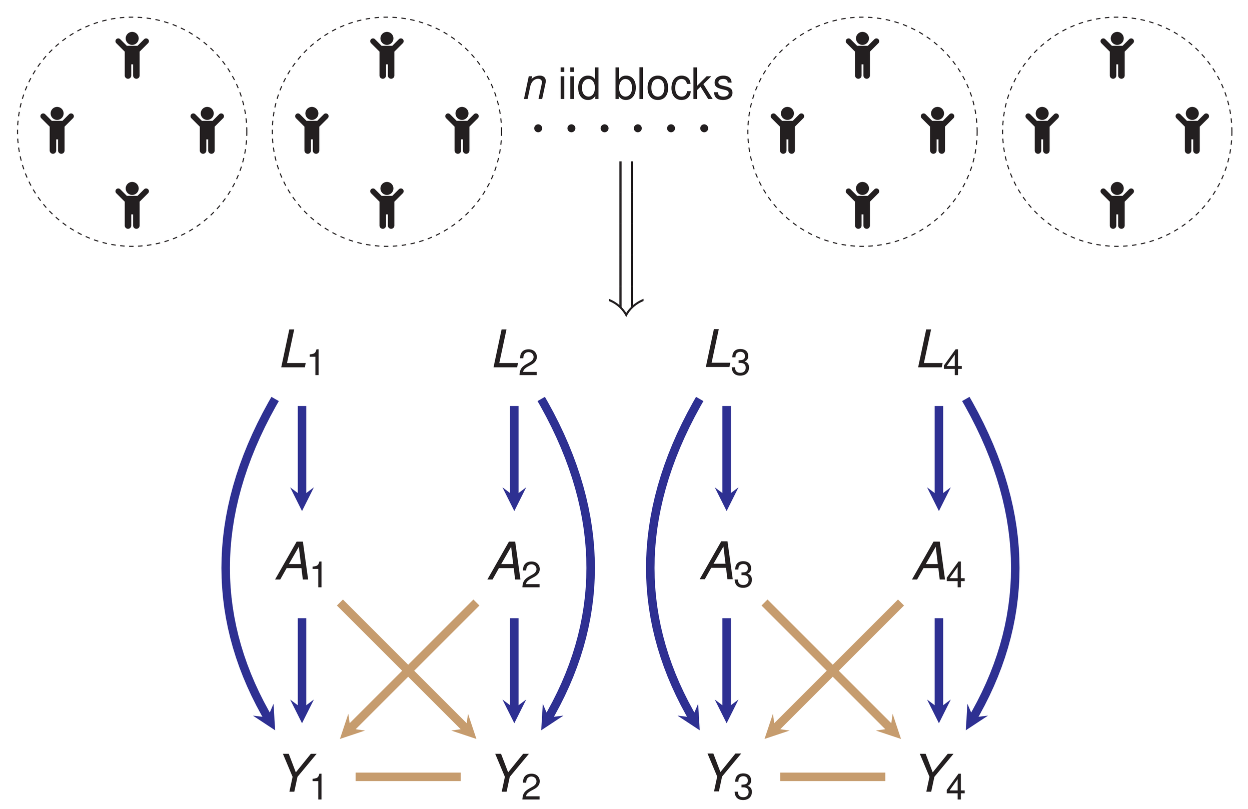
Causal Inference Under Interference And Network Uncertainty

Rohit Bhattacharya, Daniel Malinsky, Ilya Shpitser {*rbhattacharya@*, *malinsky@*, *ilyas*@cs.}jhu.edu
Department of Computer Science
Johns Hopkins University

Motivation

- Classical causal and statistical inference assumes iid data.
- **Homophily**, **contagion**, and **herd immunity** are common happenings in a network that violate independence assumptions.
- Recent work utilizes LWF **Chain Graphs** to model interference but assumes the exact dependence structure is known.
- Studies involving stigmatized communities (e.g, IV drug users) or anonymized databases involve substantial **network uncertainty**.
- We tackle this problem under **partial interference**.



Causal effects e.g., PAOE under treatment assignments $\pi_i(\mathbf{a})$:

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{\mathbf{A}}\mathbb{E}[Y_i(\mathbf{A})]\{\pi_1(\mathbf{A})-\pi_2(\mathbf{A})\}$$

Chain Graph Models And Model Selection

- Statistical models of a CG $\mathcal{G}$

$$p(\mathbf{V}) = \prod_{\mathbf{B}\in\mathcal{B}(\mathcal{G})} p(\mathbf{B} \mid \text{pa}_{\mathcal{G}}(\mathbf{B})) \text{ and}$$

$$p(\mathbf{B} \mid \text{pa}_{\mathcal{G}}(\mathbf{B})) = \frac{\prod_{\{\mathbf{C}\in\mathcal{C}((\mathcal{G}_{\text{bd}_{\mathcal{G}}(\mathbf{B})})^a):\mathbf{C}\not\subseteq\text{pa}_{\mathcal{G}}(\mathbf{B})\}}\phi_{\mathbf{C}}(\mathbf{C})}{Z(\text{pa}_{\mathcal{G}}(\mathbf{B}))}.$$

- Generalization of the g-formula for causal models of a CG:

$$p(\mathbf{V}\setminus\mathbf{A}|\text{do}(\mathbf{a})) = \prod_{\mathbf{B}\in\mathcal{B}(\mathcal{G})} p(\mathbf{B}\setminus\mathbf{A} \mid \text{pa}(\mathbf{B}), \mathbf{B}\cap\mathbf{A})\Big|_{\mathbf{A}=\mathbf{a}}$$

- Our goal is to learn $\rightarrow$, $-$ assuming we know $\rightarrow$.
- $\text{BIC} = 2\ln\mathcal{L}(\mathbf{D};\mathcal{G}) - k\ln(n)$ is a consistent model selection criterion for curved exponential families (Haughton, 1988).
- Common CG parametrizations are curved exponential.
- However, $Z(\text{pa}_{\mathcal{G}}(\mathbf{B}))$ is not easy to evaluate.
- Substitute the likelihood $\mathcal{L}$ with the pseudolikelihood $\mathcal{PL}$.
- We prove that $\text{PBIC} = 2\ln\mathcal{PL}(\mathbf{D};\mathcal{G}) - k\ln(n)$ is a consistent model selection criterion for curved exponential families.
- We also show that the PBIC is **decomposable** as follows:
- Let $\mathbf{B}_{\text{loc}} \equiv \mathbf{B}(V_j)$ if $V_i \rightarrow V_j$ and $\mathbf{B}_{\text{loc}} \equiv \mathbf{B}(\{V_i, V_j\})$ if $V_i - V_j$.
- $\text{loc}(V_i, V_j; \mathcal{G}) \equiv \bigcup_{\mathbf{C}}\{\mathbf{C}\in\mathcal{C}((\mathcal{G}_{\text{bd}_{\mathcal{G}}(\mathbf{B}_{\text{loc}})})^a) : V_i, V_j \in \mathbf{C} \not\subseteq \text{pa}_{\mathcal{G}}(\mathbf{B}_{\text{loc}})\}$
- The change in PBIC for $\mathcal{G}$ and $\mathcal{G}'$ that differ by a single edge is given by $\sum_{V\in\text{loc}(V_i,V_j;\mathcal{G})\cap\mathbf{B}_{\text{loc}}}\{s_V(\mathbf{D};\mathcal{G}) - s_V(\mathbf{D};\mathcal{G}')\}$, where $s_V(.)$ denotes the component of the score for V.

Greedy Network Search

- Exhaustive search is infeasible.
- Local greedy search is consistent for DAGs (Chickering, 2002).
- Under a fixed causal ordering $L_i \rightarrow A_j$, $L_i \rightarrow Y_j$, $A_i \rightarrow Y_j$, and tier symmetry $L_i - L_j$, $A_i - A_j$, $Y_i - Y_j$, for any units i, j; we propose the following greedy search procedure on network ties ($\rightarrow$, $-$):

Algorithm 1 GreedyNetworkSearch ($\mathcal{G}^{\text{init}}$, $\mathbf{D}$)

$\mathcal{G}^* \leftarrow \mathcal{G}^{\text{init}}$
score change $\leftarrow$ True
while score change **do**
  score change $\leftarrow$ False
  $\mathcal{E}^*_{\mathcal{N}} \leftarrow$ network ties ($\rightarrow$, $-$) in $\mathcal{G}^*$
  $E_{max} \leftarrow \text{argmax}_{E\in\mathcal{E}^*_{\mathcal{N}}} \text{PBIC}(\mathbf{D}; \mathcal{G}^*\setminus E)$
  if $\text{PBIC}(\mathbf{D}; \mathcal{G}^*\setminus E_{max}) > \text{PBIC}(\mathbf{D}; \mathcal{G}^*)$ **then**
    $\mathcal{G}^* \leftarrow \mathcal{G}^*\setminus E_{max}$
return $\mathcal{E}^*_{\mathcal{N}}$

Algorithm 2 Heterogenous ($\mathcal{G}^{\text{complete}}$, $\mathbf{D}$)

$\mathcal{G}^{\mathbf{L}}, \mathcal{G}^{\mathbf{A}}, \mathcal{G}^{\mathbf{Y}} \leftarrow$ conditional MRFs on **L**, **A**, **Y** formed from $\mathcal{G}^{\text{complete}}$
$\mathcal{E}^*_{\mathcal{N}_{\mathbf{L}}} \leftarrow \text{GreedyNetworkSearch}(\mathcal{G}^{\mathbf{L}}, \mathbf{D})$
$\mathcal{E}^*_{\mathcal{N}_{\mathbf{A}}} \leftarrow \text{GreedyNetworkSearch}(\mathcal{G}^{\mathbf{A}}, \mathbf{D})$
$\mathcal{E}^*_{\mathcal{N}_{\mathbf{Y}}} \leftarrow \text{GreedyNetworkSearch}(\mathcal{G}^{\mathbf{Y}}, \mathbf{D})$
return $\mathcal{E}^*_{\mathcal{N}_{\mathbf{L}}} \cup \mathcal{E}^*_{\mathcal{N}_{\mathbf{A}}} \cup \mathcal{E}^*_{\mathcal{N}_{\mathbf{Y}}}$

- Further efficiency can be gained through **parameter sharing**.
- For example, the pattern of connection can be assumed to be uniform for any units i, j that are adjacent.
- Alg. 3 and Alg. 4 in Bhattacharya et al. (2019) are procedures that exploit such **homogeneity**.

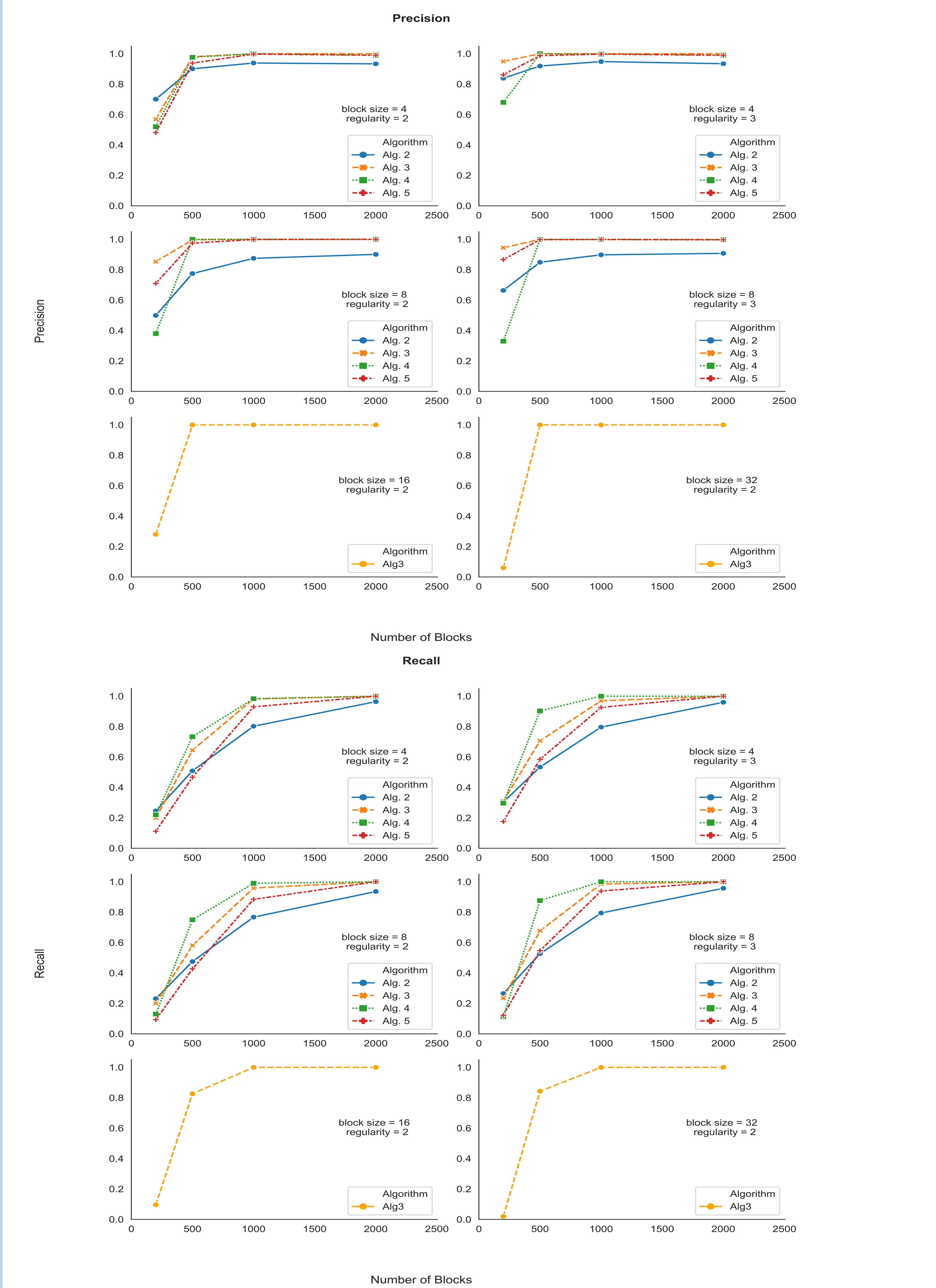
Estimation Of Causal Effects

- $p(\mathbf{Y}(\mathbf{a}))$ is identified as $\sum_{\mathbf{L}} p(\mathbf{Y}|\mathbf{A}=\mathbf{a},\mathbf{L})p(\mathbf{L})$.
- The population average overall effect (PAOE) is identified as:

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{\mathbf{L},\mathbf{A}}\mathbb{E}[Y_i|\mathbf{A},\mathbf{L}]p(\mathbf{L})\{\pi_1(\mathbf{A})-\pi_2(\mathbf{A})\}.$$

- The above can be estimated using the **auto-g computation** algorithm described by Tchetgen Tchetgen et al. (2017).
- Estimates obtained after structure learning exhibit **lower variance** than utilizing the complete graph, and are **unbiased** asymptotically.

Experiments



- Estimation using learned vs complete network.

Block Size	Complete	Homogenous	Heterogenous
4	.009, 9.2e-5	.008, 8.1e-5	.009, 9.7e-5
8	.007, 6.6e-5	.006, 4.1e-5	.006, 4.5e-5
16	.006, 3.8e-5	.005, 1.9e-5	x
32	.007, 6.1e-5	.002, 7.6e-6	x

References

Rohit Bhattacharya, Daniel Malinsky, and Ilya Shpitser. **Causal inference under interference and network uncertainty.** In *UAI*, 2019. (forthcoming).

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3 (Nov):507–554, 2002.

Dominique MA Haughton. On the choice of a model to fit data from an exponential family. *The Annals of Statistics*, 16(1): 342–355, 1988.

Eric J Tchetgen Tchetgen, Isabel Fulcher, and Ilya Shpitser. Auto-g-computation of causal effects on a network. *arXiv preprint arXiv:1709.01577*, 2017.