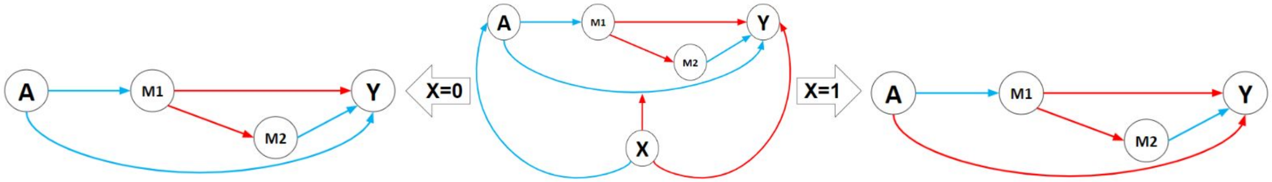
On Heterogeneous Treatment Effects in Heterogeneous Causal Graphs

Richard Watson, Hengrui Cai, Xinming An, Samuel Mclean, Rui Song

Motivation



- Causal discovery, the task of discovering the causal relations between variables in a dataset, has interesting and important applications in many areas, such as epidemiology, medicine, economics, etc.
- Yet, due to the heterogeneity of individuals in response to different events/treatments, there may not exist a uniform causal graph for everyone. This implies the existence of heterogeneous causal graphs (HCGs).
- Very few studies have been conducted to investigate heterogeneous causal effects (HCEs) in **graphical contexts** due to the lack of statistical methods.

Contributions

- To our knowledge, this is the **first** work that considers heterogeneity in terms of causal graphs. We conceptualize HCGs, by **incorporating moderators** and their interaction directly into our model
- We compute the **HCEs, directly from the HCG,** to quantify the impact of the treatment and mediators on the outcome of interest **given the moderators**.
- We propose an **interactive structural learning algorithm** to learn the complex HCGs, estimate HCEs, and compute bootstrap confidence intervals for these estimates via a debiasing process.

Whole graph LSEM

- X is directly **incorporated** into model along with interaction
- B's **sparness** can be described using prior causal information
- **No** noise distribution assumed on the errors

Weighted Adjacency Matrix

$$\boldsymbol{D} = \boldsymbol{B}^\top \boldsymbol{D} + \epsilon : \quad \begin{matrix} p \\ 1 \\ p \\ s \\ 1 \end{matrix} \begin{bmatrix} \boldsymbol{X} \\ A \\ \boldsymbol{X}A \\ \boldsymbol{M} \\ Y \end{bmatrix} = \boldsymbol{B}^\top \begin{bmatrix} \boldsymbol{X} \\ A \\ \boldsymbol{X}A \\ \boldsymbol{M} \\ Y \end{bmatrix} + \begin{bmatrix} \epsilon_{\boldsymbol{X}} \\ \epsilon_A \\ \epsilon_{\boldsymbol{X}A} \\ \epsilon_{\boldsymbol{M}} \\ \epsilon_Y \end{bmatrix}$$

Data Vector

Error Vector

Whole graph to subgraph

- The **do-operator** is used to reveal the desired subgraph once the whole graph has been estimated
- This operation applies a value to the variable in question ignoring any parents.

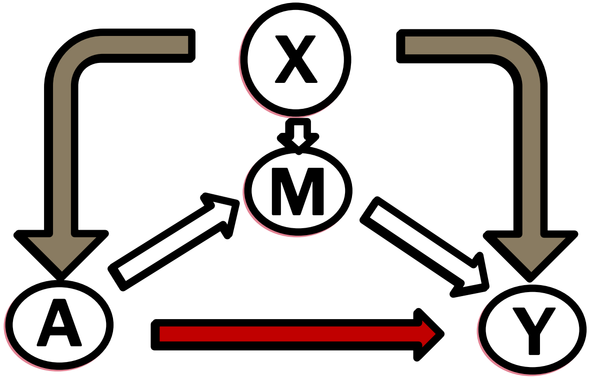
$$\boldsymbol{D}_{do(\boldsymbol{X}=\boldsymbol{x})} = \boldsymbol{B}^{\top}_{do(\boldsymbol{X}=\boldsymbol{x})} \boldsymbol{D}_{do(\boldsymbol{X}=\boldsymbol{x})} + \boldsymbol{\epsilon}' \rightarrow$$

$$\begin{bmatrix} A \\ M \\ Y \end{bmatrix} = \begin{bmatrix} \boldsymbol{\delta}_{\boldsymbol{X}} \boldsymbol{x} \\ \boldsymbol{B}^{\top}_{\boldsymbol{X}} \boldsymbol{x} \\ \boldsymbol{\gamma}_{\boldsymbol{X}} \boldsymbol{x} \end{bmatrix} + \boldsymbol{B}^{\top}_{do(\boldsymbol{X}=\boldsymbol{x})} \begin{bmatrix} A \\ M \\ Y \end{bmatrix} + \begin{bmatrix} \epsilon_A \\ \boldsymbol{\epsilon}_M \\ \epsilon_Y \end{bmatrix}$$

Heterogeneous Causal Effects

Heterogeneous Direct Effect of A on Y (HDE): The effect on Y that is due to A *and not mediated* by M given the baseline covariates, X

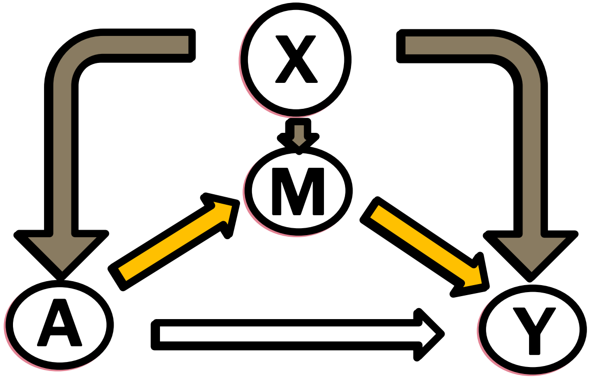
$$HDE(\boldsymbol{x}) = E\big[Y|do(A = a+1, \boldsymbol{M} = \boldsymbol{m}^{(a)}), \boldsymbol{X} = \boldsymbol{x}\big] - E\big[Y|do(A = a), \boldsymbol{X} = \boldsymbol{x}\big]$$



Heterogeneous Causal Effects

Heterogeneous Indirect Effect of A on Y (HIE): The effect on Y that is due to A *and mediated* by M given the baseline covariates, X

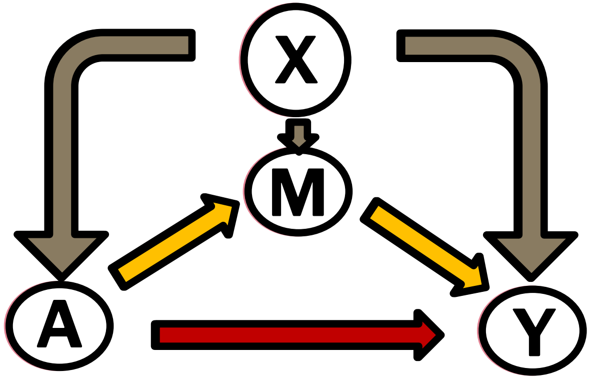
$$HIE(\boldsymbol{x}) = E\left[Y|do(A=a, \boldsymbol{M}=\boldsymbol{m}^{(a+1)}), \boldsymbol{X}=\boldsymbol{x}\right] - E\left[Y|do(A=a), \boldsymbol{X}=\boldsymbol{x}\right]$$



Heterogeneous Causal Effects

Heterogeneous Total Effect of A on Y (HTE): The total effect on Y that is due to A given the baseline covariates, X

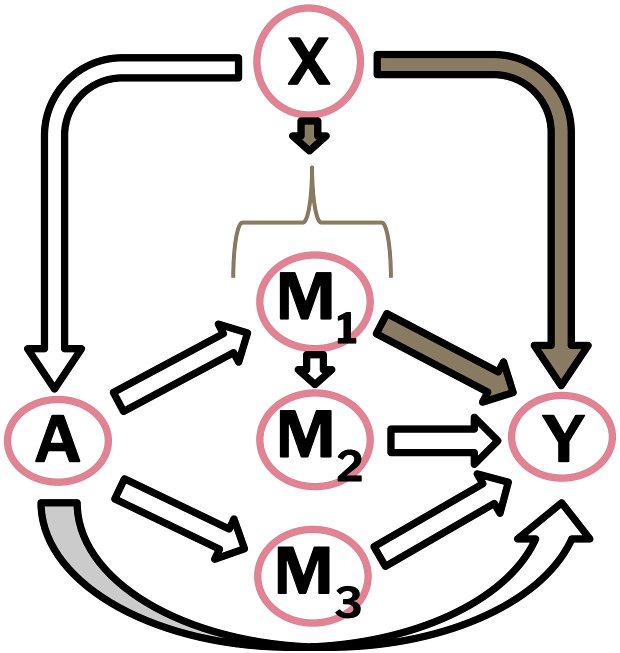
$$HTE(\boldsymbol{x}) = E[Y|do(A = a + 1), \boldsymbol{X} = \boldsymbol{x}] - E[Y|do(A = a), \boldsymbol{X} = \boldsymbol{x}]$$



Heterogeneous Causal Effects

Heterogeneous Direct Mediation Effect of M_i on Y (HDM): The effect on Y that is due to M_i *and not mediated* by other mediators given the baseline covariates, X

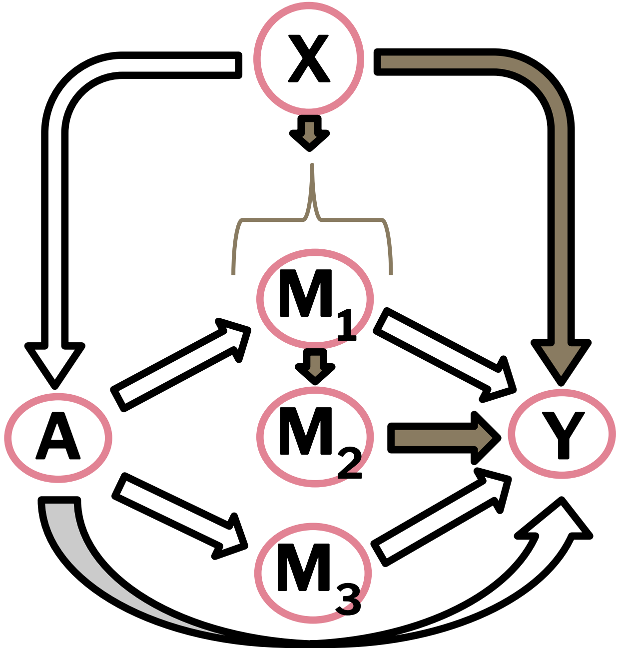
$$HDM_i(\boldsymbol{x}) = \Big[E[M_i|do(A=a+1), \boldsymbol{X}=\boldsymbol{x}] - E[M_i|do(A=a), \boldsymbol{X}=\boldsymbol{x}]\Big] \times \Big[E[Y|do(A=a, M_i=m_i^{(a)}+1, \boldsymbol{\Omega}_i=\boldsymbol{o}_i^{(a)}), \boldsymbol{X}=\boldsymbol{x}] - E[Y|do(A=a), \boldsymbol{X}=\boldsymbol{x}]\Big]$$



Heterogeneous Causal Effects

Heterogeneous Indirect Mediation Effect of M_i on Y (HIM): The effect on Y that is due to M_i *and mediated* by other mediators given the baseline covariates, X

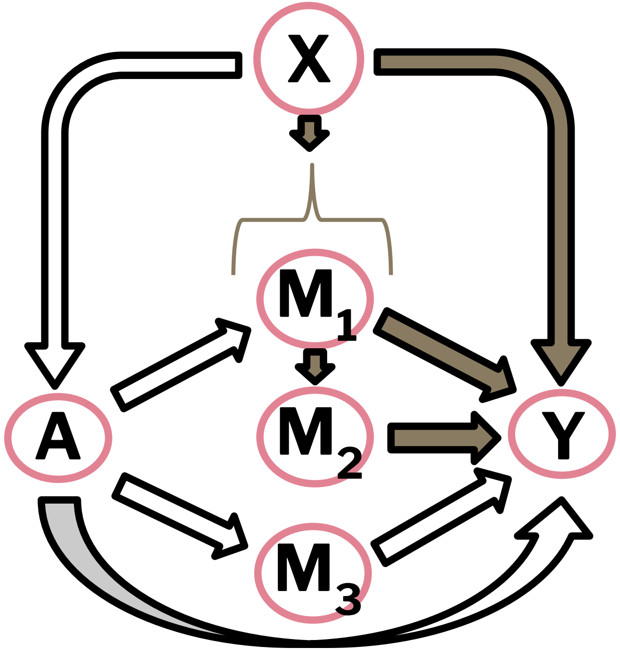
$$HIM_i(\boldsymbol{x}) = \Big[E[M_i|do(A=a+1), \boldsymbol{X}=\boldsymbol{x}] - E[M_i|do(A=a), \boldsymbol{X}=\boldsymbol{x}]\Big] \times \Big[E[Y|do(A=a, M_i=m_i^{(a)}+1), \boldsymbol{X}=\boldsymbol{x}] - E[Y|do(A=a, M_i=m_i^{(a)}+1, \Omega_i=\boldsymbol{o}_i^{(a)}), \boldsymbol{X}=\boldsymbol{x}]\Big]$$



Heterogeneous Causal Effects

Heterogeneous Total Mediation Effect of M_i on Y (HME): The total effect on Y that is due to M_i given the baseline covariates, X

$$HME_i(\boldsymbol{x}) = \left[E\{M_i|do(A=a+1),\boldsymbol{X}=\boldsymbol{x}\} - E\{M_i|do(A=a),\boldsymbol{X}=\boldsymbol{x}\}\right] \times \left[E\{Y|do(M_i=m_i+1),\boldsymbol{X}=\boldsymbol{x}\} - E\{Y|do(M_i=m_i),\boldsymbol{X}=\boldsymbol{x}\}\right]$$



Heterogeneous Causal Effects

Theorem 4.1. *Under assumptions (A1-A3) and the model described in Equation (1), we have:*
1). $HDE(\boldsymbol{x}) = \gamma_A + \boldsymbol{\gamma}_{\mathbf{X}A}\boldsymbol{x}$;
2). $HIE(\boldsymbol{x}) = \boldsymbol{\gamma}_{\boldsymbol{M}}(\boldsymbol{I}_s - \boldsymbol{B}_{\boldsymbol{M}}^\top)^{-1}(\boldsymbol{\beta}_A + \boldsymbol{B}_{\mathbf{X}A}^\top \boldsymbol{x})$;
3). $HTE(\boldsymbol{x}) = HDE(\boldsymbol{x}) + HIE(\boldsymbol{x})$, *where* $\boldsymbol{I}_s$ *is a* $s \times s$ *identity matrix and* $\boldsymbol{x}$ *is the value of* $\boldsymbol{X}$.

Theorem 4.2. *Under assumptions (A1-A3) and Model (1),*
1a). $HDM_i(\boldsymbol{x}) = \{\boldsymbol{\gamma}_{\boldsymbol{M}}\}_i\{(\boldsymbol{I}_s - \boldsymbol{B}_{\boldsymbol{M}}^\top)^{-1}(\boldsymbol{\beta}_A + \boldsymbol{B}_{\mathbf{X}A}^\top \boldsymbol{x})\}_i$;
1b). $\sum_{i=1}^{s} HDM_i(\boldsymbol{x}) = HIE(\boldsymbol{x})$;
2). $HIM_i(\boldsymbol{x}) = HTM_i(\boldsymbol{x}) - HDM_i(\boldsymbol{x})$;
3). $HTM_i(\boldsymbol{x}) = HIE(\boldsymbol{x}) - HIE_{\mathbb{G}(-i)}(\boldsymbol{x})$,
where $\{\cdot\}_i$ *is the ith element of a vector and* $HIE_{\mathbb{G}(-i)}$ *is the HIE under the causal graph* $\mathbb{G}(-i)$ *in which the ith mediator is removed from the original causal graph* $\mathbb{G}$.

$$\boldsymbol{B}^\top = \begin{bmatrix} \mathbf{0}_{p\times p} & \mathbf{0}_{p\times 1} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times s} & \mathbf{0}_{p\times 1} \\ \boldsymbol{\delta}_{\mathbf{X}} & 0 & \mathbf{0}_{1\times p} & \mathbf{0}_{1\times s} & 0 \\ \mathbf{0}_{p\times p} & \mathbf{0}_{p\times 1} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times s} & \mathbf{0}_{p\times 1} \\ \boldsymbol{B}_{\mathbf{X}}^\top & \boldsymbol{\beta}_A & \boldsymbol{B}_{\mathbf{X}A}^\top & \boldsymbol{B}_{\boldsymbol{M}}^\top & \mathbf{0}_{s\times 1} \\ \boldsymbol{\gamma}_{\mathbf{X}} & \gamma_A & \boldsymbol{\gamma}_{\mathbf{X}A} & \boldsymbol{\gamma}_{\boldsymbol{M}} & 0 \end{bmatrix}$$

- Functional versions of both theorems exist
- Sparsity of B can be characterized

Structural Learning

- Equation 1-4 describe the necessary sparsity of B
- h2 can be used alongside the acyclity constraint commonly used in score-based methods

1. $\boldsymbol{X}$ has no parents, i.e.,

$$g_1(\boldsymbol{B}) = \sum_{j=1}^{p} \sum_{i=1}^{2p+s+2} |b_{i,j}| = 0;$$

2. the only parents of A are $\boldsymbol{X}$, i.e.,

$$g_2(\boldsymbol{B}) = \sum_{i=p+1}^{2p+s+2} |b_{i,p+1}| = 0;$$

3. Y has no descendants, i.e.,

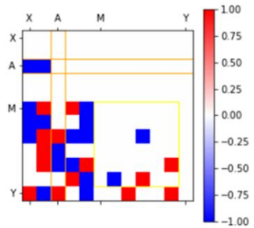
$$g_3(\boldsymbol{B}) = \sum_{i=1}^{2p+s+2} |b_{2p+s+2,i}| = 0;$$

4. the interaction $\boldsymbol{X}A$ also does not have parents, i.e.,

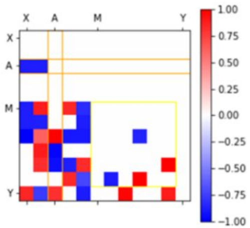
$$g_4(\boldsymbol{B}) = \sum_{j=p+2}^{2p+1} \sum_{i=1}^{2p+s+2} |b_{i,j}| = 0.$$

$$h_2(\boldsymbol{B}) = \sum_{i=1}^{4} g_i(\boldsymbol{B}) = 0$$

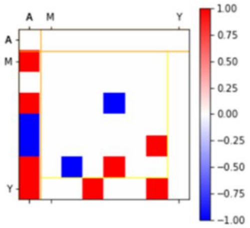
Simulation Results



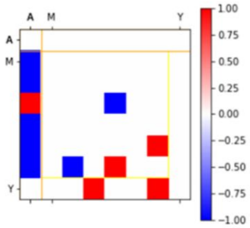
(a) True



(b) ISL($n = 500$)



(c) $\boldsymbol{X} = [1, 0]$



(d) $\boldsymbol{X} = [0, 1]$

- In total there were 8 simulation scenarios and a real data study
- To the left and bottom are the results from scenario 3
- Overall, ISL makes accurate estimates and outperforms its competitors.

	FDR	TPR	SHD
NOTEARS	0.10(0.09)	0.94(0.08)	2.52(1.97)
DAG-GNN	0.03(0.05)	0.98(0.04)	0.70(1.05)
ISL	**0.00(0.01)**	**1.00(0.01)**	**0.04(0.24)**

Thank you for listening!