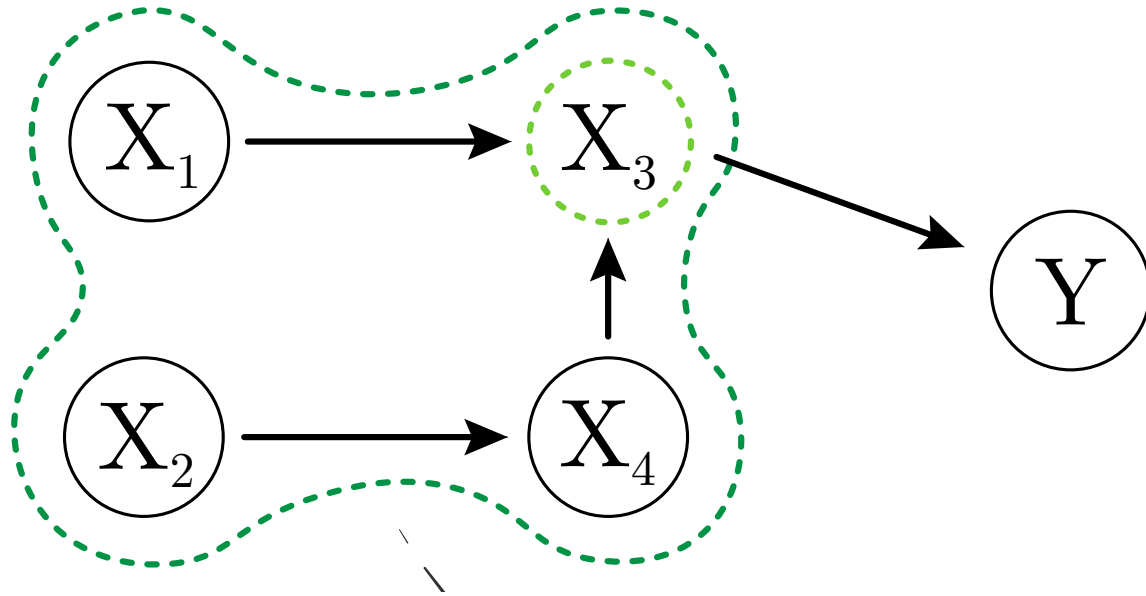


Exactly Computing do-Shapley Values

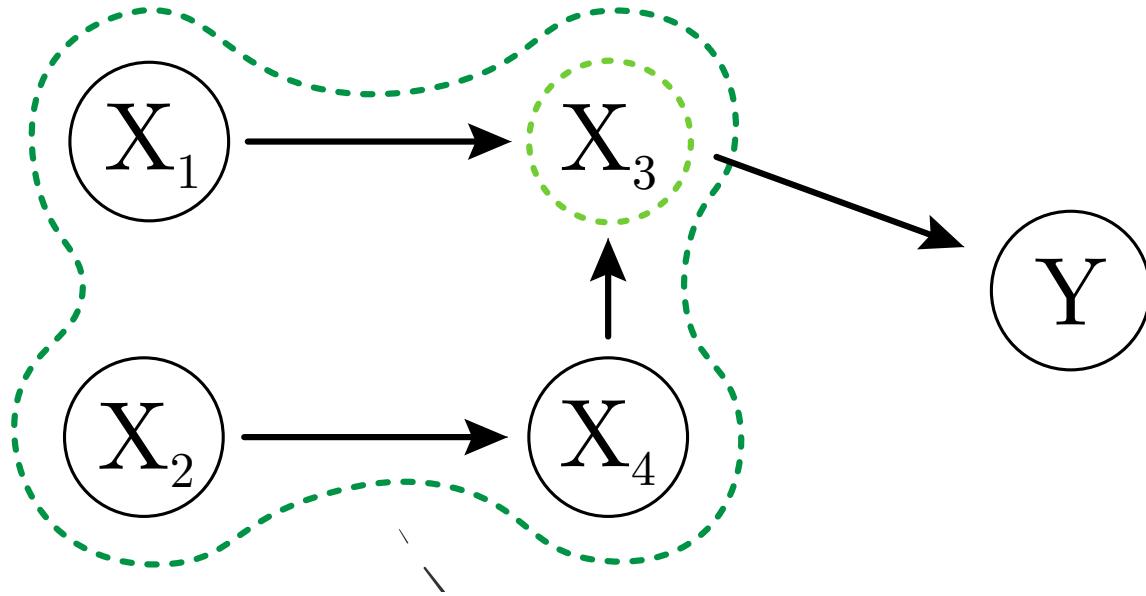
R. **Teal Witter***, Álvaro Parafita*, Tomas Garriga, Maximilian Muschalik,
Fabian Fumagalli, Axel Brando, Lucas Rosenblatt

Structural Causal Model (SCM)



- Directed Acyclic Graph (DAG)
- Nodes are variables
- Edge from X_i to X_j means X_i is a direct *cause* of X_j

do-Shapley Values



Question:

What is the *impact* of X_i on Y ?

Answer:

It depends on the values of the other variables.

do-Shapley Values

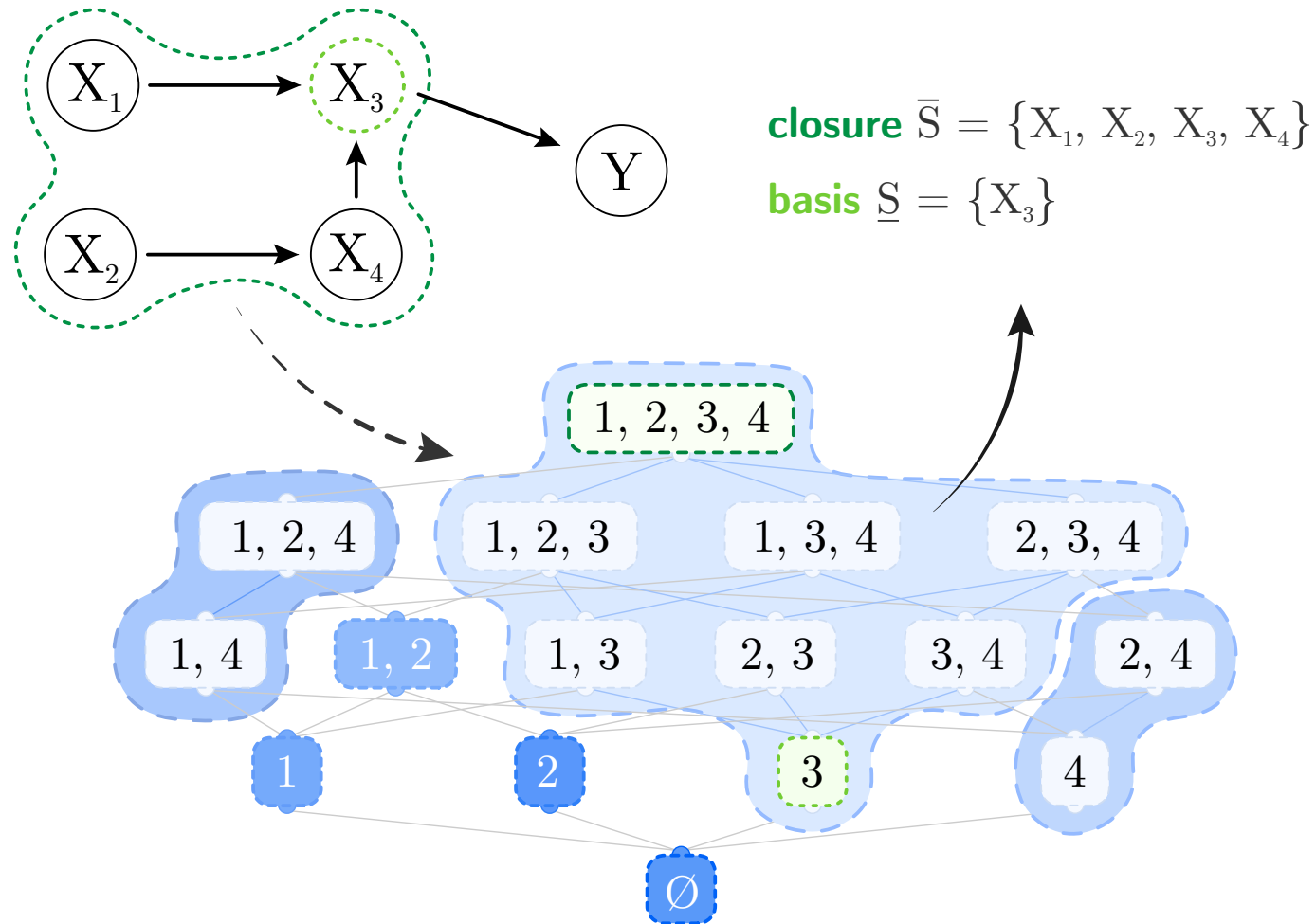
Expectation of Y when variables in S are set to \mathbf{x}_S :

$$v(S) = \mathbb{E}[Y \mid do(S = \mathbf{x}_S)]$$

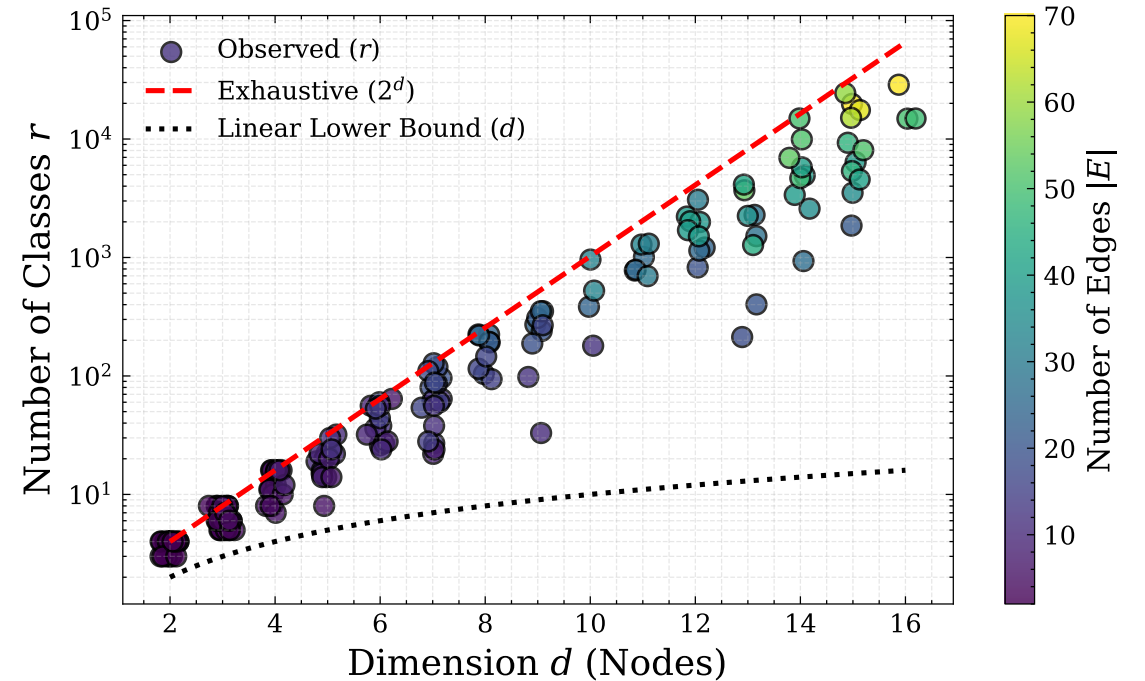
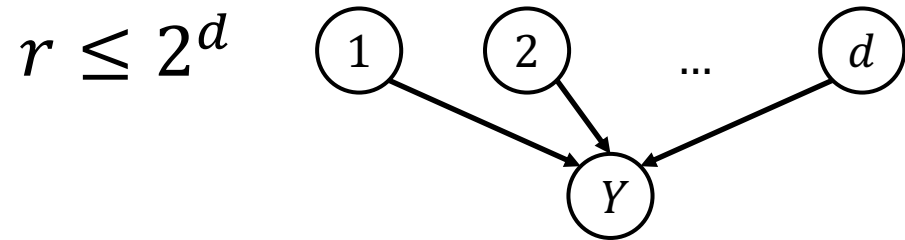
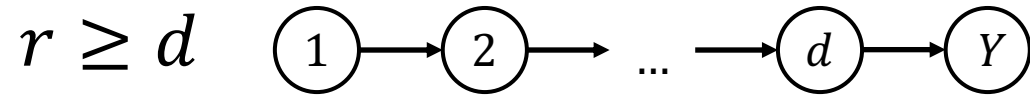
Average impact of i th variable:

$$\phi_i = \sum_{S \subseteq [d] \setminus \{i\}} [v(S \cup \{i\}) - v(S)] \cdot p_{|S|}$$

Redundant Values



Number of Equivalence Classes r



Rewriting do-Shapley Values

$$\begin{aligned}
 \phi_i &= \sum_{S \subseteq [d] \setminus \{i\}} [v(S \cup \{i\}) - v(S)] p_{|S|} \\
 &= \sum_{S \subseteq [d]} v(S) [\mathbf{1}[i \in S] p_{|S|-1} - \mathbf{1}[i \notin S] p_{|S|}] \\
 &= \sum_{j=1}^r v(c_j) \sum_{S: \underline{S} \subseteq S \subseteq \bar{S}} \mathbf{1}[i \in S] p_{|S|-1} - \mathbf{1}[i \notin S] p_{|S|} \\
 &= \sum_{j=1}^r v(c_j) w_i(c_j)
 \end{aligned}$$

$$w_i(c) = \begin{cases} \sum_{\ell=|\underline{S}|}^{|\bar{S}|} p_{\ell-1} \binom{|\bar{S}| - |\underline{S}|}{\ell - |\underline{S}|} & \text{if } i \in \underline{S} \\ - \sum_{\ell=|\underline{S}|}^{|\bar{S}|} p_{\ell} \binom{|\bar{S}| - |\underline{S}|}{\ell - |\underline{S}|} & \text{if } i \in \bar{S} \\ 0 & \text{else} \end{cases}$$

Exactly Computing do-Shapley Values

With c_1, \dots, c_r , we can *exactly* compute ϕ_i

We design algorithms for...

- finding all classes (linear in r)
- estimating ϕ_i with budget m (linear in m)

Smooth transition: When $m \geq r$, we *exactly* recover ϕ_i

Collaborators!

