Forthcoming.

TECHNICAL REPORT
R-527
March 2023

Proceedings of Machine Learning Research vol TBD:1–14, 2023 2nd Conference on Causal Learning and Reasoning

Causal Inference with Non-IID Data under Model Uncertainty

Chi Zhang ZCCC@CS.UCLA.EDU
Computer Science Department, University of California, Los Angeles

Karthika Mohan KARTHIKA.MOHAN@OREGONSTATE.EDU
Electrical Engineering and Computer Science, Oregon State University

Judea Pearl JUDEA@CS.UCLA.EDU
Computer Science Department, University of California, Los Angeles

Editors: Mihaela van der Schaar, Dominik Janzing and Cheng Zhang

Abstract

Algorithms that take data as input commonly assume that variables in the input dataset are Independent and Identically Distributed (IID). However, IID may be breached in many real world datasets that are generated by processes in which units/samples interact with one another. Typical examples include contagion that may be related to infectious diseases in public health, economic crisis in finance and risky behavior in social science. Handling non-IID data (without making additional assumptions) requires access to the true data generating process and the exact interaction pattern among units/samples, which may not be easily available. This work focuses on a specific type of interaction among samples, namely *interference* (i.e. some units' treatments affects other units' outcomes), in situations where there exists uncertainty regarding interaction patterns. The main contributions include modeling uncertain interaction using linear graphical causal models, quantifying bias when IID is incorrectly assumed, presenting a procedure to remove such bias and deriving bounds for average causal effects.

Keywords: Causal Inference, Independent and Identically Distributed (IID), Average Causal Effect, Linear Structural Causal Models

1. Introduction

Almost all Machine Learning (ML) and Causal Inference algorithms are predicated on the assumption that data are Independent and Identically Distributed (IID); however, this is not true in most real-world datasets (Kalaitzis et al., 2013; Schölkopf, 2022). For example, in the analysis of the causal effect of a drug on recovery, it is common to assume that the drug dose, health condition, and other relevant factors pertaining to a patient are *independent* of those pertaining to other patients in the dataset. But this assumption is violated, for instance, in the case of contagious diseases wherein a patient can transmit the disease to a close contact (Lin et al., 2021). IID could also be violated when the treatment (drug taken by a patient) results in lower transmission rates and thus indirectly improves the health condition of a close contact. In this work, we are interested in how causal analysis can be conducted accurately when data are not IID and full information investigating about interaction patterns is not available.

Interference analysis is a line of existing work related to non-IID causal inference that handles interactions among units, more specifically the violation of *independence* condition in IID. Interference was first defined by David Cox, as the phenomenon that a unit's outcome is causally affected by another unit's treatment. *Partial interference* is a specific type of interference that splits the population into "blocks" (usually with the same number of units per block) such that interference interactions only occur between two units that belong to the same block (Ogburn and VanderWeele, 2014). In addition, partial interference requires corresponding units in different blocks to satisfy the *'identical'* condition in IID. Thus partial interference methods assume "block IID," which is weaker than "unit IID" assumed by traditional causal methods. However, in many domains such as infectious diseases, it is unrealistic to assume that the samples in the dataset can be divided into blocks that satisfy the requirements for partial interference. For instance, if blocks pertain to families then all families may not have the same number of members and individuals in the family are likely to interact with people outside the family. Zhang et al. (2022) is a recent work that does not rely on partial interference and instead models true interactions using graphical models called interaction networks that represent general interaction patterns between units, and is not limited to interference. However, one limitation with interaction models and many other approaches (such as those in Aronow and Samii (2017); Jagadeesan et al. (2020)) is that the interaction patterns need to be known in advance. This level of detail is not easily available in real-world datasets. For example, in a drug trial, it may not be feasible to track down each participant; in an online study, it is difficult to know if participants communicated with others in the study. The question of interest is, how can we perform causal analysis given non-IID data when there is uncertainty in the interaction pattern?

The main results of this paper are given below:

1. Quantifying interaction bias when some interaction paths exist with uncertainty (Thm. 11)
2. Reducing or removing bias when some interaction paths exist with uncertainty (Thm. 12)
3. A polynomial algorithm for the bias-reduction/removal method. (Algo. 1)
4. Bounding ACE when some interaction paths exist with uncertainty. (Thm. 13 & Cor. 14)

2. Preliminaries

Independent and Identically Distributed (IID) Let X be a variable and $X_1, \dots, X_n$ be n samples of X. X is IID if $X_1, \dots, X_n$ are independent and each X_i has the same marginal distribution with CDF F(Wasserman, 2013). A dataset is IID if all variables in it are IID.

Linear Causal Models A traditional linear causal model is also known as a linear structural causal model (SCM) (Brito, 2004; Pearl, 2009; Chen and Pearl, 2014). The edge coefficients on the causal DAG represent direct effects.

A collider in a DAG is a structure with two edges connecting three nodes such that the arrows both point to the middle node. An open path is collider-free, i.e., there are no head-to-head arrows on this path. Note that if there exists an open path from W_i to V_j, it implies $W_i \not\perp\!\!\!\perp V_j$. The value of an open path in a linear model is defined as the product of the edge coefficients on that path.

Average Causal Effects In this work the query we are primarily interested in generalizing to the non-IID case is the Average Causal Effect (ACE), also named as the Average Treatment Effect (ATE) (Rubin, 1977; Holland, 1988). For consistency, we use ACE to refer to both. Given a causal model M, the average causal effect (ACE) of $X = t$ vs $X = c$ (t and c are constants) on Y for k units is defined as $ACE_{XY} = \frac{1}{k}\sum_i (Y_{iX_i=t} - Y_{iX_i=c})$. ACE is defined under the assumption that Y_i depends only on factors of unit i (including X_i) Holland (1988). Without loss of generality, we assume $t = c + 1$[1]. In linear models, ACE of X on Y can be identified as β_{YX}, the linear

1. If $t \neq c + 1$, the ACE is multiplied by the constant $(t - c)$.

regression slope of Y on X, if there is no backdoor (non-directed open paths) between X and Y (Pearl et al., 2016; Pearl, 2017). In this paper, we always use X to represent the cause/treatment and Y to represent the effect/outcome.

We refer to the variables in a traditional causal model as *generic variables*. An *explicit variable* is similar to a generic variable except that it represents an attribute/event of one specific unit (or sample or individual). For example, "treatment (X)" is a generic variable, and "the treatment of unit i (X_i)" is an explicit variable.

In the remainder of this section we present definitions from Zhang et al. (2022) that will be used in this work.

$$S^* \begin{cases} X_1 = 0.2C_1 + U_{X_1} \\ X_2 = 0.2C_2 + U_{X_2} \\ X_3 = 0.2C_3 + U_{X_3} \\ Y_1 = 5X_1 - C_2 + U_{Y_1} \\ Y_2 = 5X_2 + X_3 + U_{Y_2} \\ Y_3 = 5X_3 + U_{Y_3} \end{cases}$$

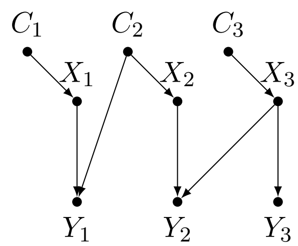

Figure 1: An interaction graph (G^*).

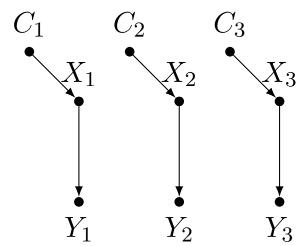

Figure 2: The isolated graph (IG^*).

Definition 1 (Interaction model $M^*(G^*, S^*)$) *An interaction model, $M^*(G^*, S^*)$, is a causal model where G^* is the interaction graph and S^* is the set of structural equations defining the data generating process of the observed explicit variables. An interaction graph, G^*, is a directed acyclic graph with each node representing an explicit variable and each directed edge $A_i \to B_j$ representing A_i causes B_j.*

An example interaction model M^* over 3 units and 9 explicit variables is shown above, with Figure 1 being the interaction graph, G^*, and the structural equations S^* are shown on the left.

Definition 2 (Isolated interaction model $IM^*(IG^*, IS^*)$) *$IM^*(IG^*, IS^*)$ is the* Isolated interaction model *of an interaction model $M^*(G^*, S^*)$ if IM satisfies the following conditions:*

1. *$IG^* = G'$ where G' is the graph obtained by removing from G^* all edges $A_i \to B_j$, $i \neq j$,*
2. *$IS^* = S'$ where S' is the set of equations obtained by removing from each equation $X_i = f(Pa(X_i))$*[2] *in S^* all terms containing any Y_j, $\forall j \neq i$.*

For example, the isolated interaction model of M^* has the interaction graph IG^* as in Figure 2, where all the unit-interacting components are removed.

Definition 3 (Balanced interaction model $M^*(G^*, S^*)$) *Let $M^*(G^*, S^*)$ be an interaction model with isolated model IM^*. M^* is a* balanced *interaction model if IM^* has the same unit-model ($IM_i^*(IG_i^*, IS_i^*)$) for every unit i.*

2. $Pa(X_i)$ denotes the parents of X_i in G^*.

For the model M^* above, after removing all the interaction terms (the orange terms in S^*), the resulting model has the same unit-model for every unit. For example, $Y_i = 5X_i + U_{Y_i}$ applies for every unit i. Hence, M^* is balanced.

Definition 4 (Ancestral same-distribution condition (ASDC)) *In the interaction network G^* a balanced interaction model, generic variable W to satisfies the ancestral same-distribution condition (ASDC) if for all unit i, 1) $Pa(W_i)$ satisfies ASDC, and 2) $Pa(W_i) \subseteq \mathcal{V}_{(i)}$, and 3) for any different unit $j \neq i$, $Pa(W_i)$ and $Pa(W_j)$ have the same set of generic variables, and their exogenous errors U_{W_i} and U_{W_j} have the same distribution. (When i=j, the condition is automatically satisfied.)*

ASDC is a graphical condition that checks if two variables are IID. In the example in Figure 1, C_1, C_2, and C_3 satisfy ASDC. If the error terms for them, U_{C_1}, U_{C_2}, and U_{C_3} are IID, then this implies C_1, C_2, and C_3 are IID. Similarly, X_1, X_2, and X_3 satisfy ASDC and are IID.

Definition 5 (True Average Causal Effect ($TACE_{XY}$)) *Let M^* be an interaction model. True average causal effect of X on Y, denoted as $TACE_{XY}$, is defined as the ACE of X on Y in the isolated interaction model IM^* corresponding to M^*.*

$TACE_{XY}$ is 5 in the example above, since $Y_i = 5X_i$ for all i.

Definition 6 (Interaction bias) *Let balanced model M^* be the true model that generated the (available) non-IID dataset D. Let Q denote the query of interest and let Q^* be its true value. Let A denote an algorithm that outputs an unbiased estimate of Q given data that are IID and the causal graph that generated the IID data. Let $G^\dagger$ denote an approximate causal graph constructed under the assumption that D is IID such that no assumption in $G^\dagger$ is refuted by D. Let $\hat{Q}$ be the estimate computed by A using $G^\dagger$ and D as input. Interaction bias is given by $||Q^* - \hat{Q}||$.*

Interaction bias denotes the size of the bias resulted from the IID assumption, in the estimation of ACE given non-IID data with interacting units. ACE is usually estimated using the ordinary least squares regression coefficient β_{YX}, for linear causal models with IID data and X and Y are not confounded (Pearl, 2017). In this paper, we also assume ordinary least squares regression is used. So the interaction bias we are trying to assess is simply $|E[\hat{\beta_{YX}}] - TACE_{XY}|$.

3. Bias Reduction for Graph with Uncertain Interactions

While the interaction modeling results proposed in Zhang et al. (2022) has the benefit of modeling general arbitrary interactions, they rely on knowing the full interaction graph structure, which is often unavailable. In this section we will generalize their results to handle uncertainty in the interaction patterns among units.

Definition 7 (Uncertain Paths) *An uncertain path between two distinct nodes A and B in a DAG is an open path between A and B that exists with probability θ, $0 < \theta < 1$.*

A *definite* path on the other hand is one that exists with probability 1.

Definition 8 (Uncertain Interaction Graphs) *An uncertain interaction graph is an interaction graph with uncertain paths.*

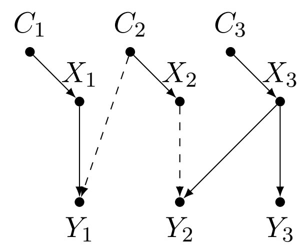

Figure 3: An uncertain interaction graph

Figure 3 shows an uncertain interaction graph, where uncertain paths are represented as dashed arrows, and definite paths are represented as solid arrows.

There are multiple ways in which units can interact, such as two units' outcomes are confounded, a unit's treatment affects its outcome through another unit's variables, etc. In this work we focus on interference, since it is one of the most common and most studied type of interactions. Interference is defined as the phenomenon that one unit's treatment affects another unit's outcome. We assume that the only form of interaction in the interaction model is via interference paths, defined below.

Definition 9 (Interference Paths) *Given an interaction graph, an interference path is a directed path from X_i to Y_j, $i \neq j$.*

We impose a few additional restrictions on the graph so it is not too arbitrary to draw useful conclusions.

Definition 10 (Balanced Graph for Uncertain Interference(b-G^U, for short)) *An interaction graph, G^U, is termed as a* balanced graph for uncertain interference *if*

1. *it is the interaction graph of a balanced interaction model M^*,*
2. *the only type of bias structures in M^* are directed paths from X_i to Y_j where all intermediate nodes belong to either unit i or j.*
3. *only definite edges exist between any two nodes A_i and B_i of unit i, for any i. Uncertain edges may exist only between nodes of distinct units i, j, for any i and j.*
4. *the sum of the values of interference paths from X_i to Y_j (if such exists) is the same as that from X_k to Y_l (if such exists), for all $i \neq j$ and $k \neq l$.*

Note that each interaction graph with or without uncertainty corresponds to an underlying interaction model that encodes the data generating process. Figure 4 is a b-G^U, if the interaction model it corresponds to is balanced. Condition 1 is satisfied. Condition 2 is satisfied since the only such path with an intermediate node is from X_2 to Y_1, with M_2 being an intermediate node, and it belongs to unit 2. Condition 3 is satisfied since the only uncertain edges $M_2 \rightarrow Y_1$ and $X_2 \rightarrow Y_3$ are both between distinct units. As for Condition 4, we can calculate the sum of the values on the three interference paths. The edge coefficients are labeled in Figure 4, and they all equal to 6. Thus, Condition 4 is also satisfied.

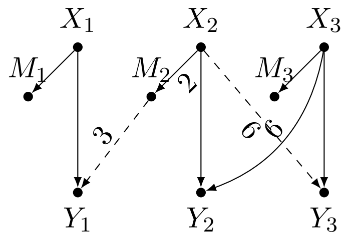

Figure 4: A balanced graph for uncertain interference.

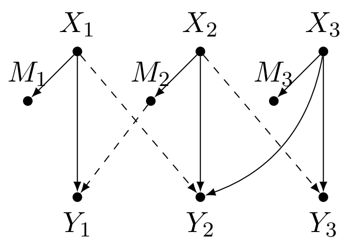

Figure 5: A balanced graph for uncertain interference.

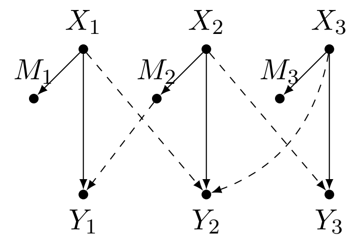

Figure 6: A balanced graph for uncertain interference with $N_d = 0$.

As is mentioned in the preliminaries, the interaction bias (Definition 6) is the bias resulted from incorrectly assuming IID to estimate the unit "true" ACE ($TACE_{XY}$). Theorem 11 below quantifies the interaction bias in an uncertain interaction graph.

Theorem 11 *Suppose $M^*, D, G^\dagger$ refer to the true model, available data and approximate graph as specified in definition 6 such that $Q = TACE_{XY}$ and $\hat{Q} = \beta_{YX}$. X_i and Y_i are not confounded by any variable of i, for all i. Let G^U be the b-G^U corresponding to M^*. For all $i \neq j$ pairs, let N_d be the number of pairs of units that have definite interference paths from i to j and let N_θ be the number of pairs of units that have uncertain interference paths from i to j with probability θ. Let the sum of the values of the interference paths from X_i to Y_j be p,[3] for all $i \neq j$. The expected interaction bias is given by*

$$E[\left|E[\hat{\beta_{YX}}] - Q\right|] = \frac{1}{n(n-1)}|p|(N_d + \theta N_\theta).$$

Figure 5 is a b-G^U with $N_d = 1$ ($X_3 \rightarrow Y_2$) and $N_\theta = 3$ ($X_2 \rightarrow M_2 \rightarrow Y_1$, $X_1 \rightarrow Y_2$, and $X_2 \rightarrow Y_3$). $n = 3$ since there are 3 units. The underlying true interaction model (unavailable) is shown in Figure 7, with the structural equations on the right. The interference effect $|p|$ is equal to 2, calculated from the structural equations. $TACE_{XY}$ is 5. We can also see that the true θ is $2/3$, i.e., out of the 3 uncertain paths, there are 2 that really exist. Although in real-world applications, θ is usually unavailable, so we need an estimate from expert knowledge about the frequency of interference in this sample. Figure 6 shows another b-G^U that corresponds to Figure 7. In Figure

3. I.e., p is equal to the causal effect of X_i on Y_j.

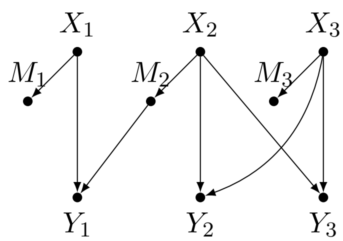

Figure 7: The underlying true interaction graph (unavailable).

$$S^* \text{ (unavailable)} \begin{cases} M_1 = 2X_1 + U_{M_1} \\ M_2 = 2X_2 + U_{M_2} \\ M_3 = 2X_3 + U_{M_3} \\ Y_1 = 5X_1 + M_2 + U_{Y_1} \\ Y_2 = 5X_2 + 2X_3 + U_{Y_2} \\ Y_3 = 5X_3 + 2X_2 + U_{Y_3} \end{cases}$$

6, there is no definite interference paths. This is in fact an interesting special case, which we will elaborate more in the next section.

The debias method in Zhang et al. (2022) selects a bias-free subset of units and uses it to unbiasedly compute TACE given the full interaction graph. When there is uncertainty, if we treat all uncertain interference paths as definite existence, we might end up selecting too small a subset, especially when there are many uncertain interference paths. One solution is to select a larger subset to maybe include some interactions, while still bound the interaction bias at a reasonable level. Theorem 12 below shows such a method.

Theorem 12 *Consider the setting in Theorem 11. Suppose we are additionally given a bias threshold τ, and the interference effect is bounded by a constant Γ times the TACE (i.e., $|p| \leq \Gamma |Q|$). If a subset $\mathcal{B}$ of units satisfies*

$$\frac{1}{|\mathcal{B}|(|\mathcal{B}|-1)}(N'_d + \theta N'_\theta)\Gamma \leq \tau,$$

then using the samples in $\mathcal{B}$, the expected interaction bias will be at most $\tau|Q|$. For all $i \neq j$ pairs with $i, j \in \mathcal{B}$, N'_d denotes the number of pairs with definite interference paths from i to j in G^*, and N'_θ denotes the number of pairs with interference paths from i to j in G^* with probability θ.

If such a subset is found, then the bias is bounded. For example, if the threshold $\tau = 0.1$, the bias will be as large as 10% of the true ACE, computed using the data from the selected subset. This theorem becomes a debias method if $\tau = 0$, since that simply implies that the bias has to be 0. Algorithm 1 is a polynomial greedy algorithm that selects such a subset given threshold τ.

Algorithm 1 goes through all the units, and select units one at a time, until the condition is no longer satisfied, and the selected subset is returned.

4. Causal Effect Estimation with Unknown Interference Structures

Next, we present a theorem for unbiased estimation of TACE. Unbiased estimation is possible if the relationship between the interference path strength and the TACE is given, and where the interference paths occur need not be known.

Theorem 13 *Consider the setting described in Theorem 11. Suppose we know the relationship between p (the interference path strength) and Q (TACE) is $p = \gamma Q$, where γ is a constant, then Q*

Algorithm 1: Select a subset $\mathcal{B}$ from an uncertain interaction graph G^U that makes the interaction bias $\leq \tau$

Input : an interaction graph G^U, probability of uncertain paths θ, interference/TACE ratio bound constant Γ, bias threshold τ
Output: a subset $\mathcal{B}$ resulting in $\leq \tau$ bias
$Units =$ randomly sorted list $1, \ldots, n$;
$\mathcal{B} = Units[1]$;
for $i = 2, \ldots, n$ **do**
  if $\mathcal{B} \cup \{Units[i]\}$ *satisfies* $1/((|\mathcal{B}+1|)|\mathcal{B}|)(N'_d + \theta N'_\theta)\Gamma \leq \tau$ **then**
    $\mathcal{B} = \mathcal{B} \cup \{Units[i]\}$;
  end
end
return $\mathcal{B}$

is unbiasedly estimated as

$$Q = \frac{E[\hat{\beta_{YX}}]}{1 - \frac{1}{n(n-1)}\gamma(N_d + \theta N_\theta)}$$

Applying Theorem 13 to generate bounds In this work there are two types of effects under consideration. First, the effect of X_i on Y_i (unit specific effect) and second, the effect of treatment applied to other units such as X_j, $j \neq i$ on Y_i (interference). In many situations such as when treatment is vaccination and outcome is disease, (i) the magnitude of unit-level treatment effects (TACE) can be safely assumed to be higher than those due to interference (p); mathematically, this translates to $|Q| > |p|$ and $0 < \gamma < 1$.

Corollary 14 *Consider the setting described in Theorem 13, if we further assume* $0 < \gamma < 1$ *and* $|Q| > |p|$ *and* $0 < \gamma < 1$ *then Q can be bounded as,*

$$\frac{E[\hat{\beta_{YX}}]}{1 - \frac{(N_d + \theta N_\theta)}{n(n-1)}} < Q < E[\hat{\beta_{YX}}]$$

Note that from Corollary 14, Q is always less than $E[\hat{\beta_{YX}}]$. This implies that when the unit specific effect and the interference effect have the same sign, then assuming IID ($E[\hat{\beta_{YX}}]$) always “overestimates” the true unit specific effect (Q).

Remark 15 *Note that there are several interesting special cases with the the results presented in Theorems 11, 12, and 13.*

1. $N_d = N_\theta = 0$. *In this case, there is no interference path (definite or uncertain) in the model, which results in a model without interaction structures. In Theorem 11, the interaction bias is 0. In Theorem 12, the inequality always hold since the l.h.s. is 0, while* τ *is positive, so we can select any subset* $\mathcal{B}$ *where* $|\mathcal{B}| > 1$*. In Theorem 13,* $Q = E[\hat{\beta_{YX}}]$*, which is consistent with an interference-free setting.*

2. $N_d = 0$. *In this case, there is no definite interference path. This special case is useful when we do not have any information about which units interact with which units in some real-world applications. All those theorems still apply.*

3. $N_\theta = 0$. *In this case, there is no uncertain interference path. This means we have all the information regarding which units interact with which units. The theorems reduce to the results in Zhang et al. (2022), where there is no uncertainty in the interaction network.*

5. Related Work

Interference was first defined by David Cox in his 1958 book, "Planning for Experiments" (Cox, 1958). Handling interference is non-trivial and majority of literature in empirical fields assume no-interference. In fact, SUTVA is a common assumption in causal inference (Rubin, 1978). However, in some fields such as epidemiology that deal with health care data and infectious diseases, ignoring interference can lead to biased outcomes and decisions that can put lives at risk; unsurprisingly, a big chunk of literature on interference comes from these fields.

Recent years have witnessed a rise in papers on interference that employ graphical models. This includes Ogburn and VanderWeele (2014); Sherman and Shpitser (2018); Bhattacharya et al. (2020); Zhang et al. (2022). Ogburn and VanderWeele (2014) was the first to model the problem of interference using DAGs. Sherman and Shpitser (2018) models interference using chain graphs which permits modeling unknown interactions between individuals. Bhattacharya et al. (2020) proposes a method to do structure learning for chain graphs.

Graham et al. (2010) estimates spillover effects with the focus on the social reallocating problems. Nabi et al. (2020) shows that interference is a problem even in applications related to adplacement and develops methods for identification and estimation of multiple queries under conditions of interference and homophily, assuming partial interference. Sobel (2006) is the first to notice the effect of interference in the housing mobility problem, and proposes causal estimands for this application.

Aronow and Samii (2017), Sussman and Airoldi (2017) model general interference (without assuming partial inference) by constructing a function to define an individual's exposure level on the number of treated neighbors they have. The methods are less restricted than partial interference methods, and allow individuals to be affected by any number of neighbors. However, the methods are limited to the type of interference where one individual's treatment affects another individual's outcome. Jagadeesan et al. (2020) proposes a quasi-coloring method to estimate direct effect under interference using experimental data, which is useful in the setting where it is allowed to design the experiment. But it does not easily generalize to observational studies. Another paper in a similar direction, Fatemi and Zheleva (2020), proposes experiment design to minimize interference bias and selection bias at the same time. Liu and Hudgens (2014) proposes a two-stage randomization design to minimize interference bias.

Tchetgen Tchetgen et al. (2021) proposes a g-computation method for general interference. Their method is the first to model general interference using DAGs (chain-graphs) and requires the interference effects to be symmetrical between individuals, and fits parameters for chain-graph models.

Hudgens and Halloran (2008) defines six types of queries in the problems involving interference. Work in interference that focuses on different queries/problems include a few as follows. VanderWeele et al. (2012) is the first to decompose the spillover effect (the effect of an individual's treatment on another's outcome (Quammen (2012))) to contagion and infectiousness effects using counterfactual mediation analysis. Shpitser et al. (2017) does decomposition for individuals with unknown and symmetrical interaction patterns. Such decomposition permits analysis of different interference paths. In our case (linear models), the contagion and infectiousness effects reduce to the interference paths from X_j to Y_i through Y_j (contagion) and not through Y_j (infectiousness), respectively. Hu et al. (2021) is the first to define and provide estimands for the average indirect effect. VanderWeele et al. (2014) develops methods for sensitivity analysis under interference.

6. Conclusions

This work focused on the problem of interference when there is uncertainty regarding the interaction patterns. We showed that bias due to interference can be quantified using the interference strength and expected number of interactions. We developed an algorithm that computes true average causal effect such that bias is guaranteed to be less than a given quantity τ. Finally, we bound the average causal effect when it is guaranteed that unit level causal effect is higher than interference.

7. Acknowledgement

This research was supported in parts by grants from the National Science Foundation [#IIS-2106908 and #IIS-2231798], Office of Naval Research [#N00014-21-1-2351], and Toyota Research Institute of North America [#PO-000897].

References

Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.

Rohit Bhattacharya, Daniel Malinsky, and Ilya Shpitser. Causal inference under interference and network uncertainty. In *Uncertainty in Artificial Intelligence*, pages 1028–1038. PMLR, 2020.

Carlos Brito. *Graphical Models for Identification in Structural Equation Models*. PhD thesis, UCLA, 2004.

Bryant Chen and Judea Pearl. Graphical tools for linear structural equation modeling. 2014.

D.R. Cox. *Planning of Experiments*. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section. Wiley, 1958. ISBN 9780471181835.

Zahra Fatemi and Elena Zheleva. Minimizing interference and selection bias in network experiment design. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1): 176–186, May 2020. URL https://ojs.aaai.org/index.php/ICWSM/article/view/7289.

Bryan S Graham, Guido W Imbens, and Geert Ridder. Measuring the effects of segregation in the presence of social spillovers: A nonparametric approach. Working Paper 16499, National Bureau of Economic Research, October 2010. URL http://www.nber.org/papers/w16499.

Paul W Holland. Causal inference, path analysis and recursive structural equations models. *ETS Research Report Series*, 1988(1):i–50, 1988.

Yuchen Hu, Shuangning Li, and Stefan Wager. Average direct and indirect causal effects under interference. *Biometrika*, 2021.

Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

Ravi Jagadeesan, Natesh S Pillai, and Alexander Volfovsky. Designs for estimating the treatment effect in networks with interference. *The Annals of Statistics*, 48(2):679–712, 2020.

Alfredo Kalaitzis, John Lafferty, Neil D Lawrence, and Shuheng Zhou. The bigraphical lasso. In *International Conference on Machine Learning*, pages 1229–1237. PMLR, 2013.

Yi-Jheng Lin, Che-Hao Yu, Tzu-Hsuan Liu, Cheng-Shang Chang, and Wen-Tsuen Chen. Positively correlated samples save pooled testing costs. *IEEE Transactions on Network Science and Engineering*, 8(3):2170–2182, 2021. doi: 10.1109/TNSE.2021.3081759.

Lan Liu and Michael G Hudgens. Large sample randomization inference of causal effects in the presence of interference. *Journal of the american statistical association*, 109(505):288–301, 2014.

Razieh Nabi, Joel Pfeiffer, Murat Ali Bayir, Denis Charles, and Emre Kıcıman. Causal inference in the presence of interference in sponsored search advertising. *arXiv preprint arXiv:2010.07458*, 2020.

Elizabeth L Ogburn and Tyler J VanderWeele. Causal diagrams for interference. *Statistical science*, 29(4):559–578, 2014.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Judea Pearl. A linear "microscope" for interventions and counterfactuals. *Journal of causal inference*, 5(1), 2017.

Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

David Quammen. *Spillover: animal infections and the next human pandemic*. WW Norton & Company, 2012.

Donald B Rubin. Assignment to treatment group on the basis of a covariate. *Journal of educational Statistics*, 2(1):1–26, 1977.

Donald B Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58, 1978.

Bernhard Schölkopf. Causality for machine learning. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 765–804. 2022.

Eli Sherman and Ilya Shpitser. Identification and estimation of causal effects from dependent data. *Advances in neural information processing systems*, 31, 2018.

Ilya Shpitser, Eric Tchetgen Tchetgen, and Ryan Andrews. Modeling interference via symmetric treatment decomposition. *arXiv preprint arXiv:1709.01050*, 2017.

Michael E Sobel. What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407, 2006.

Daniel L Sussman and Edoardo M Airoldi. Elements of estimation theory for causal effects in the presence of network interference. *arXiv preprint arXiv:1702.03578*, 2017.

Eric J Tchetgen Tchetgen, Isabel R Fulcher, and Ilya Shpitser. Auto-g-computation of causal effects on a network. *Journal of the American Statistical Association*, 116(534):833–844, 2021.

Tyler J VanderWeele, Eric J Tchetgen Tchetgen, and M Elizabeth Halloran. Components of the indirect effect in vaccine trials: identification of contagion and infectiousness effects. *Epidemiology (Cambridge, Mass.)*, 23(5):751, 2012.

Tyler J VanderWeele, Eric J Tchetgen Tchetgen, and M Elizabeth Halloran. Interference and sensitivity analysis. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 29 (4):687, 2014.

L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer New York, 2013. ISBN 9780387217369. URL https://books.google.com/books?id=qrcuBAAAQBAJ.

Chi Zhang, Karthika Mohan, and Judea Pearl. Causal inference with non-IID data using linear graphical models. In *Proceedings of Thirty-Sixth Conference on Neural Information Processing Systems*, 2022.

Appendix A. Proof

Theorem 11 *Suppose $M^*, D, G^\dagger$ refer to the true model, available data and approximate graph as specified in definition 6 such that $Q = TACE_{XY}$ and $\hat{Q} = \beta_{YX}$. X_i and Y_i are not confounded by any variable of i, for all i. Let G^U be the b-G^U corresponding to M^*. For all $i \neq j$ pairs, let N_d be the number of pairs of units that have definite interference paths from i to j and let N_θ be the number of pairs of units that have uncertain interference paths from i to j with probability θ. Let the sum of the values of the interference paths from X_i to Y_j be p,[4] for all $i \neq j$. The expected interaction bias is given by*

$$E[\Big|E[\hat{\beta_{YX}}] - Q\Big|] = \frac{1}{n(n-1)}|p|(N_d + \theta N_\theta).$$

Proof Let G^* be the true interaction graph corresponding to M^* (no uncertainty). By Theorem 1 in Zhang et al. (2022),

4. I.e., p is equal to the causal effect of X_i on Y_j.

$$\left|E[\hat{\beta_{YX}}] - Q\right| = \left|\frac{1}{n}\sum_{1\le i\le n}\sum_{p\in P[iji]} Val(p)\frac{\sigma^2_{R_p}}{\sigma^2_X} - \frac{1}{n(n-1)}\sum_{1\le i\le n}\sum_{p\in P[ji]} Val(p)\frac{\sigma^2_{R_p}}{\sigma^2_X}\right|.$$

Under our settings, there is no reflecting bias structure, but only deflecting bias structures. So the first term on the r.h.s. is 0. Theorem 1 becomes

$$\left|E[\hat{\beta_{YX}}] - Q\right| = \left| - \frac{1}{n(n-1)}\sum_{1\le i\le n}\sum_{p\in P[ji]} Val(p)\frac{\sigma^2_{R_p}}{\sigma^2_X}\right|.$$

The sole term on the r.h.s. is the sum of all deflecting bias paths' strengths multiplied by the variance factors, divided by $1/(n(n_1))$. Under our settings, interference paths are the only deflecting bias structures, and the roots of those paths are all X_i for some i. So we have $\sigma^2_{R_p}/\sigma^2_X = 1$. In addition, the summation is over all interference paths. It can be rearranged as summing over all pairs of $i \neq j$, and for each pair, sum over all the interaction paths. For each pair, the summation of all the interaction paths is the same and equal to p, from our assumptions. Hence, Theorem 1 can be further simplified as

$$\left|E[\hat{\beta_{YX}}] - Q\right| = \left| - \frac{1}{n(n-1)} Np\right|,$$

where N is the total number of ordered pairs of $i \neq j$ where there are interference paths from X_i to Y_j. The term on the r.h.s. inside of the absolute symbols except p is less than 0. So Theorem 1 becomes

$$\left|E[\hat{\beta_{YX}}] - Q\right| = \frac{1}{n(n-1)} N|p|.$$

Hence, we have

$$\begin{aligned} E[\left|E[\hat{\beta_{YX}}] - Q\right|] &= E[\frac{1}{n(n-1)} N|p|] \\ &= \frac{1}{n(n-1)}|p|E[N]. \end{aligned} \tag{1}$$

Next we evaluate $E[N]$. N is the sum of the pairs whose interference paths appear in G^U as definite paths, and the pairs whose interference paths appear in G^U as uncertain paths. The first part is simply N_d. The second part's expectation is θN_θ by the definition of θ. So we have

$$E[N] = N_d + \theta N_\theta.$$

Plugging this into Equation 1, we have the equation in the statement of Theorem 11. ■

Theorem 12 *Consider the setting in Theorem 11. Suppose we are additionally given a bias threshold τ, and the interference effect is bounded by a constant Γ times the TACE (i.e., $|p| \le \Gamma|Q|$). If a subset $\mathcal{B}$ of units satisfies*

$$\frac{1}{|\mathcal{B}|(|\mathcal{B}|-1)}(N'_d + \theta N'_\theta)\Gamma \le \tau,$$

then using the samples in $\mathcal{B}$, the expected interaction bias will be at most $\tau|Q|$. For all $i \neq j$ pairs with $i, j \in \mathcal{B}$, N'_d denotes the number of pairs with definite interference paths from i to j in G^*, and N'_θ denotes the number of pairs with interference paths from i to j in G^* with probability θ.

Proof Consider the sub-graph G_{sub} formed by projecting the true interaction graph G^* on $\mathcal{B}$. variables of units in $\mathcal{B}$. Consider an interference path from X_i to Y_j, where $i \neq j$ and $i, j \in \mathcal{B}$. It does not go through a third unit by our assumptions, and since $i, j \in \mathcal{B}$, the path remains unchanged in G_{sub}. Let G_{usub} be the uncertain sub-graph formed by projecting G^U on $\mathcal{B}$. For each pair $i \neq j$ in the original uncertain sub-graph G^U such that the interference paths from X_i to Y_j are uncertain, consider the following scenarios.

1. If $i, j \in \mathcal{B}$, then from the previous discussion, the interference paths from X_i to Y_j remains unchanged after the projection. So the probability of those paths existing is still θ.
2. If $i \in \mathcal{B}, j \notin \mathcal{B}$, then the interference paths are removed in G_{usub}.
3. If $i \notin \mathcal{B}, j \in \mathcal{B}$, if we also have interference paths from X_i to Y_k with $k \neq j$, this will result in a bidirected path between Y_j and Y_k. However, this can be ignored since it is not a bias structure, by Definitions 7 and 8 in Zhang et al. (2022).

As a result, we can apply Theorem 11 on G_{usub}, and obtain the interaction bias as

$$\frac{1}{|\mathcal{B}|(|\mathcal{B}|-1)}(N'_d + \theta N'_\theta)|p|.$$

Plugging in $|p| \leq \Gamma|Q|$, we have the interaction bias is at most $\tau|Q|$ by Theorem 12. ■

Theorem 13 *Consider the setting described in Theorem 11. Suppose we know the relationship between p (the interference path strength) and Q (TACE) is $p = \gamma Q$, where γ is a constant, then Q is unbiasedly estimated as*

$$Q = \frac{E[\hat{\beta_{YX}}]}{1 - \frac{1}{n(n-1)}\gamma(N_d + \theta N_\theta)}$$

Proof From the proof of Theorem 1 in Zhang et al. (2022), we have a slightly stronger result than Theorem 11, which is Theorem 11 without the absolute signs. We have

$$E[\hat{\beta_{YX}}] - Q = -\frac{1}{n(n-1)}p(N_d + \theta N_\theta).$$

Plugging in $p = \gamma Q$, we have the expression in the theorem statement. ■

Corollary 16 *Consider the setting described in Theorem 13, if we further assume $0 < \gamma < 1$ and $|Q| > |p|$ and $0 < \gamma < 1$ then Q can be bounded as,*

$$\frac{E[\hat{\beta_{YX}}]}{1 - \frac{(N_d + \theta N_\theta)}{n(n-1)}} < Q < E[\hat{\beta_{YX}}]$$

Proof Q is monotonic with respect to γ. Hence, Corollary 14 results from Theorem 13 by plugging in $\gamma = 0$ and $\gamma = 1$. ■