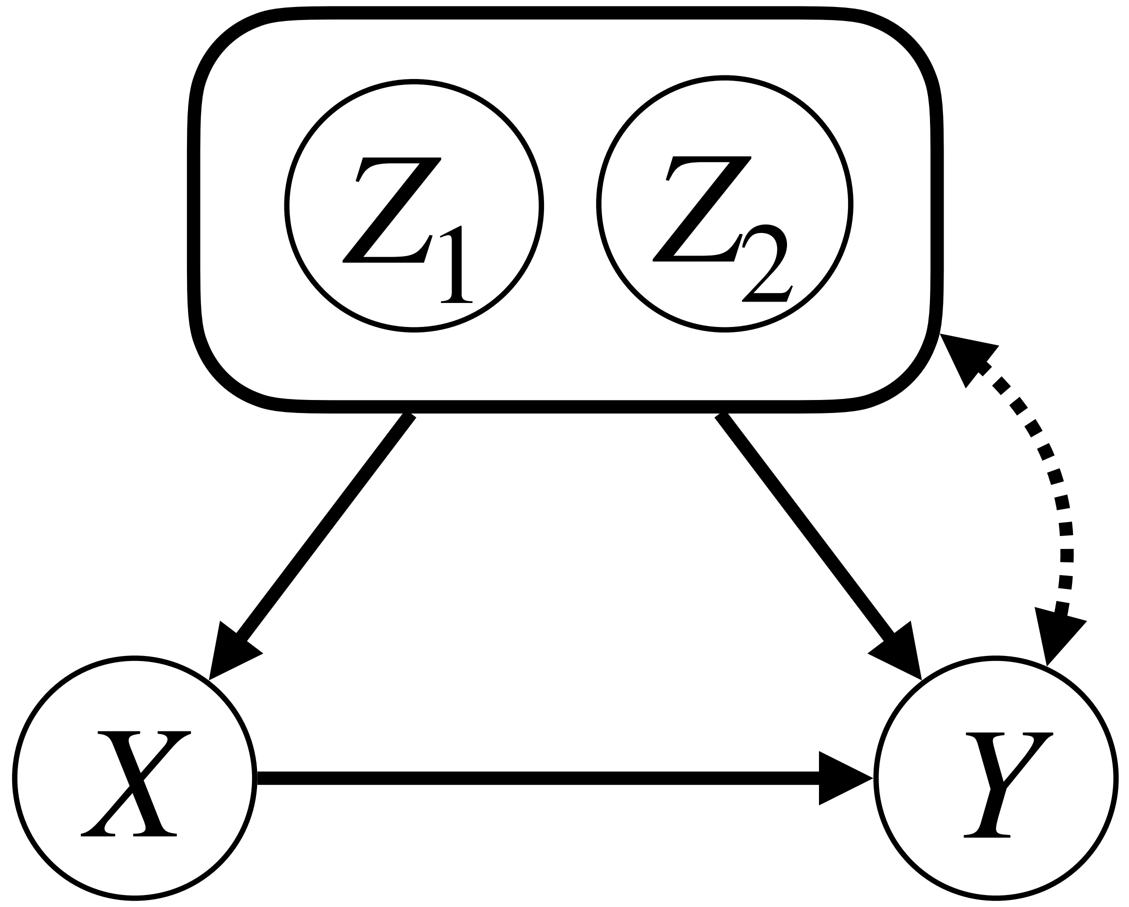
- $\mathbf{C}_i \rightarrow \mathbf{C}_j$ if $\exists V_i \in \mathbf{C}_i$ and $V_j \in \mathbf{C}_j$ such that $V_i \rightarrow V_j$
- $\mathbf{C}_i \dashleftarrow\dashrightarrow \mathbf{C}_j$ if $\exists V_i \in \mathbf{C}_i$ and $V_j \in \mathbf{C}_j$ such that $V_i \dashleftarrow\dashrightarrow V_j$

and $G_{\mathbf{C}}$ contains no cycles.

Effect Identifiabiliy given a C-DAG

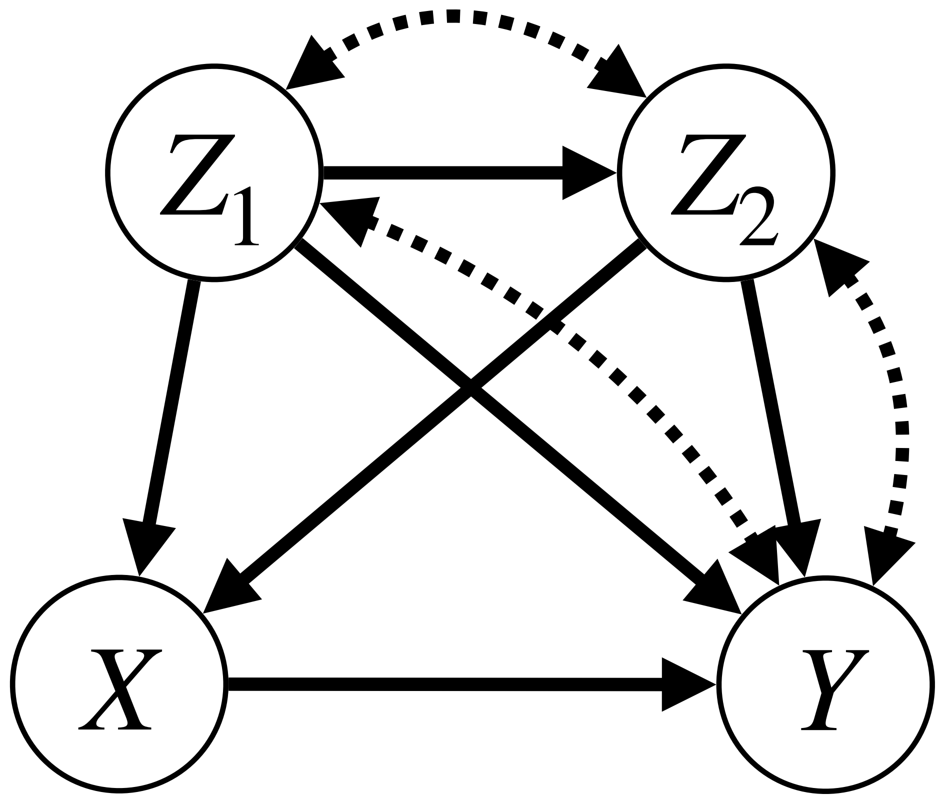
Simple evaluation of the **validity** of the Backdoor Criterion / Conditional Exchangeability

$G_{\mathbf{C}}$



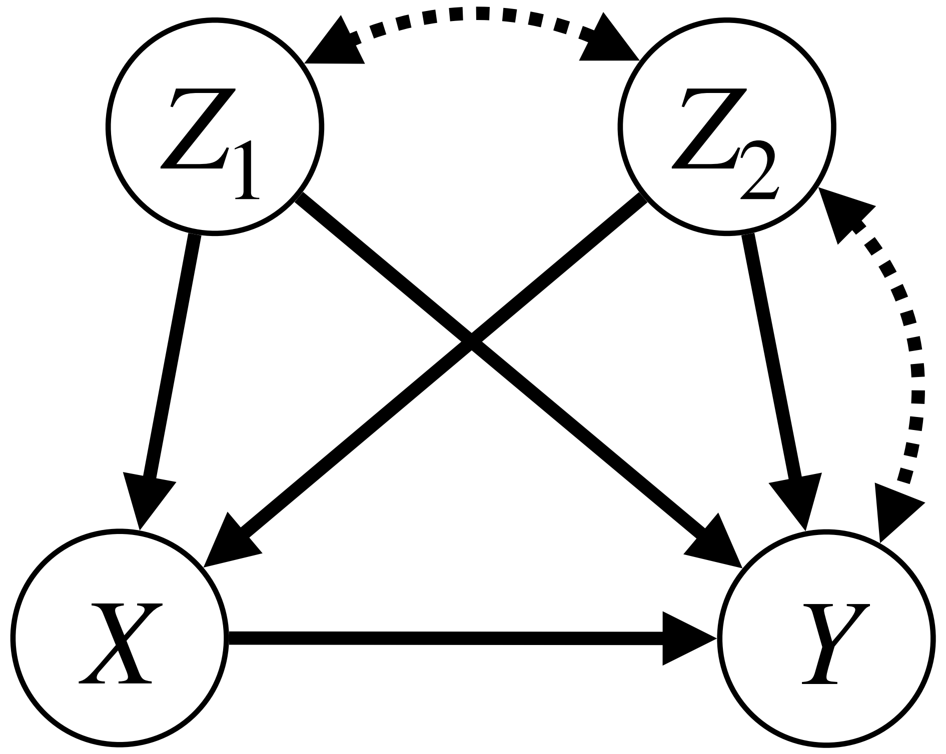
$$P(y \mid do(x)) = \sum_{\mathbf{z}} P\left(y \mid x, \mathbf{z}\right) P\left(\mathbf{z}\right)$$

G_1



$$P(y \mid do(x)) = \sum_{z_1, z_2} P(y \mid x, z_1, z_2) P(z_1, z_2)$$

G_2

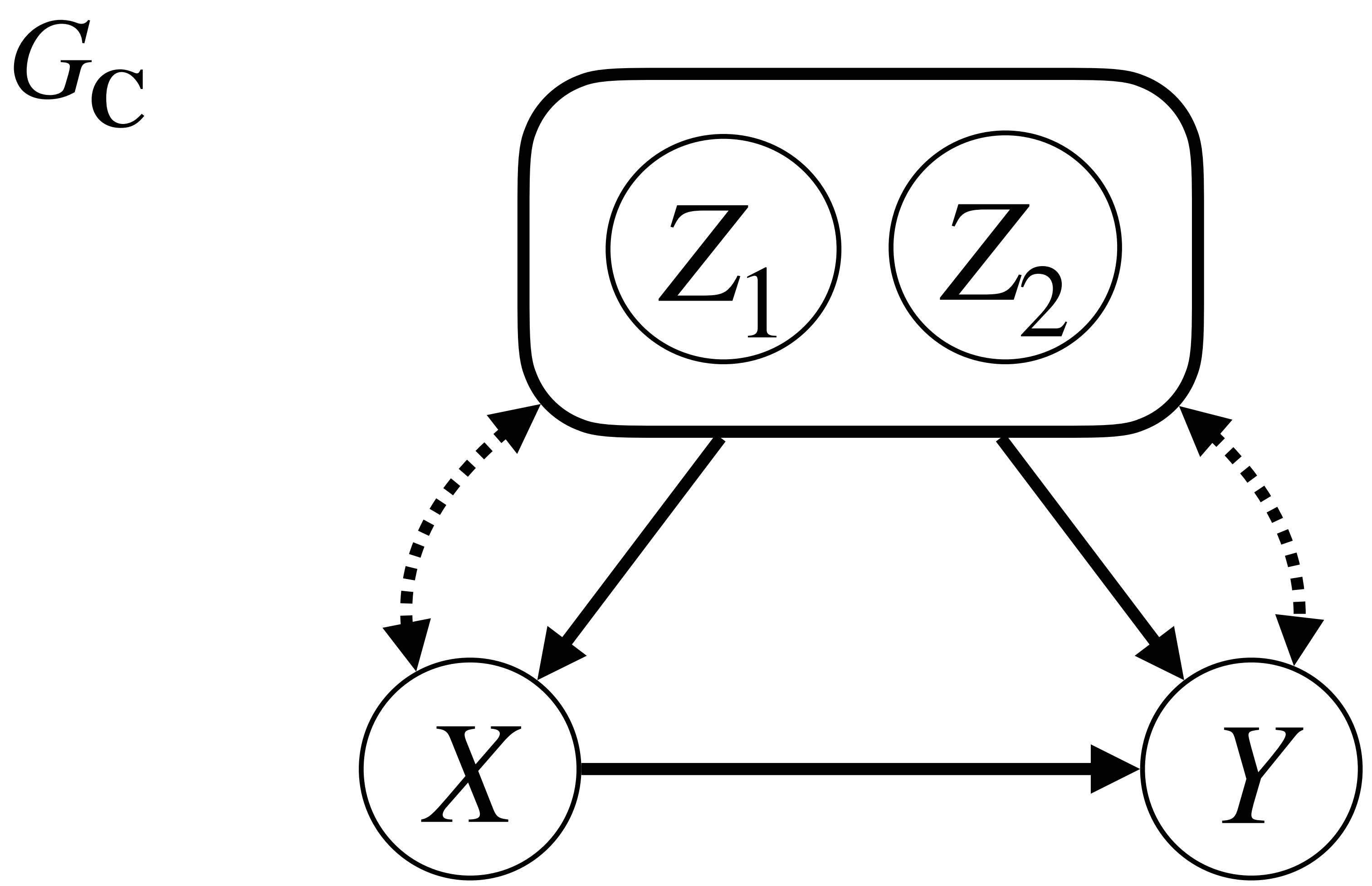


$$P(y \mid do(x)) = \sum_{z_1, z_2} P(y \mid x, z_1, z_2) P(z_1, z_2)$$

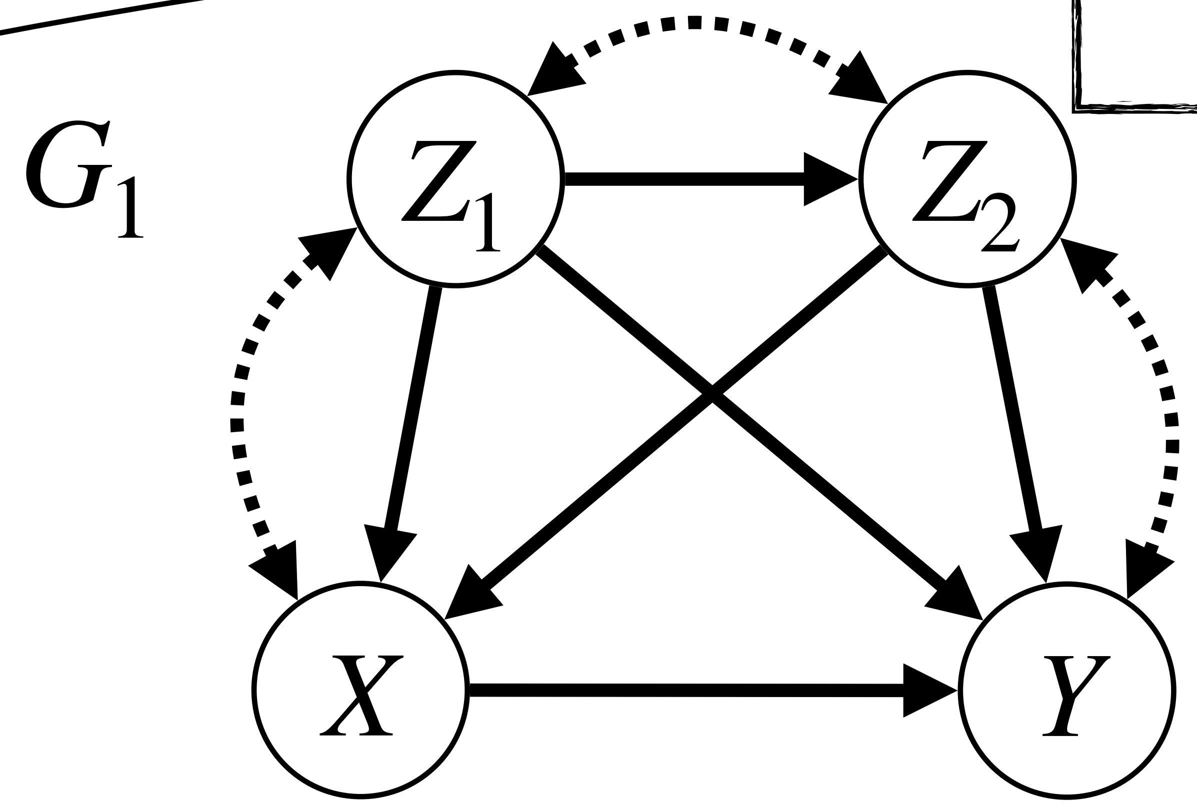
An identifiable effect in a C-DAG $G_{\mathbf{C}}$ is identifiable in all compatible causal diagrams G using the same identification formula!

Effect Non-Identifiabiliy given a C-DAG

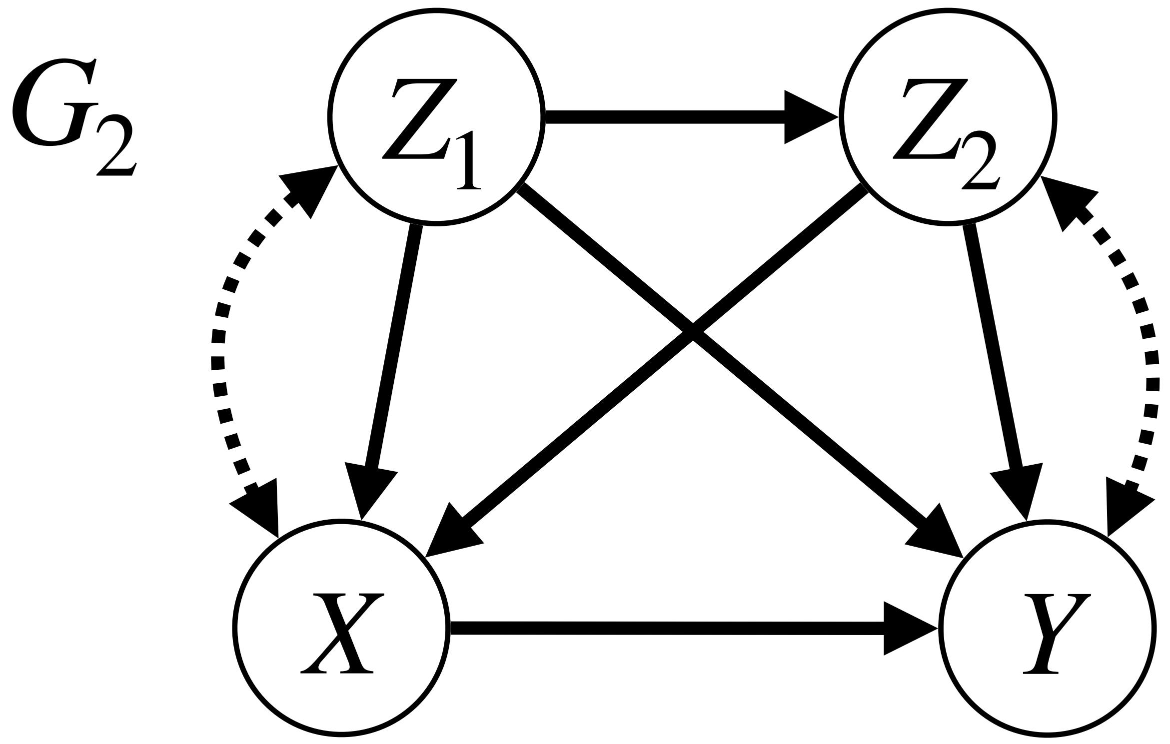
Simple evaluation of a **violation** of the Backdoor Criterion / Conditional Exchangeability



$P(y \mid do(x))$ is not identifiable



$P(y \mid do(x))$ is not identifiable



$$P(y \mid do(x)) = \sum_{z_1, z_2} P(y \mid x, z_1, z_2) P(z_1, z_2)$$

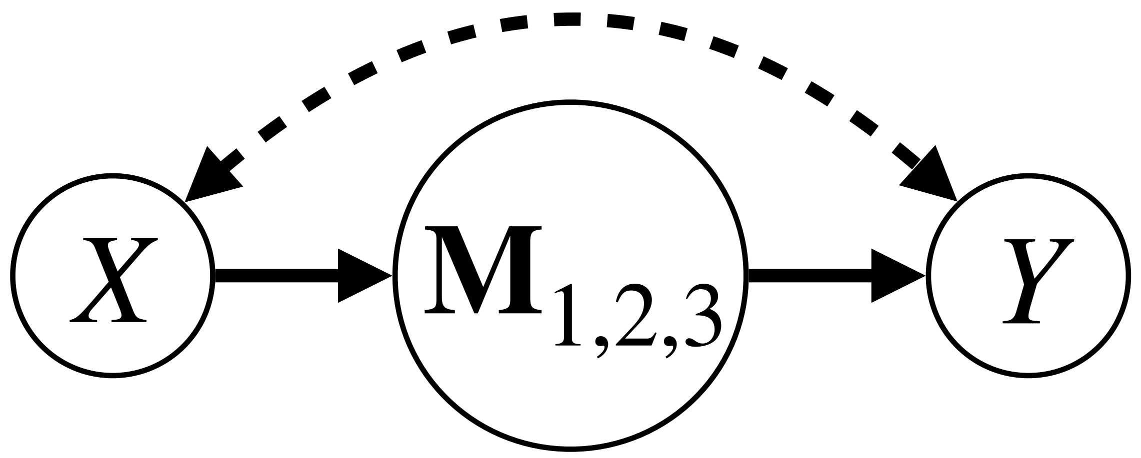
A non-identifiable effect in a C-DAG $G_{\mathbf{C}}$ implies that there exists at least one compatible causal diagrams G in which the effect is non-identifiable.

Identification of Causal Effects from C-DAGs

1 **Query**

$P(y \mid do(x))$

2 **C-DAG**



3 **Data**

$P(x, m_1, m_2, m_3, y)$

Inference Engine

Solution

yes / no

$$P(y \mid do(x)) = \sum_{\mathbf{m}_{123}} P(\mathbf{m}_{123} \mid x) \sum_{x'} P(y \mid \mathbf{m}_{123}, x')P(x')$$

Inferred (Interventional) Distribution ← Available (Observational) Distribution 👍

Anand, T. V., **Ribeiro A. H.,** Tian, J., & Bareinboim, E. (2023). Causal Effect Identification in Cluster DAGs. In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence.

Partial Identification of Causal Effects

For Discrete Treatments:

- Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. Proceedings of the 39th International Conference on Machine Learning, PMLR 162, pages 26548–26558, 2022.

For Continuous Treatments:

- Padh, Kirtan; Zeitler, Jakob; Watson, David; Kusner, Matt; Silva, Ricardo; Kilbertus, Niki; (2023) Stochastic causal programming for bounding treatment effects. In: (Proceedings) CLeaR 2023 : 2nd Conference on Causal Learning and Reasoning. (Link)

Partial Identification in C-DAGs!

What if *no* knowledge is available?

Can we learn a causal diagram $\mathscr{G}$ from observational data?

Causal Discovery (Lecture 2):

In non-parametric settings, we can't learn the true causal diagram, but algorithms such as the Fast Causal Inference (FCI) can learn a graphical representation of its *Markov equivalence class*!

Can we identify causal effects from the equivalence class?

Effect Identification:

Yes! Recently, we proposed ***generalized do-calculus and complete algorithms*** for the identification of marginal and conditional causal effect in PAGs!

Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896. Link

Jaber A., **Ribeiro A. H.,** Zhang, J., Bareinboim, E. (2022) Causal Identification under Markov Equivalence - Calculus, Algorithm, and Completeness. In Proceedings of the 36th Annual Conference on Neural Information Processing Systems, NeurIPS 2022. (Link)

Generalized Inference for PAGs

	DAGs / ADMGs	**PAGs**
$\mathscr{L}_2$ **(associational) inferences**	ID-Algorithm Conditional ID-Algorithm Do-calculus	IDP algorithm CIDP algorithm Generalized calculus
$\mathscr{L}_1$ **(associational) inferences**	d-separation	m-separation

Covered in Lecture 2!

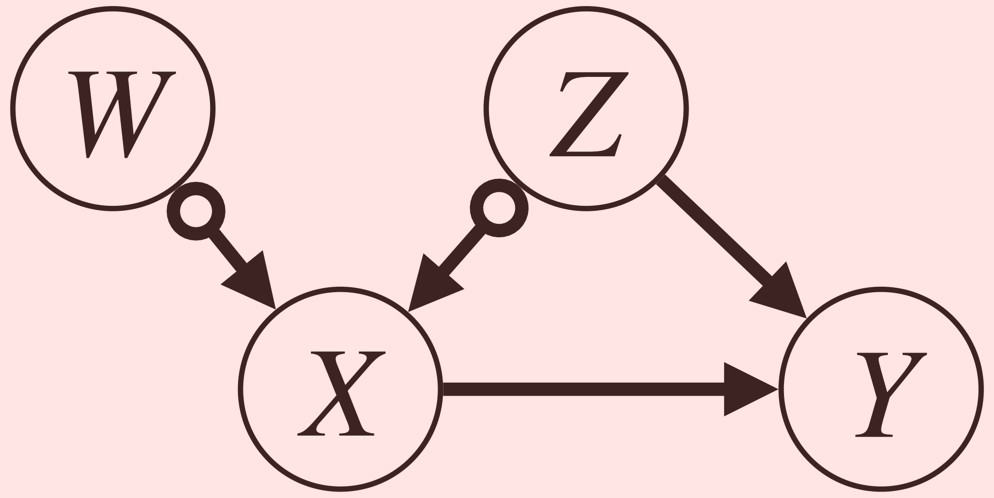
General Identification in Markov Equivalence Classes

Can be constructed in a fully data-driven way!

1 Query

$P(y \mid do(x))$

2 PAG



3 Data

$P(x, y, z, w)$ **Observational Distribution**

Inference Engine

Solution

yes / no

$$P(y \mid do(x)) = \sum_z P\left(y \mid x, z\right) P(z)$$

Inferred (Interventional) Distribution ← Available (Observational) Distribution

Jaber A., **Ribeiro A. H.,** Zhang, J., Bareinboim, E. Causal Identification under Markov Equivalence - Calculus, Algorithm, and Completeness. In Proceedings of the 36th Annual Conference on Neural Information Processing Systems, NeurIPS 2022. (Link)

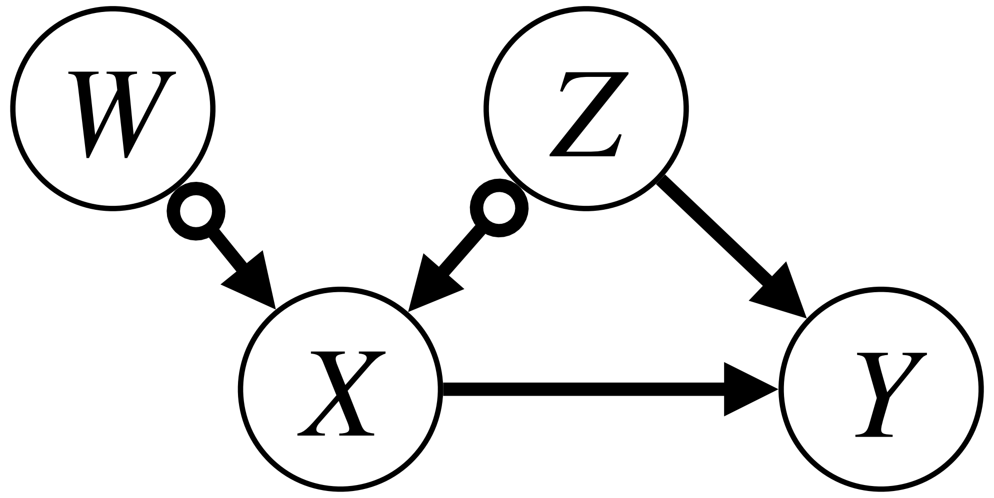
General Identification in Markov Equivalence Classes

The CIDP and IDP algorithms are available at the *PAGId* R package: https://github.com/adele/PAGId

1 Query

$P(y \mid do(x))$

2 PAG



3 Data

$P(x, y, z, w)$ **Observational Distribution**

Inference Engine

Solution

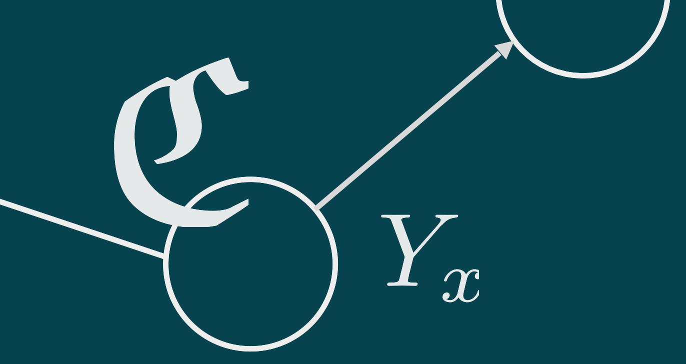
yes / no

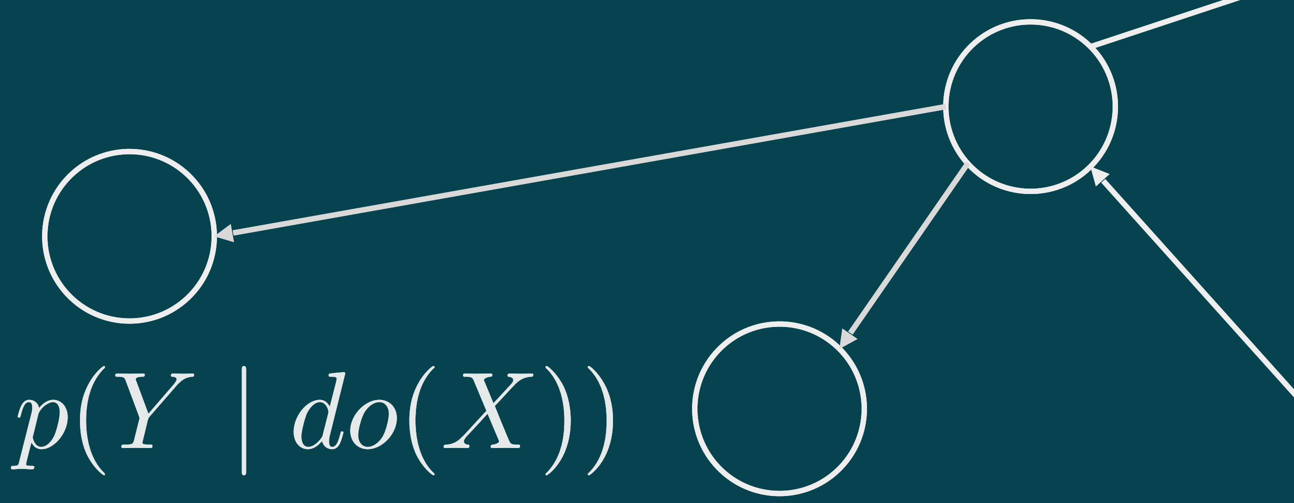
$$P(y \mid do(x)) = \sum_z P\left(y \mid x, z\right) P\left(z\right)$$

Inferred (Interventional) Distribution ← Available (Observational) Distribution

👍

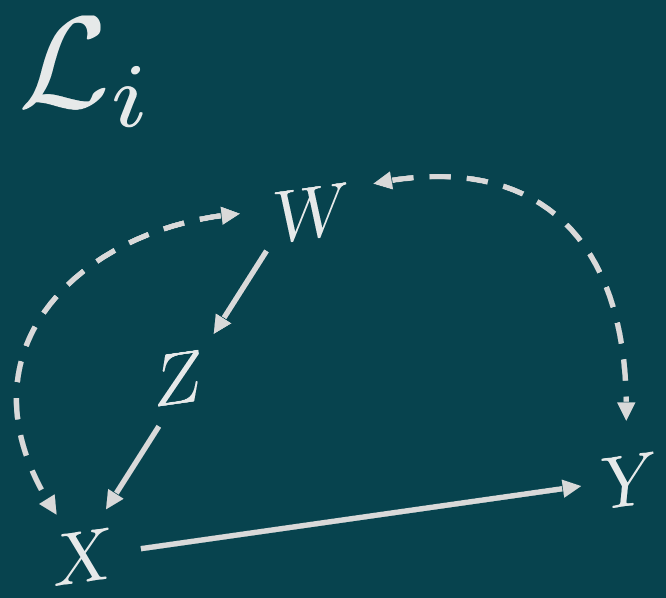
Jaber A., **Ribeiro A. H.,** Zhang, J., Bareinboim, E. Causal Identification under Markov Equivalence - Calculus, Algorithm, and Completeness. In Proceedings of the 36th Annual Conference on Neural Information Processing Systems, NeurIPS 2022. (Link)



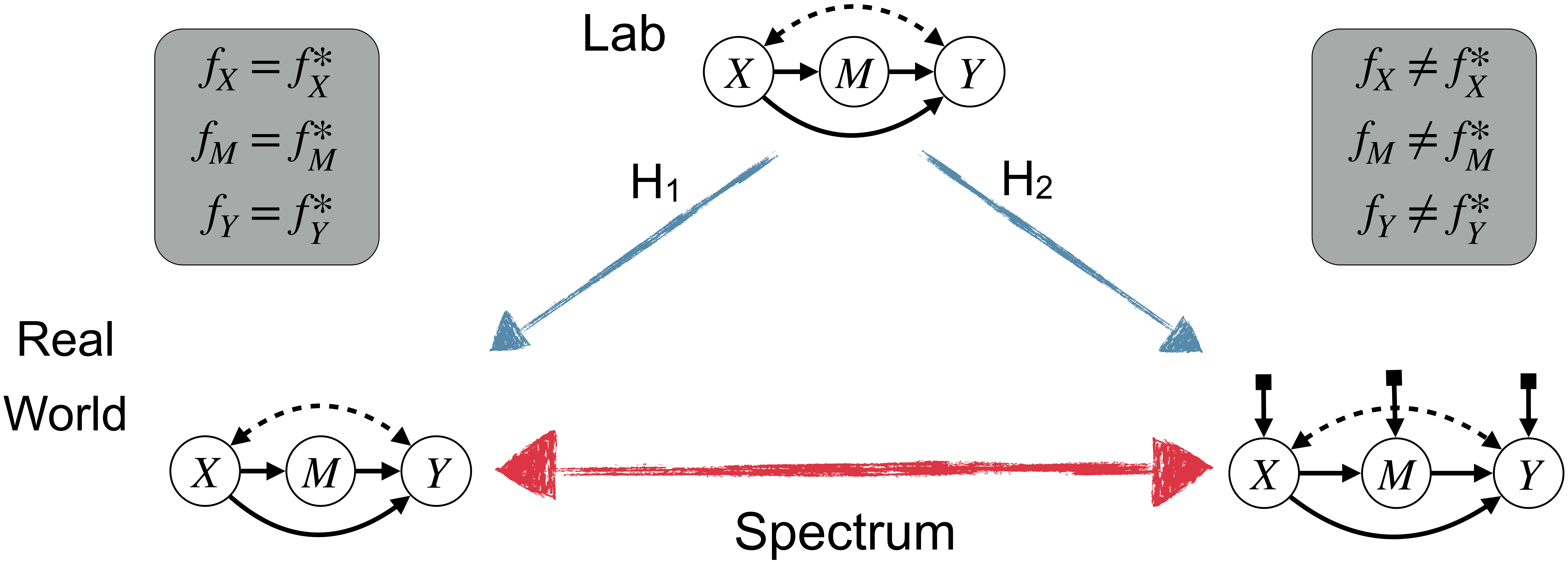


Causal Transportability

Current State of the Art



Generalizability: from the “Lab” to the “Real World”



Everything is assumed to be the same, trivially transportable!

Everything is assumed to be different, not transportable...

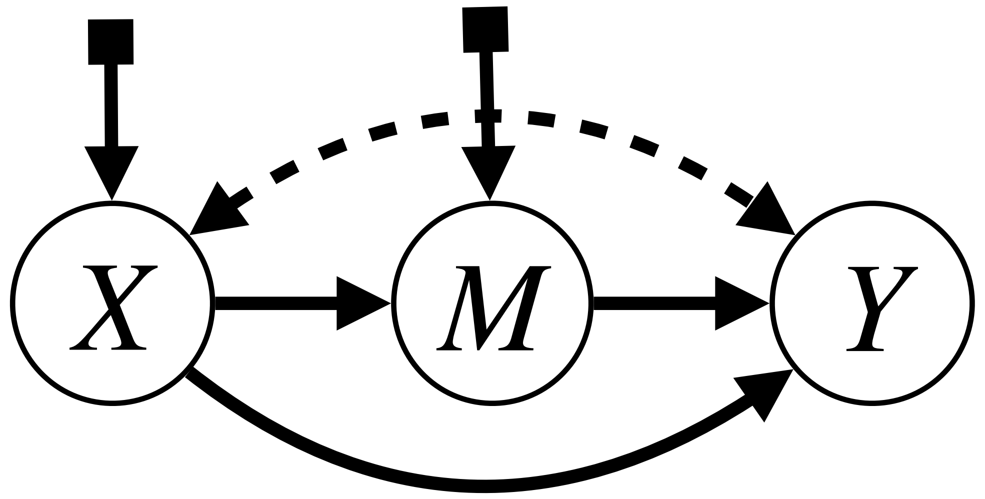
* The lab stands for any environment, population, domain, setting.

Effect Transportability - Workflow

1 **Query**

$P^*(y \mid do(x))$

2 **Causal Contraints**



3 **Data**

$P^*(x, m, y)$

$P(m, y \mid do(x))$

Inference Engine

Solution

yes / no

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z \mid x)$$

Interventional Distribution (Target Domain) ← Available Distributions

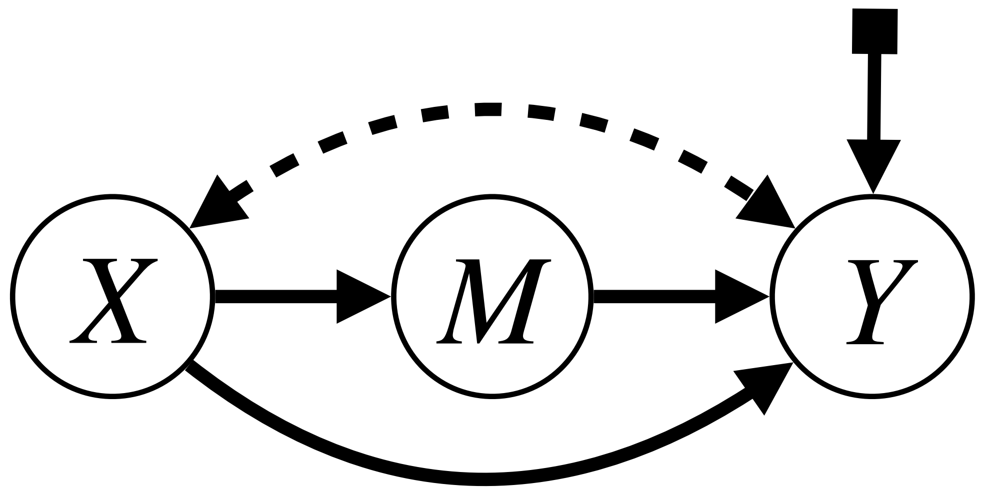
- Lee, S., Correa, J., and Bareinboim, E. (2020). General Transportability – Synthesizing Observations and Experiments from Heterogeneous Domains. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY. AAAI Press. Link

Effect Transportability - Workflow

1 **Query**

$P^*(y \mid do(x))$

2 **Causal Contraints**



3 **Data**

$P^*(x, m, y)$

$P(m, y \mid do(x))$

Inference Engine

Solution

yes / no

$P^*(y \mid do(x)) = ?$

Interventional Distribution (Target Domain) ← Available Distributions

- Lee, S., Correa, J., and Bareinboim, E. (2020). General Transportability – Synthesizing Observations and Experiments from Heterogeneous Domains. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY. AAAI Press. [Link]

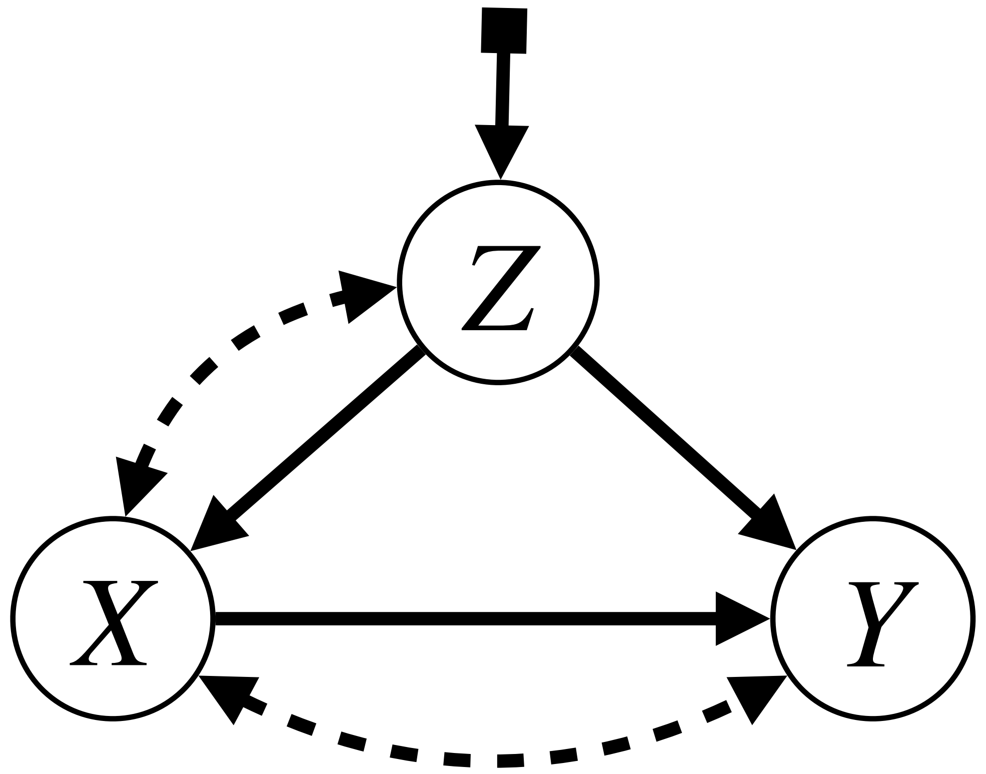
Sensitivity to the Causal Assumptions

X: Treatment

Y: Outcome

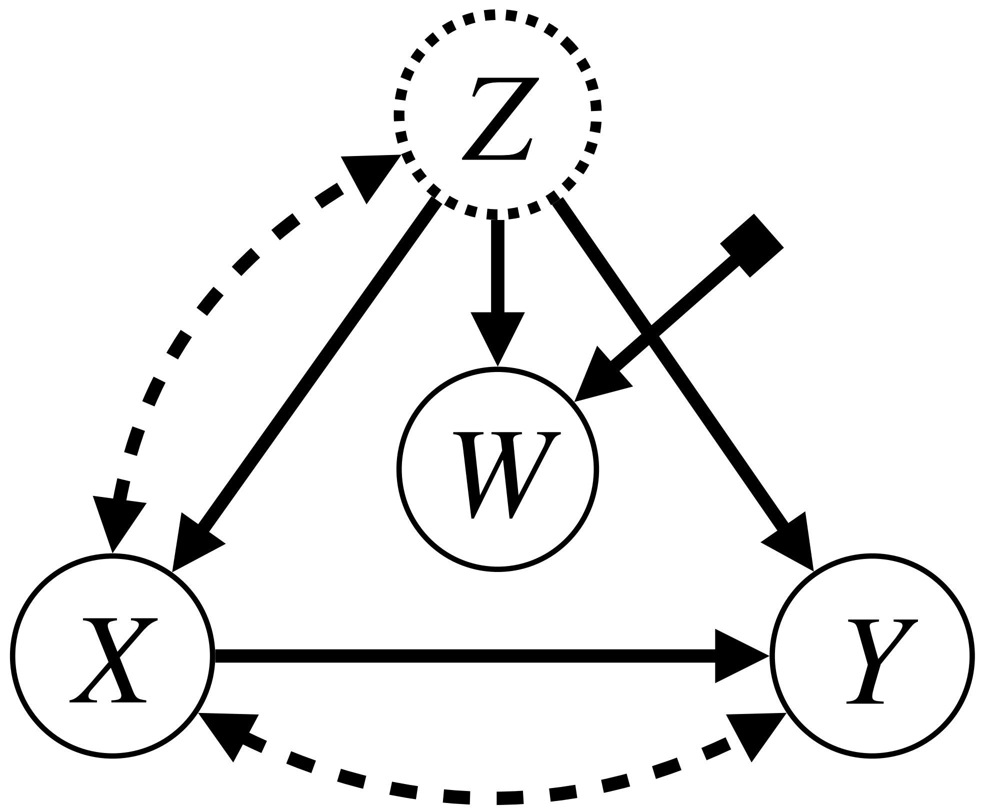
$\Pi \rightarrow \Pi^*$

USA Germany



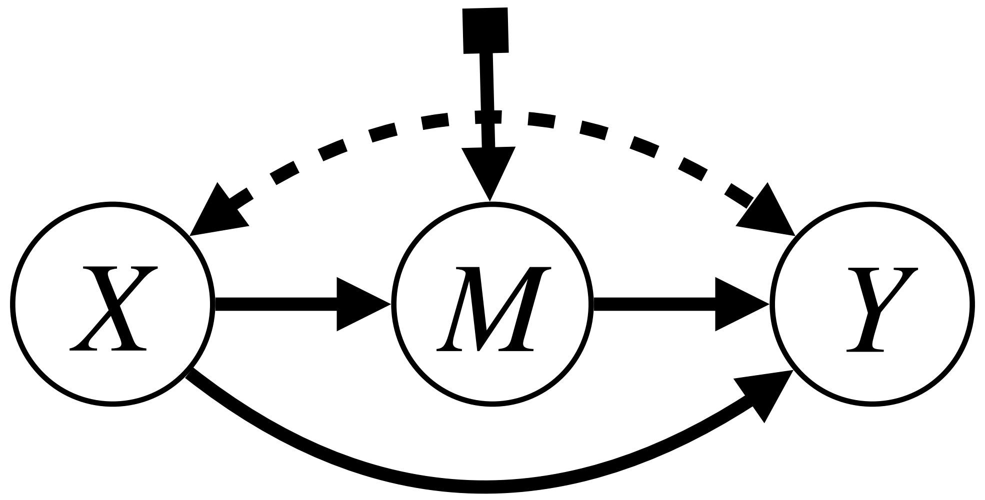
Z: Age

$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), z) P^*(z)$$



W: Education

$$P^*(y \mid do(x)) = P(y \mid do(x))$$

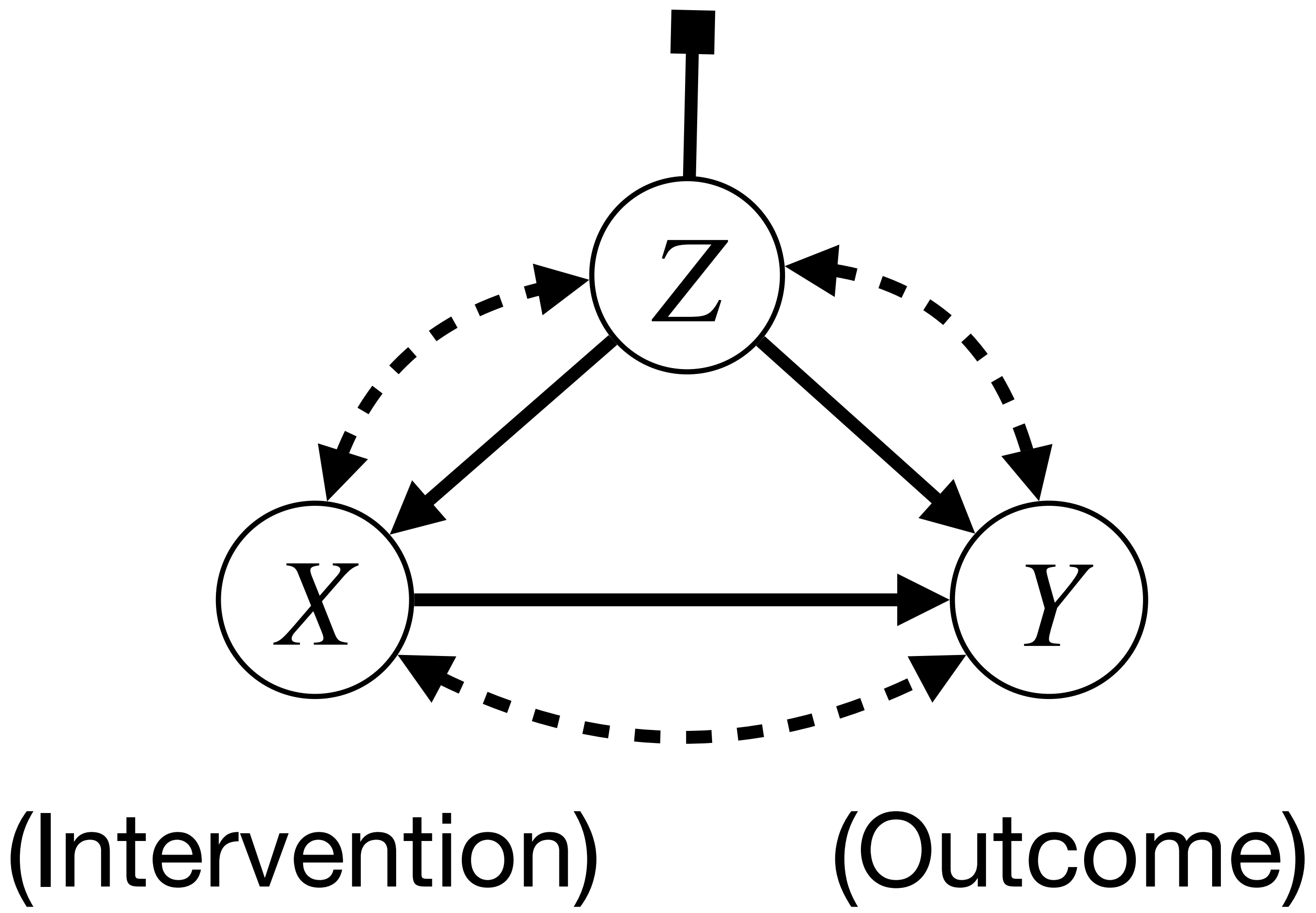


M: Biological Mediator

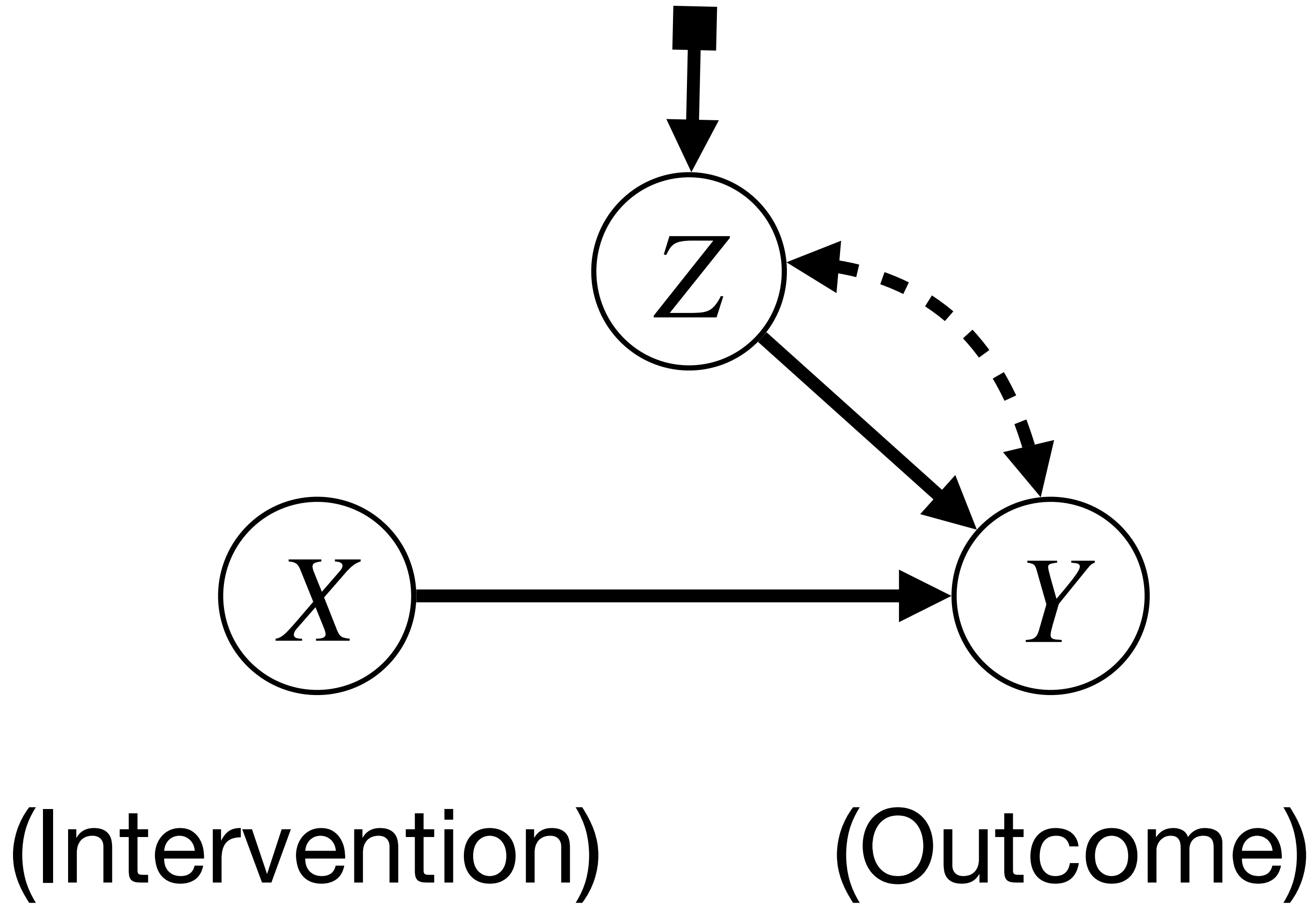
$$P^*(y \mid do(x)) = \sum_z P(y \mid do(x), m) P^*(m \mid x)$$

Is the Gold Standard Golden?

Before randomization

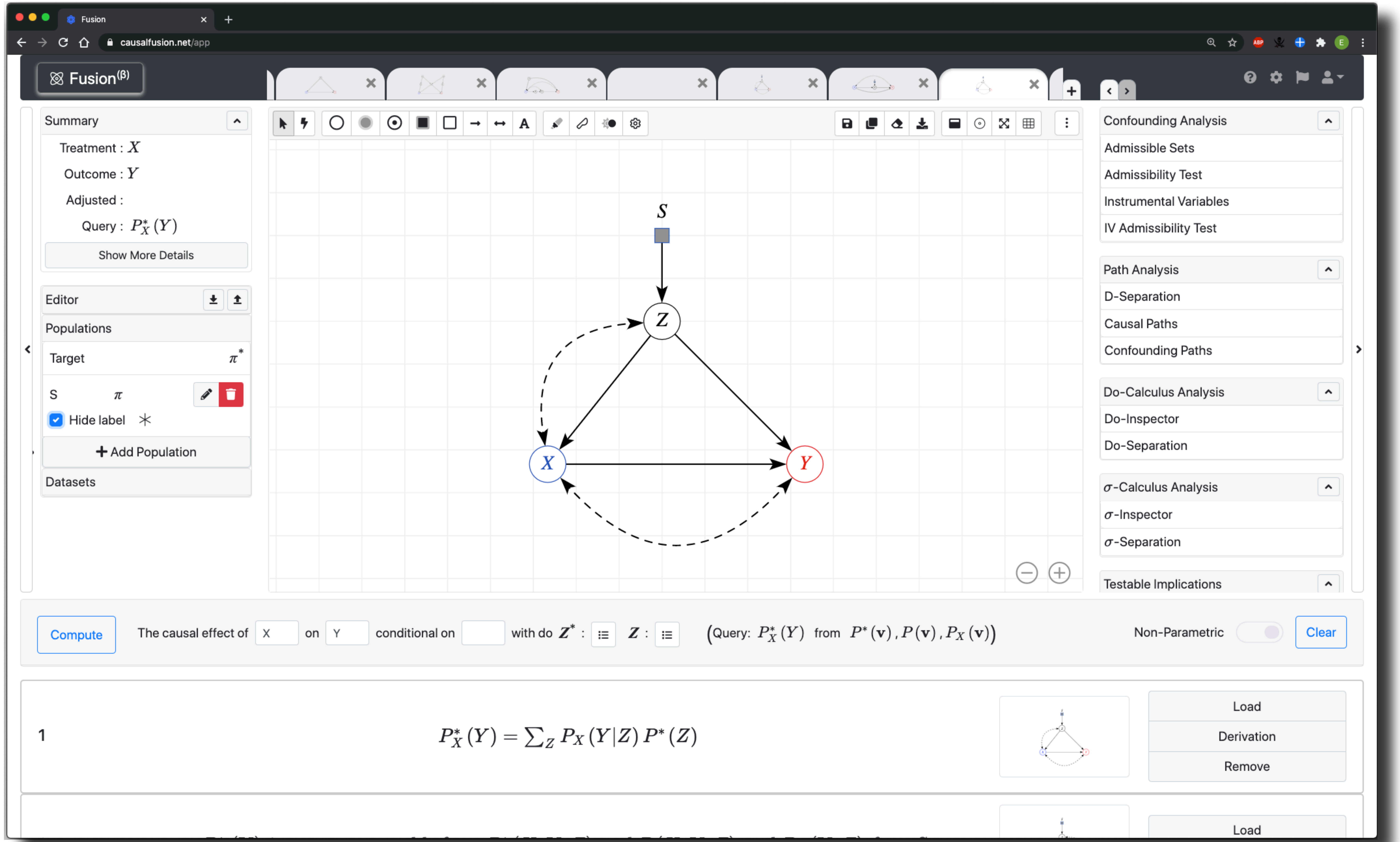


After randomization



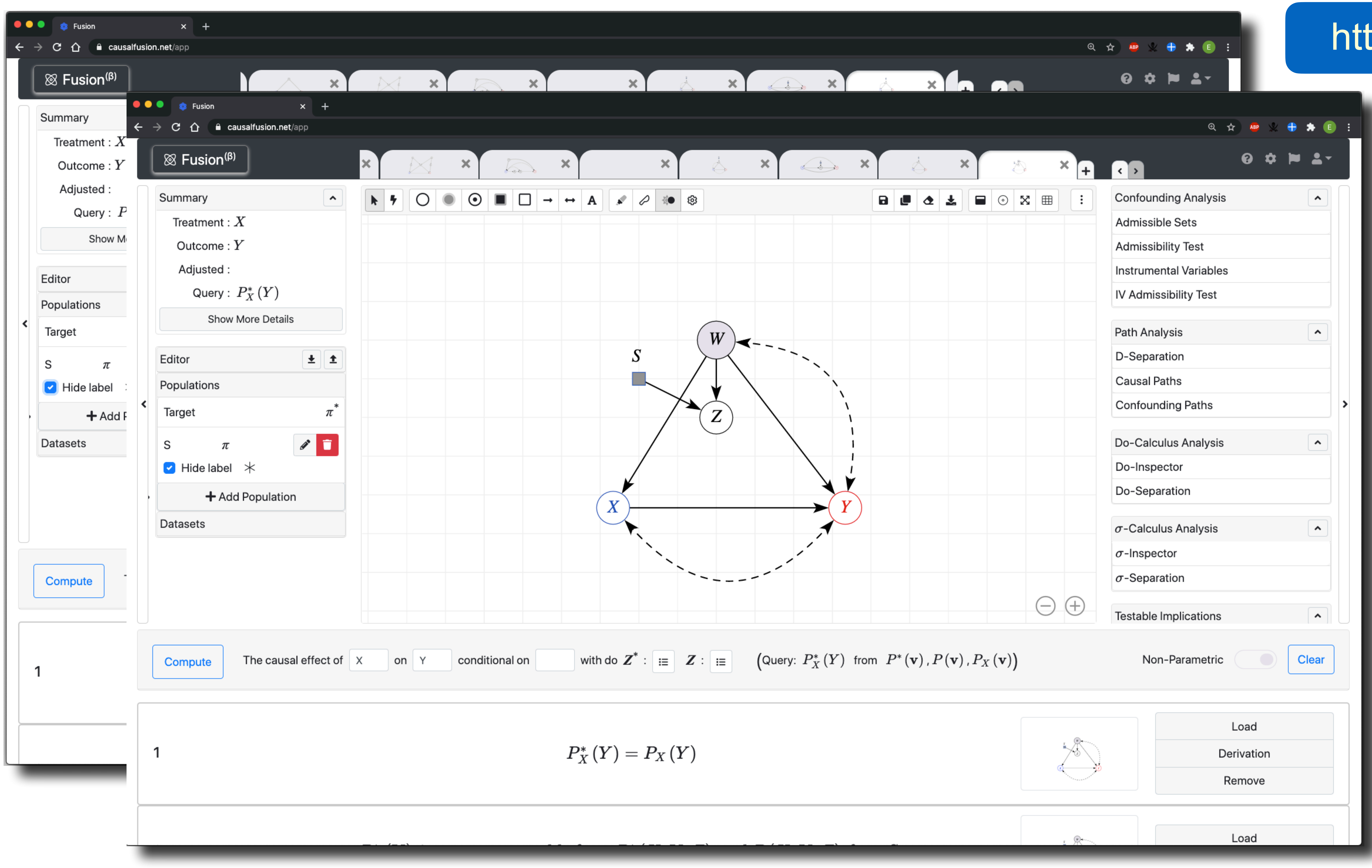
Lesson. Even if we have a perfect RCT, one still needs to go through a transportability exercise. TR theory is unavoidable.

35

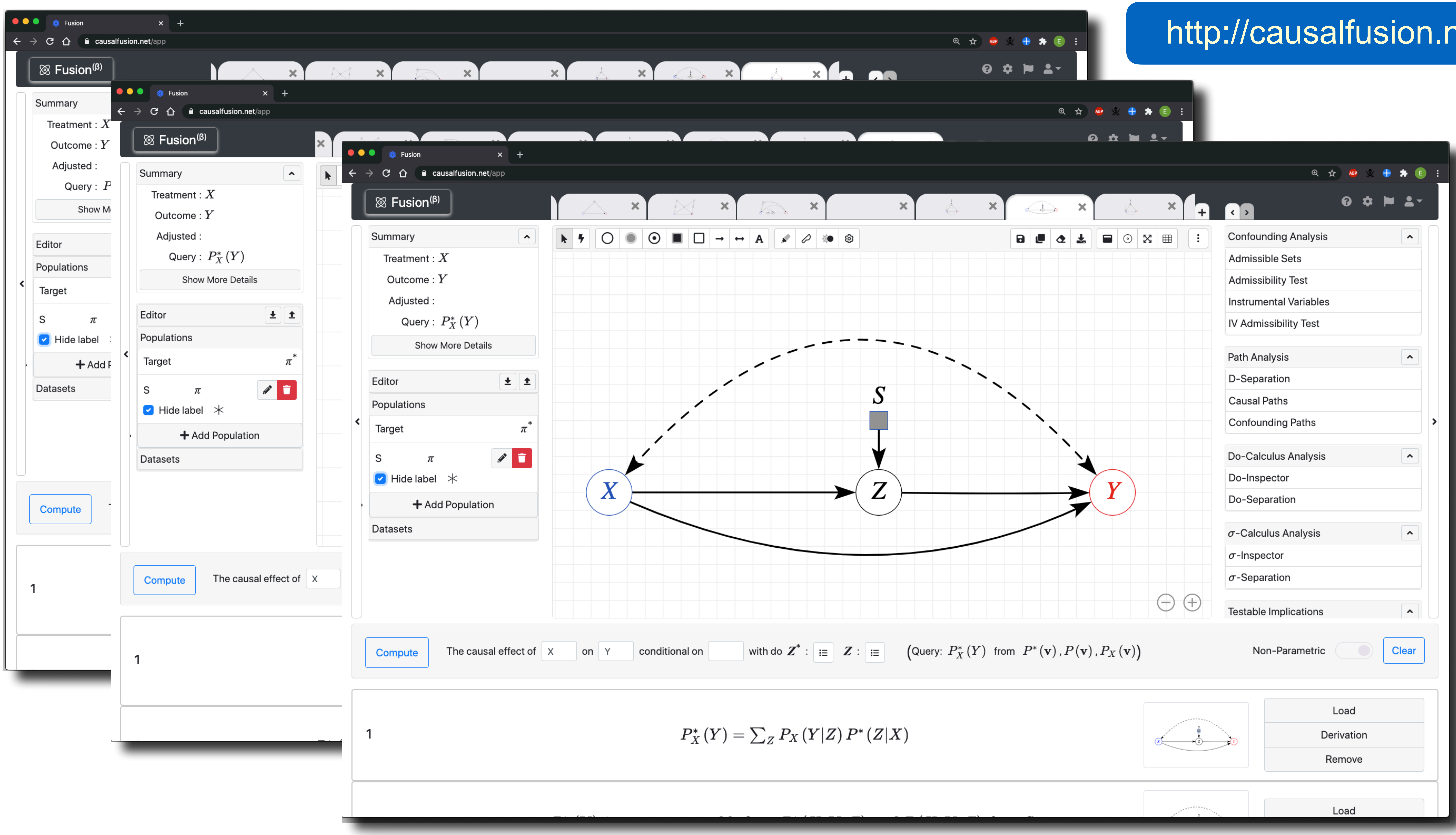


http://causalfusion.net

http://causalfusion.net



http://causalfusion.net



More on Causal Transportability

Transportability of causal effects:

Lee, S., Correa, J., and Bareinboim, E. (2020). General Transportability – Synthesizing Observations and Experiments from Heterogeneous Domains. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY. AAAI Press. Link

Transportability of stochastic/soft interventions:

Correa, J. and Bareinboim, E. (2020). General transportability of soft interventions: Completeness results. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 10902–10912, Vancouver, Canada. Curran Associates, Inc. Link

Counterfactual transportability:

Correa, J. Lee, S., Bareinboim. E. (2022). Counterfactual Transportability: A Formal Approach. In Proceedings of the 39th International Conference on Machine Learning.

Other Lines of Work and Research Questions

- Robustness agains violations of the underlying assumptions (e.g., faithfulness)
- Bayesian approaches for modeling of uncertainty
- Causal Data-Fusion:
  - Learning and inference from a combination of data + prior knowledge
  - Learning and inference from multiple heterogenous datasets and populations
- Causality and Continual Learning https://www.continualcausality.org/
- And much more…
  - Counterfactual identification https://causalai.net
  - Causal Fairness https://fairness.causalai.net
  - Causal reinforcement learning https://crl.causalai.net