A Targeted Learning Framework for Policy Evaluation with Unobserved Network Interference

Anonymous authors
Paper under double-blind review

Abstract

Estimating causal effects under network interference is a fundamental yet challenging task, especially when the network structure is represented as multiple layers or multiple views. In this paper, we consider a heterogeneous network setting, where the ties from different views of the network might achieve varying levels of interference. Meanwhile, dependence among units is allowed, due to information transmission among network ties and latent traits among units sharing ties (i.e., latent dependency). To the best of our knowledge, this setting has not been studied in literature yet. We propose a novel framework that conducts doubly robust estimation on heterogeneous networks with latent dependency. Our approach relies on a new identification strategy and integrates it with targeted maximum likelihood estimation for robust causal effect estimation from observational data. Crucially, our approach remains valid even when the outcome prediction model or data-generating process is misspecified. It also supports counterfactual inference under hypothetical network interventions using only the observed network structure. Experiments on both synthetic and real-world networks show that our approach consistently outperforms existing baselines and can provide robust estimation towards different intervention policies.

1 Introduction

Identifying and estimating causal effects in networks has received increasing attention over the past years. In real world networks, a unit's outcome may depend on not only the treatment they assigned to but also on their neighbors, violating the classical SUTVA assumptions (van der Laan, 2014). This phenomenon is referred to as *interference* in literature. A common approach to model interference is to define exposure as the proportion of treated neighbors (Forastiere et al., 2022), which allows the estimation of spillover effects, i.e., the effect of varying neighborhood exposure on a unit's outcome. In contrast to modeling exposure solely as a proportion, Toulis & Kao (2013) introduced a framework that defines exposure through the exact k-level exposure, i.e., the potential outcome of a node when exactly k of its neighbors are treated. Another way to model network interference involves modeling the transmission of information through network ties using graph neural networks (GNN) based methods (Tchetgen et al., 2019; Ma et al., 2021; 2022; Ma & Tresp, 2021) or directly modeling the confounders to address latent network dependency (Guo et al., 2019; Ma et al., 2021; Zhang et al., 2017).

However, real-world networks may be more complex. One complexity comes from the heterogeneous relationships among units. For example, in Amazon dataset(He & McAuley, 2016), ties may represent different types of interaction such as co-viewing and co-purchasing, and the interference may vary between different relational aspects. Lin et al. (2023) proposed HINITE, which aims to estimate the effect of treatment in a heterogeneous setting by aggregating the covariate information on different aspects of the networks. Related efforts also include modeling heterogeneity in peer influence using structured graph-based causal models (Adhikari & Zheleva, 2025), and modeling heterogeneity in causal pathways across subpopulations with different moderator variables (Watson et al., 2023). These methods work well when the outcome model is correctly specified for the data, however, they are usually sensitive to model misspecification.

To avoid failure caused by misspecification between the data generation process and the modeling process, the estimation of maximum likelihood is a general framework to construct an objective estimate (Van Der Laan & Rubin, 2006; Van der Laan et al., 2011). This method has been widely used in various settings (van der Laan & Gruber, 2012; Kreif et al., 2017; Chen et al., 2023; Balzer et al., 2019). Shi et al. (2019) proposed targeted regularization for binary treatment effect estimation, while Nie et al. (2021) proposed an end-to-end targeted maximum likelihood estimation (TMLE) framework to estimate the continuous treatment effect. Based on previous work, Chen et al. (2024) proposed T-Net, which models the potential outcome as a function of unit treatment and neighborhood exposure in a homogeneous network setting. In this line of research, only one method (i.e., T-Net) is designed for causal inference on networked data, but it still cannot account for heterogeneity and latent dependency in the network structure.

To close the gap, we propose a novel framework, named **M**ulti-**v**iew **d**oubly **r**obust estimator (**Mvdr**), for causal inference under heterogeneous network interference. The challenge of heterogeneity remains valid even when only a single partially observed network is available. Our goal is to estimate the average expected potential outcome under different hypothetical intervention policies over an observed network with heterogeneous ties. Our approach models interference via multi-view representations of the network and integrates these with a TMLE procedure. In this way, the learned estimators are doubly robust, i.e., they remain consistent if either the outcome model or the exposure model is correctly specified. Our Mvdr framework enables robust estimation of average potential outcomes under a variety of hypothetical interventions, such as static, dynamic, or stochastic interventions. The major contributions of this paper can be summarized as follows:

1. We develop a doubly robust estimator for causal inference on partially observed networks exhibiting latent network dependence. We prove that the estimator's efficient influence function is invariant to the admissible summary mappings, and this EIF invariance yields double robustness.
2. To examine the proposed method in real-world scenarios such as policy evaluation, we design new evaluation settings with dynamic and stochastic interference beyond the standard static setting.
3. We constructed a new DBLP dataset and conduct extensive experiments. Results demonstrate the correctness of our theory and effectiveness of our model.

The rest of this paper is organized as follows. In Section 2, we brielfy present related work on causal inference under network interference, complex interference structure, and model misspecification. Section 3 introduces preliminary and the problem settings. Section 4 presents our theoretical analysis. In Section 5, we introduce the technical details of the proposed Mvdr framework. Section 6 provides experimental results with discussions, and Section 7 concludes this paper.

2 Related Works

Causal Inference under Network Interference. Classical causal inference frameworks assume that there is no interference between units, but real-world settings often violate this assumption. The foundational work in (Halloran & Struchiner, 1995; Sobel, 2006; Hudgens & Halloran, 2008) formalized potential outcomes under interference. Manski (2013) and Aronow & Samii (2017) introduced exposure mappings, summarizing neighbor treatments into interpretable variables. Building on this, Karwa & Airoldi (2018) investigated the consequences of misspecified exposure mappings in randomized experiments, while Leung (2022) studied approximate neighborhood interference and provided inference methods under specification error. More recently, Ogburn et al. (2022) developed a semiparametric TMLE framework for causal inference on a single network, allowing dependence to grow with network degree. Sävje (2024) emphasized the distinction between exposures as effect definitions versus structural assumptions, proposing expected exposure effects that remain meaningful under misspecification.

Complex Interference Structure. This motivates the construction of proxy networks based on observed covariates. Egami (2021) considered spillover effects when the interference network is only partially observed, while Li & Wager (2022) studied network interference under latent graphon models and equilibrium settings. Parallel work in graph mining has explored multiview and multiplex networks, where multiple graphs encode different relations among the same nodes. Such methods aggregate heterogeneous views through attention or weighting. These ideas have not yet been fully integrated into causal inference. Our work is motivated by the observation that units may be dependent not only on direct social ties, but also on shared environments, socioeconomic status, or latent similarities. Thus, constructing multiple similarity-based networks from observed covariates provides a flexible way to capture such heterogeneity.

Model Misspecification. A growing literature addresses robustness when models meet with misspecification with data generating process. Leung (2022) proposed alternative estimands for approximate interference, while Sävje (2024) showed that conventional estimators remain unbiased for expected exposure effects under weakly dependent specification errors. Chao et al. (2025) studied misclassified or surrogate networks, introducing methods to correct bias using validation data. Similarly, Hoshino et al. (2024) considered causal inference with noncompliance and unknown network structures. These works highlight that perfect knowledge of the interference graph is rarely available, and meaningful causal estimands can still be defined under approximation. Our approach draws directly on these insights. By treating similarity-based graphs as proxies rather than true causal pathways, we target expected exposure effects that remain interpretable and estimable even if the constructed networks do not fully align with the underlying social process.

3 Preliminary

In this section, we first introduce the notations used in this paper and then describe the preliminary of our work, such as the structural equation model and assumptions. Finally, we present the problem setting and introduce the causal estimands of interest in detail.

3.1 Notations

Let $x_i \in \mathbb{R}^d$ denote the covariates of unit i, $t_i \in \{0,1\}$ denote the binary treatment assignment, and $y_i \in \mathbb{R}$ denote the observed outcome for unit i under treatment t_i. We then use $X = (x_1, \ldots, x_n) \in \mathbb{R}^{n\times d}$, $T = (t_1, \ldots, t_n) \in \{0,1\}^n$, and $Y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ denote the covariates, treatment assignments, and observed outcomes for all n units in a network, respectively.

Suppose that the observed network G is of size n and denote the adjacency matrix of G as $A \in \{0,1\}^{n\times n}$. If there exists an edge between unit i and j, $A_{ij} = 1$; otherwise, $A_{ij} = 0$. In practice, network G can be represented as multiple layers or multiple views (e.g., with adjacency matrices $A^{(1)}, \cdots, A^{(K)}$ for K views), and each view captures a specific relationship between units. The observed network G can be seen as an aggregation of these multi-view networks. In this work, we consider the case of undirected edges, where influence between connected units is assumed to be symmetric. A table of all notations used in this paper is provided in Appendix A.

3.2 Structural Equation Model

We assume that the observed data arise from a complete network of size n. Units in this network may exhibit latent variable dependence, meaning that unobserved traits shared by connected units can induce correlations in their covariates, treatments, or outcomes (Shalizi & Thomas, 2011). This type of dependence is typically stronger between units who are closer in the network topology, which is dependent on the network information but not necessarily fully observed from the adjacency matrix.

To formalize the data generating process, we adopt the structural equation model (SEM) framework (Pearl, 2012) and assume that the data are generated by sequentially evaluating the following set of equations:

$$\begin{aligned} &X_i = f_X[\epsilon_{X_i}],\ i = 1, ..., n \\ &T_i = f_T[\{X_j : A_{ij} = 1\}, \epsilon_{T_i}],\ i = 1, ..., n \\ &Y_i = f_Y[\{T_j : A_{ij} = 1\}, \{X_j : A_{ij} = 1\}, \epsilon_{Y_i}],\ i = 1, ..., n \end{aligned} \tag{1}$$

where f_X, f_T, and f_Y are unknown and unspecified functions that may depend on the unit i. ϵ_{X_i}, ϵ_{T_i}, and ϵ_{Y_i} are all exogenous, unobserved errors for the unit i. The errors may be correlated across units. This SEM

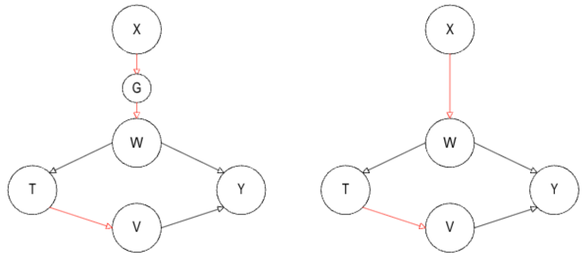

Figure 1: Structural equation model for the causal assumptions. The left one is which our model built on, where the G represents learned graph structure.

function implies our assumption, i.e., the outcome of unit i depends on the treatments and covariates of i's neighborhoods through a function $f_Y(\cdot)$.

Based on this SEM model, we can then define summary functions s_X and s_T as well as random variables $W_i = s_{X,i}(\{X_j : A_{ij} = 1\})$ and $V_i = s_{T,i}(\{T_j : A_{ij} = 1\})$. Then, the model in Eq. (1) can be rewritten as:

$$\begin{aligned} X_i &= f_X[\epsilon_{X_i}],\ i = 1, ..., n \\ T_i &= f_T[W_i, \epsilon_{T_i}],\ i = 1, ..., n \\ Y_i &= f_Y[V_i, W_i, \epsilon_{Y_i}],\ i = 1, ..., n \end{aligned} \tag{2}$$

Eq. (2) means that the outcome Y_i depends on $(\mathbf{X}, \mathbf{T})$ only through summary functions $(s_{X,i}, s_{T,i})$. The summary function $s_{X,i}$ implies that the exposure and outcome of node i only depend on some function of i and its neighbors' covariates. Analogously, $s_{T,i}$ also aggregate the information for T.

3.3 Assumptions

Based on the SEM model, we make the following assumptions:

Assumption 1 (Pre-treatment network construction): Each view-specific network $A^{(m)} = G_m(C)$ is constructed solely from pre-treatment covariates C. In particular, edges are functions of the observed baseline features and do not depend on treatment T or outcome Y.

Assumption 2 (Ignorability): Error vectors $(\epsilon_{T_1}, ..., \epsilon_{T_N})$, $(\epsilon_{Y_1}, ..., \epsilon_{Y_N})$, and $(\epsilon_{X_1}, ..., \epsilon_{X_N})$ are mutually independent.

Assumption 3 (Exchangeability): Errors $\epsilon_{T_1}, ..., \epsilon_{T_N}$ and $\epsilon_{X_1}, ..., \epsilon_{X_N}$ are identically distributed.

Assumption 4 (Local Latent Dependence Structure): The error terms $\varepsilon_{X_i} \not\perp \varepsilon_{X_j}$ (and analogously for $\varepsilon_{Y_i}, \varepsilon_{C_i}$) are permitted to be dependent if and only if nodes i and j are connected through a common neighbor, i.e., if $\exists\, k$ such that $A_{ik} = A_{kj} = 1$.

Assumption 5 (Positivity): For all i, $P(V_i = v \mid s_{X,i}(X) = s_{X,i}(x)) > 0$ for all x in the support of X and all v in the support of V.

In addition, the SEM model in Eq. (2) encodes the assumption that $\mathbf{X}$ suffices to control for the confounding effect of $\mathbf{T}$ on $\mathbf{Y}$. This is a version of the *conditional ignorability* assumption that is typical in *i.i.d* settings. Assumption 1 avoids post-treatment bias and ensures that the network can be regarded as fixed prior to interventions. Assumption 4 allows for the existence of latent dependency among network but limited within two hops. Assumption 5 is critical for the identifiability of treatment effects.

3.4 Problem Setting and Estimands

Our goal is to estimate the causal effects of hypothetical intervention policy T^* applied to the network. We can construct an average expected outcome that evaluates the causal effect of an observed network receiving

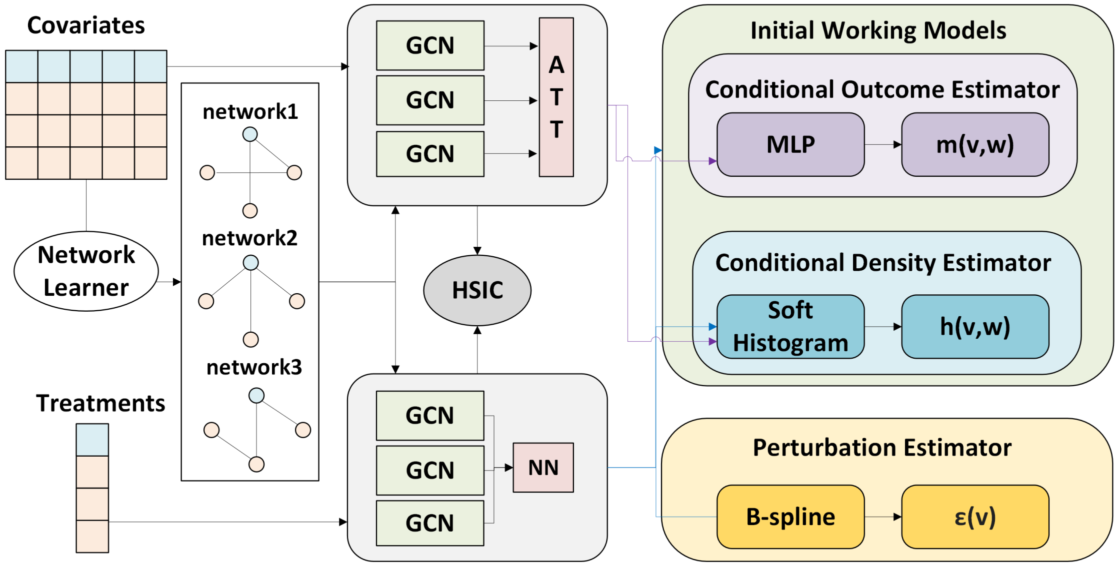

Figure 2: Illustration of our Mvdr framework. Mvdr contains three modules. The representation module learns the summary function of $\mathbf{X}$ and aggregate the information from heterogeneous network. The initial working model contains one conditional outcome estimator m and a conditional density estimator h. The perturbation estimator is used to conduct TMLE.

certain interventions based on this independent relationship (Van der Laan, 2014):

$$E[\bar{Y}_n^*] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i^*] = \frac{1}{n}\sum_{i=1}^{n} E[m(s_{T,i}(T^*), s_{X,i}(x))] = \frac{1}{n}\sum_{i=1}^{n}\sum_{w} m(v_i^*, w)h_i^*(v_i^*.w), \tag{3}$$

where h refers to the conditional distributions of W and V: $h_i(v \mid w) = P(V_i = v \mid W_i = w)$ and $h_i(v, w) = P(V_i = v, W_i = w)$. $m(v, w) = \sum_y y p_Y(y \mid v, w)$ is the conditional expectation of the outcome Y given $V = v$ and $W = w$.

The causal estimands of interest are identified by functionals of the observed data distribution $P(X, T, Y)$, which depends on N and A in our assumption. This is also known as the identifying function. It is a parameter of the observed data distribution for a network of size N, that is, a parameter of the data generating distribution that gave rise to the data at hand. It is an unknown parameter rather than an observed quantity, because the data we observe comprise a single random draw from $P(X, T, Y)$.

The identifying function of our estimands can be written as:

$$\mathcal{M} = \mathcal{M}_h \times \mathcal{M}_m,$$

where $\mathcal{M}$ represents the class of functions that can be used to estimate the two parameters h and m. m is an initial estimation model of the potential outcome, which can be seen as a function of the treatment summary V and covariate summary W. This identifying function shows that our estimand depends on two model classes, i.e., m model class to estimate the potential outcome given a certain variable (v_i, w_i), and a conditional density estimator h that computes the probability of a certain realization (v_i, w_i).

4 Theoretical Analysis

4.1 Efficient Influence Curve

The efficient influence curve is a key ingredient in semi-parametric efficient estimation, which defines the linear approximation of any efficient and regular asymptotically linear estimator. Therefore, it provides an asymptotic bound for the variance of all regular asymptotically linear estimations. Under the assumptions made in Section 3.3, the influence function for the causal estimand of interest, which is denoted as ψ_N, can be written as:

$$D_N(o) = \frac{1}{n}\sum_{i=1}^{n}(E[m(V_i^*, W_i) \mid X = x] - \psi_n + \frac{\bar{h}^*(v_i, w_i)}{\bar{h}(v_i, w_i)}\{y_i - m(v_i, w_i)\}), \tag{4}$$

where $\bar{h}(v_i, w_i) = \frac{1}{n}\sum_{j=1}^{n} h_j(v_i, w_i)$ and $\bar{h}^*(v_i, w_i) = \frac{1}{n}\sum_{j=1}^{n} h_j^*(v_i, w_i)$. $D_n(o)$ will have an expected value of 0 at the true ψ_n.

Estimating equations for the parameters indexing a working model for m and h are stacked with the influence function estimating equation for ψ_n (Kennedy, 2016). The key idea of the targeted maximum likelihood estimation (TMLE) is to make the influence function equal to 0. The simplified steps of TMLE include: (1) estimate an initial working model m to predict the potential outcome Y based on (V, W), and conditional density for the observed treatment h and target intervention h^*; (2) fit the influence function with initial estimation of m and $\frac{h^*}{h}$ to make it equal to 0 to estimate ϵ; (3) use estimated ϵ to calculate final results $\widetilde{Y}$.

The estimators with this influence function are doubly robust. The right hand side of Eq. (4) has an expected value equal to 0 if $m(\cdot)$ is replaced with an arbitrary function of V or if $h(\cdot)$ is replaced with an arbitrary function of W, as long as one of the two remains correctly specified.

Let $O = (W, V, Y)$ denote the observed variables, where W and V are aggregated summaries constructed from multi-view networks. Write $m(v, w) = \mathbb{E}[Y \mid V = v, W = w]$, $\bar{h}(v \mid w) = \Pr(V = v \mid W = w)$ (or density if V is continuous), and p_W the marginal law of W. We consider two common targets.

Target A (Static neighbor intervention). For a fixed v_0 in the support of V, define

$$\psi \;=\; \Psi(P) \;=\; \mathbb{E}\{\, m(v_0, W) \,\} \;=\; \int m(v_0, w)\, p_W(dw).$$

Under the nonparametric model $M = M_m \times M_h \times M_{p_W}$, the efficient influence function (EIF) is:

$$D^*(O) \;=\; \frac{\mathbb{1}\{V = v_0\}}{\bar{h}(v_0 \mid W)}\big\{Y - m(v_0, W)\big\} \;+\; m(v_0, W) - \psi. \tag{5}$$

If V is continuous, we can replace the indicator ratio by a Radon–Nikodým derivative using a kernel density estimator.

Target B (Stochastic/dynamic neighbor intervention). Let $h^*(\cdot \mid w)$ be a stochastic intervention on V that may depend on W. Define

$$\psi \;=\; \Psi(P) \;=\; E_W\Big[\, E_{V^* \sim h^*(\cdot \mid W)}\{\, m(V^*, W) \,\}\Big] \;=\; \iint m(v, w)\, h^*(dv \mid w)\, p_W(dw).$$

The efficient influence function is

$$D^*(O) \;=\; \frac{\bar{h}^*(V \mid W)}{\bar{h}(V \mid W)}\big\{Y - m(V, W)\big\} \;+\; \int m(v, W)\, h^*(dv \mid W) - \psi. \tag{6}$$

When V is discrete, $\int m(v, W)\, h^*(dv \mid W) = \sum_v m(v, W) h^*(v \mid W)$.

The multi-view construction only affects the definitions of V and W. We now establish the large-sample properties of our estimator. Under Assumptions 1–5, the target parameter is identifiable and the TMLE estimator is consistent and asymptotically normal. In particular, Assumptions 1–3 ensure identifiability, Assumption 5 guarantees positivity so that causal contrasts are well-defined with finite variance, and Assumption 4 restricts dependence to local neighborhoods, allowing the use of network central limit theorem arguments. As is standard in semiparametric theory, these results hold under mild regularity conditions such as bounded moments and smoothness of the relevant estimators. The detailed conditions we used to obtain the CAN of our estimator are provided in Appendix C.

Theorem 4.1 (Consistency and Asymptotic Normality of TMLE). *Let $\hat{\psi}_n$ denote the TMLE for the target parameter ψ. Under Assumptions 1–3 and Regularity Conditions (1) and (2),*

$$\sqrt{C_n}\,(\hat{\psi}_n - \psi) \;\xrightarrow{d}\; N(0, \sigma^2),$$

where C_n is an effective sample size satisfying $n/K^2_{\max,n} \leq C_n \leq n$. The asymptotic variance σ^2 equals the variance of the efficient influence function under the true data-generating process.

5 Methodology

In this section, we introduce our *Mvdr* framework, which estimates the causal estimands of interest discussed above. As illustrated in Figure 2, our framework consists of three components: confounder representation learning with multi-view feature balancing, density estimator, and outcome prediction. Our motivation is to find a way to integrate information from different views of the network. Inspired by the SEM framework, we propose a novel way to present the transmission of information to units through summary functions and construct them as multi-view networks. Technically, our method offers a semi-parametric way to estimate two classes of parameters h and m included in the function class $\mathcal{M}$ that decides our model. More specifically, the outcome predictor m is a parametric model dependent on the covariate summary w and the exposure summary v, and the density estimator h is estimated through every realization of (w, v).

5.1 Multi-view Representation of Network

Based on the assumptions of causal relationship, the outcome Y is affected by covariates X through the summary function W and by exposure T through another summary function V. In the confounder representation learning module, we use K graph convolutional network (Kipf & Welling, 2017) blocks to represent the summary function $s_{x,i}$ that aggregates information of covariates of unit i and its neighbors on multi-view networks (i.e. X_i, $A^{(k)}$, $k = 1, ..., K$):

$$h_i^{(k)} = \sigma(\sum_{j \in N_i^{(k)}} \frac{1}{\sqrt{d_i^{(k)} d_j^{(k)}}} W_k^T x_j; \theta_k), \tag{7}$$

where $\sigma(\cdot)$ is a sigmoid activation function, $d_i^{(k)}$ is the degree of unit i in the k-th network, and W_k is the weight matrix of GCN parameterized by θ_k. To integrate multiple views of structural information and identify the importance of different views of the networks, we use an attention-based fusion module. Given node embeddings $h_i^{(k)} \in \mathbb{R}^d,\ k = 1, \cdots K$ from the GCNs, the scores can be computed as:

$$\alpha_{ik} = \frac{\exp(\mathbf{w}^T h_i^{(k)} + b)}{\sum_{j=1}^K \exp(\mathbf{w}^T h_i^{(k)} + b)}, \tag{8}$$

and the final multi-view representation of unit i can be written as:

$$w_i = \sum_{k=1}^K \alpha_{ik} \cdot h_i^{(k)}. \tag{9}$$

For the exposure summary, we use another stack of GCN layers to learn the aggregation of neighborhood treatment, and then an MLP layer is adopted to learn the weights of exposure from different views. The final exposure mapping V is then combined with two variables: unit treatment T and multi-view neighborhood exposure S.

To ensure that the GCNs for different views of the network can learn different aspects of information, we use a ***multi-view network representation discrepancy*** module to maximize the information learned from the multiple GCNs. Specifically, we hope that the summary functions W_i, which is sued to learn the information from covariates, and S_i, which aims to aggregate information from multiple views, can learn different information for the final estimation. In particular, we design a loss function based on the Hilbert-Schmidt independence criterion (HSIC) as follows:

$$L_{HSIC}(W,S) = \frac{1}{n^2} tr(KMLM),\ M = \mathbb{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T, \tag{10}$$

where n is the number of training units, $\mathbb{I}_n$ represents the identity matrix, and $\mathbf{1}_n$ is the vector of all ones. K and L represent the Gaussian kernel applied to W and S, respectively, i.e.,

$$K_{ij} = \exp(-\frac{||w_i - w_j||_2^2}{2}),\ L_{ij} = \exp(-\frac{(s_i - s_j)^2}{2}). \tag{11}$$

5.2 Essential Predictors

As we discussed earlier, three essential parts are used to achieve our doubly robust estimator, including the conditional outcome estimator, the conditional density estimator, and the perturbation estimator.

For the ***conditional outcome estimator*** m, given the covariate representations W and treatment representations V obtained from the mlti-view representation module, we use two multi-layer perceptron (MLP) blocks that consist of three feed-forward layers to infer the outcome y_i for the treated and control groups respectively. This potential outcome predictor is used for initial outcome prediction, representing m in the identifying function of the SEM model. Specifically, m_0 denotes the predictor for the treated units, and m_1 denotes the predictor for the controlled units. These potential outcome predictors are optimized by the mean square error (MSE) between the predicted outcomes and observed outcomes:

$$L_1 = \frac{1}{n}\sum_{i=1}^{n}(m_{t_i}(v_i, w_i) - y_i)^2. \tag{12}$$

The ***conditional density estimator*** h is the conditional distribution of the joint exposure mapping $V = (T,S)$ given covariates W, denoted by:

$$h((t,s) \mid w) \ = \ \Pr(T = t \mid W = w) \ \cdot \ f_{S|T,W}(s \mid t, w). \tag{13}$$

This factorization separates the discrete ego treatment T from the continuous neighbor summary S.

Ego Treatment Model. We estimate $\Pr(T = 1 \mid W = w)$ using an MLP classifier. Given input w, the network outputs $\hat{e}(w) = \sigma(g_\theta(w))$, where g_θ is an MLP and σ is the sigmoid function. The model is trained by minimizing the binary cross-entropy loss:

$$L_e(\theta) \ = \ -\frac{1}{n}\sum_{i=1}^{n}\Big[T_i \log \hat{e}(W_i) + (1 - T_i)\log(1 - \hat{e}(W_i))\Big]. \tag{14}$$

Neighbor Summary Model. For the second factor $f_{S|T,W}$, we approximate the conditional density of S given (T,W). Following the conditional histogram approach (Díaz Muñoz & Van Der Laan, 2011), we normalize v into $[0,1]$ and partition its support into C intervals. For each interval, we fit a multinomial logistic regression model that outputs bin-wise probabilities:

$$\pi_i = \mathrm{softmax}\big(W_{\mathrm{dens}}^\top(x_i \| w_i) + b\big) \in \mathbb{R}^C, \tag{15}$$

where W_{dens} are learnable parameters and $\|$ denotes concatenation. The conditional density at v_i is then obtained by interpolating between the corresponding lower and upper bins:

$$\hat{h}(v_i \mid x_i, w_i) = \pi_{i,L_i} + (v_i \cdot C - L_i) \cdot \big(\pi_{i,U_i} - \pi_{i,L_i}\big), \tag{16}$$

where L_i and U_i denote the indices of the bins bracketing v_i. This estimator can be viewed as a flexible, data-adaptive approximation to the Radon–Nikodým derivative of the target intervention distribution with respect to the observed distribution. It is trained by minimizing the negative log-likelihood:

$$L_n = -\sum_{i=1}^{n}\log \hat{h}(v_i \mid x_i, w_i). \tag{17}$$

Putting all the two density models together, the final density model is trained by minimizing the negative log-likelihood:

$$L_2 = L_{\mathrm{e}} + L_{\mathrm{n}} = -\sum_{i=1}^{n} \log \hat{h}((x_i, s_i) \mid w_i). \tag{18}$$

To achieve doubly robust estimation, we use an MLP block to learn ϵ that makes the influence function equal to 0. The loss function of our ***perturbation estimator*** is:

$$L_3 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{m}_i - \frac{\epsilon}{\hat{h}_i})^2. \tag{19}$$

Combining the three modules for our estimation together, we have the final loss function:

$$Loss = \sum_{i=1}^{N} (y_i - \widetilde{y}_i)^2 - \sum_{i=1}^{n} \log \hat{h}((x_i, s_i) \mid w_i) + \sum_{i=1}^{n} (y_i - \hat{m}_i - \frac{\epsilon}{h_i})^2 + \lambda L_{HSIC}. \tag{20}$$

5.3 Inference under Interventions Policies

After training the initial outcome predictor m and the density estimator h, we purpose the procedure of predicting average potential outcome under certain intervention policy. Suppose our intervention towards the network is $T^* \in \{0,1\}^n$. We first calculate the initial potential outcome $\hat{y}_i = m_{t_i^*}(v_i^*, w_i)$. Secondly, to achieve the robust estimator of the final result, we need to compute the auxiliary weights as:

$$H_i = \frac{\hat{h}^*(v_i^*, w_i)}{\hat{h}(v_i, w_i)} \tag{21}$$

For the last step, we construct the final outcome prediction with targeted learning. The final estimation for the unit i can be written as:

$$\widetilde{Y}_i^* = m_{t_i^*}(v_i^*, w_i) + \epsilon(v_i^*, w_i) \cdot H_i, \tag{22}$$

and the average potential outcome under the intervention T^* is given by:

$$\bar{\widetilde{Y}}^* = \frac{1}{n} \sum_{i=1}^{n} \widetilde{Y}_i^*. \tag{23}$$

This estimator is doubly robust, as it will be consistent for the inference function if either h or m is correctly specified for the model.

6 Experiment

In this section, we validate the proposed method *Mvdr* on two commonly used semi-synthetic datasets. We also construct a new semi-synthetic dataset based on the DBLP dataset. We verify the effectiveness of our method and further evaluate the correctness of our analysis using these semi-synthetic datasets. In particular, we aim to answer the following research questions (RQ):

- RQ1: How does the proposed method compare with existing methods in terms of effect estimation performance?
- RQ2: How does the perturbation estimator module affect the performance of our methods?
- RQ3: Does our method stably perform well under different types of misspecification?

In the following, we first introduce the experimental setup and then answer the questions above by conducting the corresponding experiments.

Datasets. In our experiment, we utilize two widely used semi-synthetic datasets, i.e., ***BlogCatalog*** and ***Flickr***. Another contribution of this paper is that we construct a new semi-synthetic dataset based on the ***DBLP*** dataset. The DBLP dataset contains rich information of conference and journal publications, such as detailed information of authors. We construct a coauthorship network by matching the author information and the publication information. Detailed statistics of three datasets are provided in Appendix B.

Multi-view network construction. A simple idea of constructing multi-view networks is to compute the similarity scores of covariates between units and construct the relational networks based on their similarity. We follow the multi-view network construction method from (Lin et al., 2023), in which multi-view networks were constructed for the BlogCatalog and Flickr datasets based on the similarity of covariates. The constructed latent networks can then be potential for the upcoming multi-view information aggregation. Notice that these latent multi-view networks are learned from the covariates and are not directly observed from the network data.

Since it is impossible to observe the counterfactual outcomes, we need to mimic the unknown output. Following prior studies on ITE and ATE, we simulate outcomes as ground-truth values for counterfactual outcomes that are not available using the following equation:

$$Y_i = \beta_T \cdot T_i + \beta_X \cdot \left(\mathbf{w}^\top \mathbf{x}_i\right) + \beta_{NX} \cdot \sum_{k=1}^{3} \frac{1}{|\mathcal{N}_i^{(k)}|} \sum_{j \in \mathcal{N}_i^{(k)}} \mathbf{w}^\top \mathbf{x}_j + \beta_{NT} \cdot \left(\sum_{k=1}^{3} \alpha_k \cdot \frac{\sum_{j \in \mathcal{N}_i^{(k)}} T_j}{|\mathcal{N}_i^{(k)}|} \right) + \varepsilon_i. \tag{24}$$

This construction method assumes that the exposure mapping is a simple function of the ratio of treated neighbors across views. Here, $\mathbf{x}_i \in \mathbb{R}^d$ denotes the covariate vector for unit i, and $T_i \in \mathbb{R}$ is the assigned treatment. The neighborhood sets $\mathcal{N}_i^{(k)}$ are derived from three adjacency matrices $A^{(1)}, A^{(2)}, A^{(3)}$, representing multi-view network structures. Each network layer $k \in \{1, 2, 3\}$ contributes to the potential outcome through aggregated covariate signals and treatment exposures, weighted by coefficients α_k. We set $\alpha_1 = \alpha_2 = 1$ and $\alpha_3 = 0.5$ in our experiments to reflect the varying network influence. Lastly, $\varepsilon_i \sim \mathcal{N}(0, 1)$ introduces stochastic variation.

6.1 Three Types of Interventions

To evaluate the stability of our method and baselines under different types of interventions, we design three interventions to mimic different hypothetical treatment policies in real life. **Static intervention** assigns all the units in the network as treated or not treated. **Dynamic intervention** assigns exposures as a user-specified, deterministic function of covariates. This scenario often appears when there are some constraints in resources. For example, the economic incentive was resource constrained and could only be allocated to up to 10% of the whole group. In our experiments, we allocate the top 10% most connected members of the community. **Stochastic intervention** assigns exposures as a user-specified, random function. In our experiments, we assign each unit to treatment with a constant probability of 0.35.

6.2 Baselines and Metrics

Baselines. We compare our method with several representative baselines in causal inference, including methods that deal with a single observed network and multi-view networks. CFR (Johansson et al., 2023) and ND (Guo et al., 2019) are modified with additionally inputting the exposure. TARNet (Shalit et al., 2017) has a similar model architecture as CFR but removes the balance term. HINITE (Lin et al., 2023) considers several observed heterogeneous networks. NetEst (Jiang & Sun, 2022) is designed for treatment effect estimation under interference. TNet (Chen et al., 2024) uses the proportion of treated neighbors as the exposure mapping, which is a doubly robust method.

Evaluation Metric. The main purpose of our model is to predict the average expected outcome of hypothetical intervention policies. We evaluate the accuracy of the model predicting the average expected outcome of a group receiving treatment under different intervention policies, written as:

$$\varepsilon_{intervention} = |\widetilde{Y}_n^* - Y_n^*|, \text{ where } Y_n^* = \frac{1}{n} \sum_{i=1}^{n} Y_i^*.$$

Table 1: Comparison of methods on causal effect of network estimand with different interventions. 'Static0' sets all units to control; 'Static1' sets all to treated; 'Stochastic' applies treatment to 35% randomly; 'Dynamic' assigns treatment based on covariate-dependent probabilities. Best performance in each row is **bolded**. Results are reported as the mean and standard deviation over five runs for each setting.

Dataset		Intervention	CFR	CFR+z	ND	ND+z	TARNet	TARNet+z	NetEst	HINITE	TNet	Mvdr(w/o. $\mathcal{L}_3$)	Mvdr
BlogCata	With-in-Sample	Static0	$0.4428_{0.2618}$	$\mathbf{0.4034}_{0.2625}$	$0.4466_{0.2508}$	$0.4093_{0.2576}$	$0.4457_{0.2442}$	$0.4082_{0.2726}$	$0.3861_{0.2576}$	$1.0292_{0.7670}$	$16.346_{17.3932}$	$0.4968_{0.2366}$	$0.4891_{0.1899}$
		Static1	$0.5381_{0.1829}$	$0.4302_{0.2203}$	$0.5766_{0.1959}$	$0.3953_{0.2357}$	$0.5873_{0.2077}$	$0.5045_{0.1933}$	$0.8412_{0.2356}$	$1.2021_{0.9884}$	$7999.1908_{12974.5585}$	$0.3035_{0.1633}$	$\mathbf{0.2574}_{0.1515}$
		Stochastic	$0.5833_{0.2865}$	$0.5747_{0.2845}$	$0.5821_{0.3006}$	$0.5797_{0.2920}$	$0.5881_{0.3175}$	$0.5813_{0.3119}$	$0.2751_{0.1782}$	$0.6227_{0.2099}$	$81.2073_{123.1866}$	$0.2818_{0.1699}$	$\mathbf{0.2668}_{0.1847}$
		Dynamic	$0.4858_{0.2952}$	$0.4869_{0.3001}$	$0.4910_{0.3025}$	$0.4974_{0.3033}$	$0.4919_{0.3108}$	$0.5065_{0.3007}$	$0.5001_{0.3032}$	$0.5585_{0.2657}$	$23.1303_{47.0434}$	$0.4486_{0.3433}$	$\mathbf{0.4401}_{0.3636}$
	Out-of-Sample	Static0	$0.4251_{0.2345}$	$0.3723_{0.2388}$	$0.4316_{0.1992}$	$\mathbf{0.3283}_{0.2780}$	$0.4758_{0.2521}$	$0.5224_{0.2921}$	$0.4005_{0.348}$	$0.9796_{0.7588}$	$15.6192_{14.4018}$	$0.4509_{0.2701}$	$0.4174_{0.3391}$
		Static1	$0.5011_{0.1543}$	$0.3958_{0.2316}$	$0.4179_{0.4720}$	$0.5031_{0.3988}$	$0.4843_{0.3408}$	$0.3502_{0.3040}$	$0.7291_{0.3454}$	$1.1732_{0.9409}$	$5735.9332_{9219.6814}$	$0.2940_{0.2183}$	$\mathbf{0.2279}_{0.1549}$
		Stochastic	$0.4755_{0.2916}$	$0.4590_{0.2745}$	$0.4478_{0.3245}$	$0.4914_{0.3420}$	$0.5313_{0.3889}$	$0.5067_{0.4134}$	$0.4446_{0.3048}$	$0.5065_{0.3123}$	$112.701_{133.1668}$	$0.1974_{0.1821}$	$\mathbf{0.1847}_{0.1784}$
		Dynamic	$0.6421_{0.3101}$	$0.6385_{0.3204}$	$0.6324_{0.3334}$	$0.5830_{0.3148}$	$0.6721_{0.3047}$	$0.6414_{0.3250}$	$0.6623_{0.3347}$	$0.7148_{0.3062}$	$18.7178_{39.9793}$	$0.4785_{0.3798}$	$\mathbf{0.4500}_{0.3886}$
Flickr	With-in-Sample	Static0	$0.1710_{0.0766}$	$\mathbf{0.0887}_{0.0986}$	$0.1747_{0.0768}$	$0.1002_{0.0231}$	$0.1816_{0.0700}$	$0.1700_{0.0676}$	$0.1014_{0.0747}$	$0.3486_{0.3168}$	$5.9266_{8.5908}$	$0.1311_{0.0371}$	$0.1086_{0.0896}$
		Static1	$0.1751_{0.1185}$	$\mathbf{0.0994}_{0.0327}$	$0.2522_{0.1844}$	$0.0923_{0.0613}$	$0.2683_{0.1981}$	$0.1347_{0.0902}$	$0.4473_{0.0706}$	$1.1022_{0.8934}$	$18.1979_{16.6327}$	$0.3878_{0.0494}$	$0.3788_{0.0658}$
		Stochastic	$0.1410_{0.0777}$	$0.1368_{0.0967}$	$0.1647_{0.0974}$	$0.1908_{0.1021}$	$0.1779_{0.1059}$	$0.2181_{0.0993}$	$0.3243_{0.0986}$	$0.2045_{0.1504}$	$10.8568_{11.5352}$	$0.1995_{0.0742}$	$\mathbf{0.1316}_{0.0375}$
		Dynamic	$0.1793_{0.1497}$	$0.2085_{0.1607}$	$0.2109_{0.1635}$	$0.2305_{0.0678}$	$0.2279_{0.1714}$	$0.1905_{0.0503}$	$0.5980_{0.1442}$	$0.1436_{0.1705}$	$2934.7736_{3022.9666}$	$0.0401_{0.041}$	$\mathbf{0.0338}_{0.0325}$
	Out-of-Sample	Static0	$\mathbf{0.1241}_{0.0995}$	$0.1407_{0.1073}$	$0.1419_{0.0812}$	$0.1391_{0.1172}$	$0.1642_{0.0738}$	$0.1289_{0.1133}$	$0.1376_{0.0892}$	$0.3111_{0.2683}$	$14.9494_{19.1093}$	$0.2731_{0.0744}$	$0.2163_{0.0486}$
		Static1	$0.1927_{0.1098}$	$\mathbf{0.1565}_{0.1069}$	$0.2960_{0.1432}$	$0.2018_{0.0665}$	$0.3354_{0.1231}$	$0.2048_{0.0773}$	$0.2352_{0.0867}$	$1.0141_{0.8732}$	$26.4602_{21.5615}$	$0.634_{0.1683}$	$0.4648_{0.1063}$
		Stochastic	$0.1424_{0.1079}$	$0.1598_{0.0834}$	$0.2542_{0.1024}$	$0.1864_{0.0732}$	$0.1471_{0.1169}$	$0.2180_{0.1140}$	$0.3173_{0.0793}$	$0.2248_{0.1756}$	$4.87_{5.5107}$	$0.1245_{0.0409}$	$\mathbf{0.1172}_{0.0718}$
		Dynamic	$0.1642_{0.1354}$	$0.1703_{0.1496}$	$0.2032_{0.1661}$	$0.2905_{0.1676}$	$0.2036_{0.1606}$	$0.1293_{0.0574}$	$0.7368_{0.1807}$	$0.3059_{0.2397}$	$2141.9216_{1912.0244}$	$0.0470_{0.0152}$	$\mathbf{0.0402}_{0.0287}$
DBLP	With-in-Sample	Static	$0.6801_{0.6428}$	$0.7406_{0.6214}$	$0.8354_{0.5585}$	$0.6755_{0.5753}$	$0.9776_{0.8482}$	$0.7969_{0.6272}$	$0.9383_{0.593}$	$0.8162_{0.5422}$	$7.2816_{17.7863}$	$0.7142_{0.4798}$	$\mathbf{0.5933}_{0.5328}$
		Static1	$0.6128_{0.4203}$	$0.6682_{0.4753}$	$0.6194_{0.3356}$	$0.7100_{0.6311}$	$\mathbf{0.5597}_{0.3182}$	$0.6378_{0.4699}$	$0.697_{0.2554}$	$1.4879_{1.1256}$	$25.8228_{50.5348}$	$1.1797_{0.5966}$	$1.1800_{0.5476}$
		Stochastic	$0.6548_{0.5236}$	$0.7344_{0.5142}$	$0.7434_{0.4676}$	$0.8192_{0.5609}$	$0.7264_{0.4199}$	$0.8145_{0.5024}$	$0.6705_{0.427}$	$1.0153_{0.6379}$	$28.2031_{124.7104}$	$0.8348_{0.5777}$	$\mathbf{0.6602}_{0.4815}$
		Dynamic	$0.5924_{0.5628}$	$0.5944_{0.5645}$	$0.5956_{0.4799}$	$0.5701_{0.6118}$	$0.7413_{0.6965}$	$0.7391_{0.7125}$	$0.6482_{0.6376}$	$0.7816_{0.7768}$	$7.0901_{25.3574}$	$0.7038_{0.5519}$	$\mathbf{0.5633}_{0.5301}$
	Out-of-Sample	Static	$0.6520_{0.6234}$	$0.7282_{0.5922}$	$0.8258_{0.5324}$	$0.756_{0.5221}$	$0.8693_{0.7477}$	$0.8637_{0.5882}$	$0.976_{0.5703}$	$1.0222_{0.4386}$	$5.0336_{16.0037}$	$0.6208_{0.4500}$	$\mathbf{0.5369}_{0.5328}$
		Static1	$0.6532_{0.4305}$	$0.7056_{0.4685}$	$0.6401_{0.4714}$	$0.6800_{0.4572}$	$0.6391_{0.4214}$	$\mathbf{0.5417}_{0.3613}$	$0.6841_{0.3559}$	$1.3416_{1.3356}$	$11.0473_{13.9739}$	$0.7730_{0.5636}$	$0.7018_{0.5807}$
		Stochastic	$0.8120_{0.5061}$	$0.7946_{0.4904}$	$0.7804_{0.4153}$	$0.7830_{0.5038}$	$0.7610_{0.3402}$	$0.7142_{0.4665}$	$0.976_{0.5703}$	$0.7249_{0.6312}$	$26.2156_{117.6624}$	$0.7861_{0.5641}$	$\mathbf{0.6711}_{0.4551}$
		Dynamic	$0.5776_{0.6066}$	$0.5801_{0.6092}$	$0.5981_{0.5365}$	$0.6501_{0.6210}$	$0.6001_{0.7037}$	$0.7048_{0.7051}$	$0.7342_{0.672}$	$0.8196_{0.6017}$	$1.7650_{2.5869}$	$0.6913_{0.5283}$	$\mathbf{0.5628}_{0.4488}$

Table 2: Stochastic Policy Evaluation. The proportion of treated units ranges from 0.20 to 0.80, reflecting real-world policy settings where the treatment size may fluctuate according to budget constraints. Best performance in each row is **bolded**.

Dataset	Proportion	CFR	CFR+z	ND	ND+z	TARNet	TARNet+z	NetEst	HINITE	TNet	Mvdr
BlogCata	0.20	$0.3197_{0.305}$	$0.3562_{0.3169}$	$0.5202_{0.4264}$	$0.5617_{0.4721}$	$0.4834_{0.364}$	$0.4491_{0.3787}$	$0.4672_{0.4064}$	$0.6987_{0.2483}$	$57.2356_{57.4217}$	$\mathbf{0.3016}_{0.2646}$
	0.50	$0.2598_{0.2665}$	$0.2552_{0.261}$	$0.3095_{0.2291}$	$0.3846_{0.2485}$	$0.3743_{0.3218}$	$0.2980_{0.3026}$	$0.313_{0.4055}$	$0.5063_{0.2674}$	$8.5232_{11.3317}$	$\mathbf{0.2948}_{0.2602}$
	0.65	$0.4678_{0.3421}$	$0.494_{0.3619}$	$0.4957_{0.2699}$	$0.6111_{0.3331}$	$0.4117_{0.2646}$	$0.4747_{0.2324}$	$0.5033_{0.3818}$	$1.0180_{0.5412}$	$7.3843_{6.2696}$	$\mathbf{0.3990}_{0.3953}$
	0.80	$0.2743_{0.175}$	$0.2728_{0.1868}$	$0.2859_{0.2169}$	$0.3844_{0.2375}$	$0.2581_{0.1462}$	$0.3422_{0.1296}$	$0.4369_{0.3256}$	$0.9691_{0.6849}$	$3033.8552_{2888.6561}$	$\mathbf{0.2300}_{0.1513}$
Flickr	0.20	$0.0956_{0.0763}$	$0.1181_{0.0968}$	$0.1745_{0.0364}$	$0.1864_{0.0732}$	$0.0992_{0.0541}$	$0.1294_{0.055}$	$0.1724_{0.0824}$	$0.4369_{0.3551}$	$320.0_{634.0}$	$\mathbf{0.0872}_{0.0694}$
	0.50	$0.1150_{0.0419}$	$0.1161_{0.0507}$	$0.1455_{0.0486}$	$0.1277_{0.0537}$	$0.1414_{0.0649}$	$0.1293_{0.0574}$	$0.2080_{0.1165}$	$0.1928_{0.2158}$	$320.5_{545.6}$	$\mathbf{0.0914}_{0.0688}$
	0.65	$0.1028_{0.0692}$	$0.0966_{0.0706}$	$0.1118_{0.0111}$	$0.1002_{0.0527}$	$0.1071_{0.0273}$	$0.1036_{0.026}$	$0.2436_{0.0739}$	$0.4003_{0.3475}$	$333.2976_{295.757}$	$\mathbf{0.0908}_{0.0614}$
	0.80	$0.1196_{0.0886}$	$0.0933_{0.0495}$	$0.1781_{0.0703}$	$0.137_{0.0744}$	$0.2265_{0.110}$	$0.2008_{0.0558}$	$0.5136_{0.1013}$	$0.5119_{0.3149}$	$1073.9346_{1198.002}$	$\mathbf{0.0850}_{0.0700}$
DBLP	0.20	$0.6884_{0.6434}$	$0.4900_{0.6684}$	$0.5301_{0.6866}$	$0.4749_{0.7014}$	$0.5244_{0.6407}$	$0.5582_{0.7820}$	$0.6222_{0.7431}$	$0.8006_{0.7482}$	$121.7685_{240.3945}$	$\mathbf{0.4284}_{0.5858}$
	0.50	$0.4794_{0.3694}$	$0.4749_{0.3688}$	$0.7325_{0.5254}$	$0.5381_{0.4129}$	$0.5989_{0.4951}$	$0.5228_{0.3430}$	$0.5057_{0.3798}$	$0.8234_{0.6519}$	$3.0800_{4.7400}$	$\mathbf{0.4363}_{0.3654}$
	0.65	$0.5841_{0.5569}$	$0.6119_{0.5669}$	$0.5515_{0.2723}$	$0.5838_{0.6110}$	$0.5397_{0.2044}$	$0.5593_{0.4905}$	$0.5499_{0.4583}$	$1.4967_{1.0754}$	$17.701_{32.4764}$	$\mathbf{0.5140}_{0.4782}$
	0.80	$0.6107_{0.6391}$	$0.6028_{0.6532}$	$0.7433_{0.2997}$	$0.5800_{0.6448}$	$0.7180_{0.2523}$	$0.6690_{0.5468}$	$0.6785_{0.3982}$	$1.7752_{1.2415}$	$9.3042_{7.7427}$	$\mathbf{0.4026}_{0.5439}$

6.3 Results and Ablation Studies

Comparisons with Baselines. Results in Table 1 show that our Mvdr method outperforms baselines in most cases, especially for the stochastic and dynamic policy settings, where the treatment assignment is related to units' covariates and neighborhood exposure. Under static interventions, strong i.i.d. baselines can be competitive or best, suggesting that static policies are less sensitive to misspecification of the exposure mechanism. We also report the results of some other metrics in Appendix D.

The Perturbation Estimator. As shown in Table 1, we have conducted experiments to compare the performance between Mvdr and Mvdr(w/o. $\mathcal{L}_3$) across the three datasets. Overall, Mvdr outperforms Mvdr(w/o. $\mathcal{L}_3$), which is as expected. Results show that the TMLE framework introduces double robustness to Mvdr and mitigates bias through the three modules. When the initial outcome model includes bias, the conditional density works as a re-biasing step to produce robust results.

The Stochastic Policy. We evaluate the performance when the treated proportion of stochastic assignment policy changes from 0.20 to 0.80. As shown in Table 2, Mvdr outperforms baselines across different assignment policy settings.

Unbiasness under Misspecification. To show that our model can perform stably when either the working model m and h is misspecified, we simulate data accordingly to mimic the different misspecification problem. (1) To model the misspecification of treatment mechanism, we include data that randomly flip treatment with rate from 0.25 to 1. (2) For the misspecification of outcome mechanism, we include higher nonlinearity by

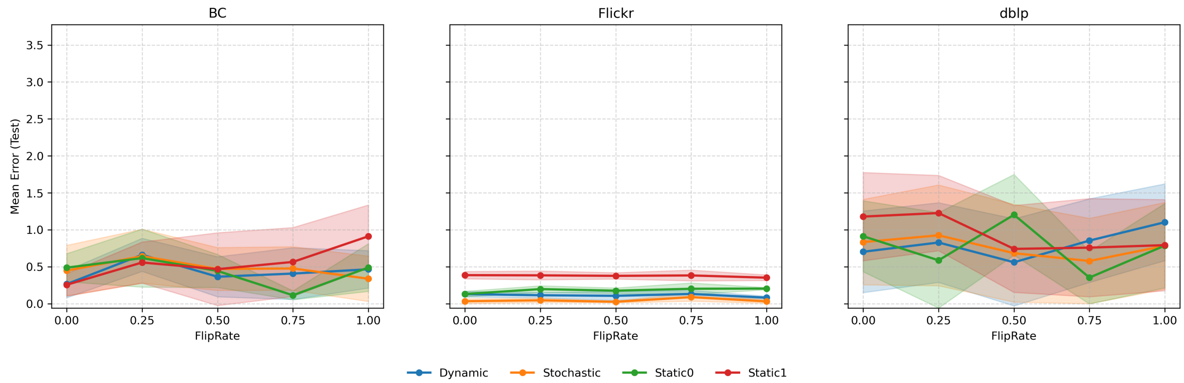

Figure 3: Evaluation on bias caused by misspecification on treatment mechanism.

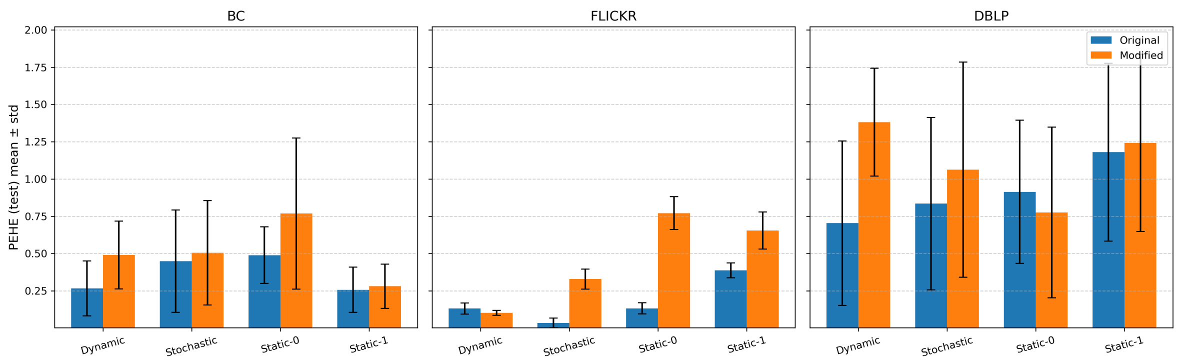

Figure 4: Evaluation on bias caused by misspecification on outcome model.

introducing the tanh function toward covariates to outcome function. This will lead to the exposure summary being no longer sufficient for outcome prediction but still following the treatment generation mechanism. Figure 3 and Figure 4 show that the prediction with misspecification is close to the original results.

7 Discussions and Conclusions

This paper proposes a framework to estimate the average expected outcome on an observed heterogeneous network with latent network dependency through a targeted maximum likelihood estimation method. We provided theorems showing the consistency and double robustness. Experiments show that our method is able to perform well on estimating average expected outcome under hypothetical intervention policies on networks. Our method aims to solve the problem of latent network dependency in causal inference under heterogeneous network setting, by modeling potential outcome as conditional expectation of two network related variable, W and V. Though it achieves relatively good performance, we acknowledge that this method has its limitations. For example, the exposure mapping, where we limited its dimension to 1, may not be sufficient to cover the interference mechanism in complex network settings. We plan to explore other semi-parametric methods to model the exposure mapping and achieve better identification of interference in our future work.

References

Shishir Adhikari and Elena Zheleva. Inferring individual direct causal effects under heterogeneous peer influence. *Machine Learning*, 114(4), March 2025. ISSN 1573-0565. doi: 10.1007/s10994-024-06729-2. URL `http://dx.doi.org/10.1007/s10994-024-06729-2`.

Peter M Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. 2017.

Laura B Balzer, Wenjing Zheng, Mark J van der Laan, and Maya L Petersen. A new approach to hierarchical data analysis: Targeted maximum likelihood estimation for the causal effect of a cluster-level exposure. *Statistical methods in medical research*, 28(6):1761–1780, 2019.

Ariel Chao, Donna Spiegelman, Ashley Buchanan, and Laura Forastiere. Estimation and inference for causal spillover effects in egocentric-network randomized trials in the presence of network membership misclassification. *Biostatistics*, 26(1):kxaf009, 2025.

David Chen, Maya L Petersen, Helene Charlotte Rytgaard, Randi Grøn, Theis Lange, Søren Rasmussen, Richard E Pratley, Steven P Marso, Kajsa Kvist, John Buse, et al. Beyond the cox hazard ratio: a targeted learning approach to survival analysis in a cardiovascular outcome trial application. *Statistics in Biopharmaceutical Research*, 15(3):524–539, 2023.

Weilin Chen, Ruichu Cai, Zeqin Yang, Jie Qiao, Yuguang Yan, Zijian Li, and Zhifeng Hao. Doubly robust causal effect estimation under networked interference via targeted learning, 2024. URL `https://arxiv.org/abs/2405.03342`.

Iván Díaz Muñoz and Mark J. Van Der Laan. Super Learner Based Conditional Density Estimation with Application to Marginal Structural Models. *The International Journal of Biostatistics*, 7(1):1–20, January 2011. ISSN 1557-4679. doi: 10.2202/1557-4679.1356. URL `https://www.degruyter.com/document/doi/10.2202/1557-4679.1356/html`.

Naoki Egami. Spillover effects in the presence of unobserved networks. *Political Analysis*, 29(3):287–316, 2021.

Laura Forastiere, Fabrizia Mealli, Albert Wu, and Edoardo M Airoldi. Estimating causal effects under network interference with bayesian generalized propensity scores. *Journal of Machine Learning Research*, 23(289):1–61, 2022.

Ruocheng Guo, Jundong Li, and Huan Liu. Learning Individual Causal Effects from Networked Observational Data, December 2019. URL `http://arxiv.org/abs/1906.03485`. arXiv:1906.03485 [cs].

M Elizabeth Halloran and Claudio J Struchiner. Causal inference in infectious diseases. *Epidemiology*, pp. 142–151, 1995.

Ruining He and Julian McAuley. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pp. 507–517, Republic and Canton of Geneva, CHE, 2016. International World Wide Web Conferences Steering Committee. ISBN 9781450341431. doi: 10.1145/2872427.2883037. URL `https://doi.org/10.1145/2872427.2883037`.

Tadao Hoshino, , and Takahide Yanagi. Causal Inference with Noncompliance and Unknown Interference. *Journal of the American Statistical Association*, 119(548):2869–2880, October 2024. ISSN 0162-1459. doi: 10.1080/01621459.2023.2284413. URL `https://doi.org/10.1080/01621459.2023.2284413`. Publisher: ASA Website _eprint: https://doi.org/10.1080/01621459.2023.2284413.

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

Song Jiang and Yizhou Sun. Estimating Causal Effects on Networked Observational Data via Representation Learning. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 852–861, Atlanta GA USA, October 2022. ACM. ISBN 978-1-4503-9236-5. doi: 10.1145/3511808.3557311. URL `https://dl.acm.org/doi/10.1145/3511808.3557311`.

Fredrik D. Johansson, Uri Shalit, Nathan Kallus, and David Sontag. Generalization bounds and representation learning for estimation of potential outcomes and causal effects, 2023. URL `https://arxiv.org/abs/2001.07426`.

Vishesh Karwa and Edoardo M Airoldi. A systematic investigation of classical causal inference strategies under mis-specification due to network interference. *arXiv preprint arXiv:1810.08259*, 2018.

Edward H. Kennedy. Semiparametric theory and empirical processes in causal inference, 2016. URL `https://arxiv.org/abs/1510.04740`.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017. URL `https://arxiv.org/abs/1609.02907`.

Noémi Kreif, Linh Tran, Richard Grieve, Bianca De Stavola, Robert C Tasker, and Maya Petersen. Estimating the comparative effectiveness of feeding interventions in the pediatric intensive care unit: a demonstration of longitudinal targeted maximum likelihood estimation. *American journal of epidemiology*, 186 (12):1370–1379, 2017.

Michael P Leung. Causal inference under approximate neighborhood interference. *Econometrica*, 90(1): 267–293, 2022.

Shuangning Li and Stefan Wager. Random graph asymptotics for treatment effect estimation under network interference. *The Annals of Statistics*, 50(4):2334–2358, 2022.

Xiaofeng Lin, Guoxi Zhang, Xiaotian Lu, Han Bao, Koh Takeuchi, and Hisashi Kashima. Estimating Treatment Effects Under Heterogeneous Interference, September 2023. URL `http://arxiv.org/abs/2309.13884`. arXiv:2309.13884 [cs].

Jing Ma, Ruocheng Guo, Chen Chen, Aidong Zhang, and Jundong Li. Deconfounding with Networked Observational Data in a Dynamic Environment. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pp. 166–174, Virtual Event Israel, March 2021. ACM. ISBN 978-1-4503-8297-7. doi: 10.1145/3437963.3441818. URL `https://dl.acm.org/doi/10.1145/3437963.3441818`.

Jing Ma, Mengting Wan, Longqi Yang, Jundong Li, Brent Hecht, and Jaime Teevan. Learning Causal Effects on Hypergraphs, July 2022. URL `http://arxiv.org/abs/2207.04049`. arXiv:2207.04049 [cs].

Yunpu Ma and Volker Tresp. Causal Inference under Networked Interference and Intervention Policy Enhancement, May 2021. URL `http://arxiv.org/abs/2002.08506`. arXiv:2002.08506 [stat].

Charles F Manski. Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23, 2013.

Lizhen Nie, Mao Ye, Qiang Liu, and Dan Nicolae. VCNET AND FUNCTIONAL TARGETED REGULARIZA- TION FOR LEARNING CAUSAL EFFECTS OF CONTINU- OUS TREATMENTS. 2021.

Elizabeth L Ogburn, Ilya Shpitser, and Youjin Lee. Causal inference, social networks and chain graphs. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 183(4):1659–1676, 2020.

Elizabeth L. Ogburn, Oleg Sofrygin, Ivan Diaz, and Mark J. van der Laan. Causal inference for social network data, 2022. URL `https://arxiv.org/abs/1705.08527`.

Judea Pearl. The causal foundations of structural equation modeling. *Handbook of structural equation modeling*, pp. 68–91, 2012.

Fredrik Sävje. Causal inference with misspecified exposure mappings: separating definitions and assumptions. *Biometrika*, 111(1):1–15, 2024.

Uri Shalit, Fredrik D. Johansson, and David Sontag. Estimating individual treatment effect: generalization bounds and algorithms, 2017. URL `https://arxiv.org/abs/1606.03976`.

Cosma Rohilla Shalizi and Andrew C. Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, 40(2):211–239, May 2011. ISSN 1552-8294. doi: 10.1177/0049124111404820. URL `http://dx.doi.org/10.1177/0049124111404820`.

Claudia Shi, David Blei, and Victor Veitch. Adapting neural networks for the estimation of treatment effects. *Advances in neural information processing systems*, 32, 2019.

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.

Eric J. Tchetgen Tchetgen, Isabel Fulcher, and Ilya Shpitser. Auto-g-computation of causal effects on a network, 2019. URL `https://arxiv.org/abs/1709.01577`.

Panos Toulis and Edward Kao. Estimation of causal peer influence effects. In Sanjoy Dasgupta and David McAllester (eds.), *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pp. 1489–1497, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR. URL `https://proceedings.mlr.press/v28/toulis13.html`.

Mark J Van der Laan. Causal inference for a population of causally connected units. *Journal of Causal Inference*, 2(1):13–74, 2014.

Mark J. van der Laan. Causal Inference for a Population of Causally Connected Units. *Journal of causal inference*, 2(1):13–74, March 2014. ISSN 2193-3677. doi: 10.1515/jci-2013-0002. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4500386/`.

Mark J van der Laan and Susan Gruber. Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The international journal of biostatistics*, 8(1), 2012.

Mark J Van Der Laan and Daniel Rubin. Targeted maximum likelihood learning. *The international journal of biostatistics*, 2(1), 2006.

Mark J Van der Laan, Sherri Rose, et al. *Targeted learning: causal inference for observational and experimental data*, volume 4. Springer, 2011.

Richard A Watson, Hengrui Cai, Xinming An, Samuel McLean, and Rui Song. On heterogeneous treatment effects in heterogeneous causal graphs, 2023. URL `https://arxiv.org/abs/2301.12383`.

Kun Zhang, Biwei Huang, Jiji Zhang, Clark Glymour, and Bernhard Schölkopf. Causal discovery from nonstationary/heterogeneous data: Skeleton estimation and orientation determination. In *IJCAI: Proceedings of the Conference*, volume 2017, pp. 1347, 2017.

A Notations

Symbol	Type / domain	Description
Units, data, and network		
$i \in \{1, \dots, n\}$	index	Unit (node) index; n is network size.
x_i	$\mathbb{R}^d$	Covariates of unit i (column vector).
t_i	$\{0, 1\}$	Binary treatment of unit i.
y_i	$\mathbb{R}$	Observed outcome for unit i under t_i.
X	$\mathbb{R}^{n \times d}$	Covariate matrix $X = (x_1, \dots, x_n)^\top$.
T	$\{0, 1\}^n$	Treatment vector $T = (t_1, \dots, t_n)^\top$.
Y	$\mathbb{R}^n$	Outcome vector $Y = (y_1, \dots, y_n)^\top$.
G	graph on n nodes	Observed (aggregated) network.
A	$\{0, 1\}^{n \times n}$	Adjacency matrix of G (symmetrized). $A_{ij} = 1$ iff edge (i, j) exists.
$A^{(m)}$	$\{0, 1\}^{n \times n}$	Adjacency of the m-th view/layer, $m = 1, \dots, K$.
$N_i^{(m)}$	subset of $\{1, \dots, n\}$	Neighbor set of node i in view m.
$d_i^{(m)}$	$\mathbb{N}$	Degree of node i in view m (in $A^{(m)}$).
Structural equation model (SEM)		
f_X, f_T, f_Y	functions	Unknown structural functions in the SEM.
$\varepsilon_{X_i}, \varepsilon_{T_i}, \varepsilon_{Y_i}$	r.v.	Exogenous errors for unit i (may be dependent across units).
$s_{X,i}$	summary map	Summary of $\{X_j : A_{ij} = 1\}$ for unit i.
$s_{T,i}$	summary map	Summary of $\{T_j : A_{ij} = 1\}$ for unit i.
W_i	r.v.	Covariate summary $W_i = s_{X,i}(\{X_j : A_{ij} = 1\})$.
S_i	r.v.	Neighbor exposure summary of treatments (continuous or discrete).
V_i	r.v.	Joint exposure summary, typically $V_i = (T_i, S_i)$.
Statistical models and conditionals		
$m(v, w)$	$\mathbb{R}$	Outcome regression: $m(v, w) = \mathbb{E}[Y \mid V = v, W = w]$.
$h(v \mid w)$	density / pmf	Exposure mechanism: $h(v \mid w) = \mathbb{P}(V = v \mid W = w)$.
$h(v, w)$	joint	Joint law of (V, W); $h(v, w) = \mathbb{P}(V = v, W = w)$.
$h^*(\cdot \mid w)$	intervention kernel	Target (stochastic) intervention on V given $W = w$.
p_W	law on $\mathcal{W}$	Marginal distribution of W.
O	tuple	Observed variables $O = (W, V, Y)$.
$\bar{h}(\cdot)$	average	Node-wise average: e.g. $\bar{h}(v, w) = \frac{1}{n} \sum_{j=1}^n h_j(v, w)$.
$\bar{h}^*(\cdot)$	average under h^*	$\bar{h}^*(v, w) = \frac{1}{n} \sum_{j=1}^n h_j^*(v, w)$.
Targets / estimands		
T^*	$\{0, 1\}^n$	Hypothetical intervention assignment over the network.
V^*	r.v.	Exposure under intervention, e.g. $V^* \sim h^*(\cdot \mid W)$.
v_0	value in $\mathsf{supp}(V)$	Fixed (static) neighbor exposure level.
ψ	$\mathbb{R}$	Target parameter (e.g., $\mathbb{E}[m(v_0, W)]$ or $\mathbb{E}_W \int m(v, W) h^*(dv \mid W)$).
H_i	weight	"Clever covariate": $\frac{\bar{h}^*(V_i \mid W_i)}{\bar{h}(V_i \mid W_i)}$ (stochastic) or $\frac{\mathbf{1}\{V_i = v_0\}}{\bar{h}(v_0 \mid W_i)}$ (static).
$D^*(O)$	r.v.	Efficient influence function (EIF) for ψ.
Asymptotics / dependence		
$K_{\max,n}$	$\mathbb{N}$	Maximum degree of (aggregated) network.
C_n	$\mathbb{N}$	Effective sample size with $n/K_{\max,n}^2 \leq C_n \leq n$.

B Data Statics

Dataset	Graph	Nodes	Edges	AvgDeg	StdDeg	MinDeg	MaxDeg	Density	Components
BC	A1	5196	171743	66	54.83	5	769	0.012725	1
	A2	5196	32036	12	1.77	10	20	0.002374	1
	A3	5196	34957	13	2.12	10	23	0.002590	1
Flickr	A1	7301	238086	65	132.66	2	1834	0.008934	1
	A2	7301	45727	12	2.07	10	30	0.001716	1
	A3	7301	48203	13	2.18	10	32	0.001809	1
DBLP	A1	2828	3988527	2820	91.90	2	2824	0.997786	2
	A2	2828	17868	12	2.24	10	23	0.004470	7
	A3	2828	17615	12	1.95	10	22	0.004407	1

C Theory

Lemma C.1 (Identification invariance under aggregation)**.** *Let (W, V) be any Borel-measurable aggregations of $(X, T, A^{(1)}, \dots, A^{(M)})$ with W depending only on pre-treatment inputs. Under Assumptions 1–3, both targets $\psi_{\text{stat}}, \psi_{\text{stoch}}$ are identified from $P(W, V, Y)$ via*

$$\psi_{\text{stat}} = \int m(v_0, w)\, p_W(dw), \qquad \psi_{\text{stoch}} = \iint m(v, w)\, h^*(dv \mid w)\, p_W(dw).$$

Proof. Multiple view-specific networks $\{A^{(m)} : m = 1, \dots, M\}$ are constructed from baseline covariates. For each unit i, we form view-specific summaries

$$W_i^{(m)} = s_C^{(m)}(\{C_j : A_{ij}^{(m)} = 1\}), \quad V_i^{(m)} = s_X^{(m)}(\{X_j : A_{ij}^{(m)} = 1\}),$$

and aggregate them into

$$W_i = \text{Agg}_W(W_i^{(1)}, \dots, W_i^{(M)}), \quad V_i = \text{Agg}_V(V_i^{(1)}, \dots, V_i^{(M)}).$$

Importantly, both constructions produce (V_i, W_i) that are functions of baseline covariates C and treatment X of neighbors. As such, the statistical model remains nonparametric in (V, W), and the target parameter is defined entirely through $m(v, w) = E[Y \mid V = v, W = w]$ and $h(v \mid w) = \Pr(V = v \mid W = w)$. Therefore the pathwise derivative and canonical gradient retain the same form given in Equations equation 5–equation 6. Only the dimension and support of (V, W) change under different aggregation rules; the EIF expressions remain unchanged.

□

Condition 1 (Bounded dependence / degree growth)**.** *The maximum degree of the (aggregated) network grows slowly with sample size:*

$$K_{\max,n}^2 / n \to 0.$$

This ensures that each unit's dependence neighborhood does not become too large relative to the overall sample, permitting a central limit theorem under Stein's method (Ogburn et al., 2022).

Condition 2. *Either the outcome regression $m(v, w)$ or the exposure mechanism $h(v \mid w)$ is consistently estimated at a sufficiently fast rate, or their estimation errors satisfy*

$$\|\hat{m} - m\| \cdot \|\hat{h} - h\| = o_p(n^{-1/2}).$$

This guarantees that the second-order remainder in the influence function expansion vanishes.

Proposition C.1 (EIF form is invariant; only $m, \bar{h}$ change)**.** *Under Assumptions 1–3 and positivity, the efficient influence functions are*

$$D_{\text{stat}}^*(O) = \frac{\mathbb{1}\{V = v\}}{\bar{h}(v \mid W)}\{Y - m(v, W)\} + \big(m(v, W) - \psi_{\text{stat}}\big),$$

$$D_{\text{stoch}}^*(O) = \frac{\bar{h}^*(V \mid W)}{\bar{h}(V \mid W)}\{Y - m(V, W)\} + \Big(\int m(v, W) h^*(dv \mid W) - \psi_{\text{stoch}}\Big),$$

with expectations 0 and finite variances. The expressions are identical to the single-view case, with $m, \bar{h}$ computed for the aggregated (W, V).

Lemma C.2 (Dependence and effective sample size under aggregation)**.** *Let $G^{(m)}$ have maximum degree $K^{(m)}_{\max,n}$. Define the aggregated support-graph $\tilde{G}$ by connecting $i \sim j$ if they are connected in at least one view. Then*

$$K_{\max,n}(\tilde{G}) \;\le\; \sum_{m=1}^{M} K^{(m)}_{\max,n}.$$

Under Assumption 4 (two-hop local dependence per view), the dependency graph for $\{D^*(O_i)\}_{i=1}^n$ on $\tilde{G}$ has maximum degree

$$\Delta_n \;\lesssim\; K_{\max,n}(\tilde{G})^2 \;\le\; \Big(\sum_{m=1}^{M} K^{(m)}_{\max,n}\Big)^2.$$

By Lemma C.1 and Proposition C.1, the multi–view reparametrization preserves both the target functional and the EIF, so the EIF coincides with the forms in Ogburn et al. (2020). Hence, under the stated regularity (positivity, bounded moments) and $K^2_{\max,n}/n \to 0$, the CAN result can be obtained.

Proposition C.2 (CLT with multi-view aggregation)**.** *Let $Z_i := D^*(O_i)$ with $\mathbb{E}|Z_i|^3 < \infty$. Assume $K_{\max,n}(\tilde{G})^2/n \to 0$. Then, with $C_n := n/(1+\Delta_n)$ and $\Delta_n \lesssim K_{\max,n}(\tilde{G})^2$,*

$$\sqrt{C_n}\,(\mathbb{P}_n - P_0)D^*(O) \;\Rightarrow\; \mathcal{N}(0,\sigma^2), \qquad \sigma^2 = \mathrm{Var}\{D^*(O)\}.$$

Theorem C.1 (CAN with multi-view aggregation)**.** *Under Assumptions 1–5, Lemma C.1, Proposition C.1, the CLT of Proposition C.2, and the nuisance-rate condition $\|\hat{m} - m\|\,\|\hat{\bar{h}} - \bar{h}\| = o_p(C_n^{-1/2})$, the TMLE $\hat{\psi}_n$ for either target satisfies*

$$\sqrt{C_n}\,(\hat{\psi}_n - \psi) \;\Rightarrow\; \mathcal{N}(0,\sigma^2), \qquad C_n = \frac{n}{1+\Delta_n}, \quad \Delta_n \lesssim K_{\max,n}(\tilde{G})^2.$$

D Additional Experimental Results

The main purpose of this paper is to evaluate the average potential outcome $E[\bar{Y}_n^*]$, which means the outcome for an observed network with latent dependency if certain hypothetical intervention is assigned to it. In Table 3, we also report the results for average main effect, average spillover effect, average total effect and individual main effect, individual spillover effect, individual total effect by pre-defining the exposure mapping v as the average ratio of treated neighbors across views.
Define ψ_i as $\psi_i(t,z) = m_t(z,w_i) + \epsilon(z,w_i)\cdot H_i$, and ψ as $\psi(t,z) = \frac{1}{n}\sum_{i=1}^n \psi_i(t,z)$. Then we have the following definitions:

- The average main effect (AME) measures the difference in the mean outcomes between units assigned to $T=t, Z=0$ and assigned to $T=t', z=0$: $\tau^{(t,0),(t',0)} = \psi(t,0) - \psi(t',0)$.
- The average spillover effect (ASE) measures the difference in mean outcomes between units assigned to $T=0, Z=z$ and assigned to $T=0, Z=z'$: $\tau^{(0,z),(0,z')} = \psi(0,z) - \psi(0,z')$.
- The average total effect (ATE) measures the difference in mean outcomes between units assigned to $T=t, Z=z$ and assigned $T=t', Z=z'$: $\tau^{(t,z),(t',z')} = \psi(t,z) - \psi(t',z')$.
- The individual main effect (IME) measures the difference in mean outcomes of a particular unit x_i assigned to $T=t, Z=0$ and assigned $T=t', Z=0$: $\tau_i^{(t,0),(t',0)} = \psi_i(t,0) - \psi_i(t',0)$.
- The individual spillover effect (ISE) measures the difference in mean outcomes of a particular unit x_i assigned to $T=0, Z=z$ and assigned $T=0, Z=z'$: $\tau_i^{(t,z),(t',z')} = \psi_i(0,z) - \psi_i(0,z')$.
- The individual total effect (ITE) measures the difference in mean outcomes of a particular unit x_i assigned to $T=t, Z=z$ and assigned $T=t', Z=z'$: $\tau_i^{(t,z),(t',z')} = \psi_i(t,z) - \psi_i(t',z')$.

For the average effects, we use mean absolute error (MAE) as metric: $\epsilon_{average} = |\hat{\tau} - \tau|$. For the individual effects, we use Precision in Estimation of Heterogeneous Effect (PEHE) as metric: $\epsilon_{individual} = \frac{1}{n}\sum_{i=1}^n (\tau_i - \hat{\tau}_i)^2$. Specifically, HINITE does not indentify spillover effects and main effects, so the results are not reported.

Table 3: Comparison of estimation errors across different models and datasets. Best performance in each row is **bolded**. ′−′ represents the results are not reported by that method.

Split	**Metric**	**dataset**	**effect**	CFR	CFR+z	ND	ND+z	TARNet	TARNet+z	NetEst	HINITE	TNet	MVDR
With-in-Sample	$\varepsilon_{average}$	BlogCata	AME	$\mathbf{0.1425}_{0.1516}$	$0.2549_{0.1329}$	$0.1590_{0.0859}$	$0.3080_{0.1620}$	$0.1550_{0.0683}$	$0.3335_{0.1443}$	$0.3533_{0.1410}$	–	$40.3076_{92.4736}$	$0.1720_{0.0995}$
			ASE	$0.1735_{0.1516}$	$0.2509_{0.2208}$	$0.1735_{0.1516}$	$0.2430_{0.2278}$	$\mathbf{0.1734}_{0.0696}$	$0.2241_{0.2090}$	$0.2191_{0.2231}$	–	$145.1494_{214.8604}$	$0.1905_{0.1016}$
			ATE	$0.2449_{0.1704}$	$0.1650_{0.0786}$	$0.1326_{0.1208}$	$\mathbf{0.1105}_{0.1060}$	$0.1326_{0.1076}$	$0.1308_{0.1385}$	$0.5076_{0.1546}$	$0.6230_{0.4354}$	$5707.6897_{9251.3347}$	$\mathbf{0.1639}_{0.1457}$
		Flickr	AME	$0.2691_{0.1097}$	$0.1692_{0.1467}$	$0.2726_{0.2432}$	$0.4251_{0.1537}$	$0.2918_{0.2166}$	$0.3492_{0.0897}$	$0.7693_{0.0524}$	–	$12.0787_{11.3391}$	$\mathbf{0.0593}_{0.0470}$
			ASE	$\mathbf{0.0887}_{0.0211}$	$0.1767_{0.0620}$	$0.1078_{0.0822}$	$0.1964_{0.0883}$	$0.1078_{0.0822}$	$0.2736_{0.0700}$	$0.1948_{0.0679}$	–	$158.5197_{175.1475}$	$0.2296_{0.1129}$
			ATE	$0.2617_{0.1099}$	$0.1887_{0.1861}$	$0.3048_{0.1872}$	$0.3013_{0.1364}$	$0.2833_{0.2067}$	$0.3701_{0.1785}$	$1.0245_{0.1374}$	$\mathbf{0.1751}_{0.1004}$	$2923.5093_{3001.1783}$	$0.4908_{0.2424}$
		DBLP	AME	$0.2153_{0.1557}$	$0.2606_{0.1441}$	$0.5965_{0.2617}$	$0.2916_{0.1325}$	$1.0216_{0.6321}$	$0.3672_{0.2286}$	$0.8758_{0.2556}$	–	$3.5683_{11.7916}$	$\mathbf{0.0336}_{0.0348}$
			ASE	$0.2606_{0.1769}$	$0.329_{0.2398}$	$0.3108_{0.1768}$	$0.3311_{0.1755}$	$0.1381_{0.0438}$	$\mathbf{0.1376}_{0.0429}$	$0.2619_{0.1827}$	–	$7.5224_{11.0523}$	$\mathbf{0.2611}_{0.0314}$
			ATE	$\mathbf{0.076}_{0.0833}$	$0.1739_{0.0943}$	$0.8423_{0.2742}$	$0.4265_{0.0574}$	$1.1592_{0.5784}$	$0.2585_{0.1698}$	$1.0077_{0.0836}$	$0.6815_{0.3553}$	$11.982_{16.2407}$	$0.4408_{0.0527}$
	$\varepsilon_{individual}$	BlogCata	IME	$3.0597_{0.1232}$	$3.0358_{0.1232}$	$2.9224_{0.094}$	$2.9359_{0.0842}$	$2.9427_{0.0894}$	$2.9570_{0.0808}$	$2.9412_{0.0865}$	–	$140.5076_{127.1403}$	$\mathbf{2.8966}_{0.0583}$
			ISE	$2.8973_{0.0845}$	$2.9072_{0.0845}$	$2.8973_{0.0852}$	$2.9069_{0.0842}$	$2.8973_{0.0826}$	$2.9044_{0.0836}$	$2.9088_{0.0800}$	–	$270.6221_{341.4499}$	$\mathbf{2.7668}_{0.0404}$
			ITE	$3.0540_{0.097}$	$3.0182_{0.0970}$	$2.9004_{0.0510}$	$2.8967_{0.0496}$	$2.9216_{0.0529}$	$2.9185_{0.0591}$	$2.9353_{0.0535}$	$3.3113_{0.2458}$	$12038.2843_{15704.8492}$	$\mathbf{2.8692}_{0.0482}$
		Flickr	IME	$9.3637_{5.1163}$	$2.9457_{0.0596}$	$6.0954_{1.8223}$	$2.9607_{0.0690}$	$2.9683_{0.0648}$	$2.9124_{0.0499}$	$2.9537_{0.0498}$	–	$35.5421_{32.8715}$	$\mathbf{0.6581}_{0.2529}$
			ISE	$2.8980_{0.0439}$	$2.8965_{0.0976}$	$3.0014_{0.0237}$	$3.0342_{0.0964}$	$2.9228_{0.0547}$	$2.8865_{0.0989}$	$2.8896_{0.1010}$	–	$440.6491_{463.7988}$	$\mathbf{0.6886}_{0.2660}$
			ITE	$9.3395_{5.1269}$	$3.0072_{0.641}$	$6.1200_{1.8291}$	$2.9525_{0.0830}$	$2.9506_{0.0622}$	$2.9470_{0.0449}$	$3.0514_{0.0463}$	$3.0427_{0.0662}$	$9314.1054_{9633.2401}$	$0.8652_{0.2978}$
		DBLP	IME	$3.3294_{0.2079}$	$3.3316_{0.2094}$	$3.0448_{0.0829}$	$2.6218_{0.0341}$	$2.8543_{0.2530}$	$2.6366_{0.0541}$	$3.4669_{0.2156}$	–	$6.3263_{11.4395}$	$\mathbf{0.0506}_{0.0353}$
			ISE	$3.4736_{0.1455}$	$3.483_{0.1517}$	$2.9793_{0.0928}$	$2.6923_{0.0325}$	$2.6921_{0.0321}$	$2.6921_{0.0322}$	$3.508_{0.1563}$	–	$12.3642_{21.6718}$	$\mathbf{0.2852}_{0.0557}$
			ITE	$3.3944_{0.0849}$	$3.3985_{0.0839}$	$3.0996_{0.0805}$	$2.5758_{0.0772}$	$3.2055_{0.1500}$	$2.5984_{0.1022}$	$3.5871_{0.0966}$	$2.9779_{0.2013}$	$13.5564_{15.9518}$	$\mathbf{0.4881}_{0.0460}$
Out-of-Sample	ε_{ATE}	BlogCata	AME	$0.2103_{0.154}$	$0.203_{0.1861}$	$0.5117_{0.2551}$	$0.6367_{0.3633}$	$0.4508_{0.3217}$	$0.5380_{0.4065}$	$0.2312_{0.2088}$	–	$40.3076_{92.4736}$	$\mathbf{0.1289}_{0.0416}$
			ASE	$0.1277_{0.0965}$	$0.1882_{0.1349}$	$0.1277_{0.0965}$	$0.1825_{0.1322}$	$0.1277_{0.0965}$	$0.1736_{0.1373}$	$\mathbf{0.1209}_{0.098}$	–	$145.1494_{214.8604}$	$0.1678_{0.0751}$
			ATE	$0.2449_{0.1704}$	$0.2726_{0.2062}$	$0.5484_{0.3592}$	$0.7826_{0.4412}$	$0.4874_{0.2824}$	$0.6466_{0.3515}$	$0.3724_{0.2632}$	$0.3574_{0.2483}$	$5707.6897_{9251.3347}$	$\mathbf{0.1639}_{0.1055}$
		Flickr	AME	$0.2691_{0.1097}$	$0.2908_{0.2120}$	$0.1662_{0.1123}$	$0.4614_{0.2634}$	$\mathbf{0.1626}_{0.1219}$	$0.4107_{0.2165}$	$0.7582_{0.1641}$	–	$9.0943_{10.9997}$	$0.2605_{0.0590}$
			ASE	$0.0887_{0.0211}$	$0.2074_{0.0858}$	$\mathbf{0.0887}_{0.0211}$	$0.2475_{0.1138}$	$0.0887_{0.0211}$	$0.2106_{0.1111}$	$0.1406_{0.1115}$	–	$92.6766_{87.5788}$	$0.2426_{0.0766}$
			ATE	$0.2617_{0.1099}$	$\mathbf{0.1940}_{0.1716}$	$0.2669_{0.1859}$	$0.2878_{0.1566}$	$0.3395_{0.1229}$	$0.2618_{0.1089}$	$0.7807_{0.1527}$	$0.1674_{0.1618}$	$2152.0205_{1924.3221}$	$0.5516_{0.3424}$
		DBLP	AME	$0.2153_{0.1557}$	$0.2606_{0.1441}$	$0.7047_{0.2382}$	$0.284_{0.1825}$	$0.8104_{0.4552}$	$0.4669_{0.1738}$	$0.8758_{0.2556}$		$3.5683_{11.7916}$	$\mathbf{0.0506}_{0.0353}$
			ASE	$\mathbf{0.2606}_{0.1769}$	$0.329_{0.2398}$	$0.2606_{0.1769}$	$0.2718_{0.1793}$	$0.2606_{0.1769}$	$0.2627_{0.176}$	$0.2619_{0.1827}$	–	$7.5224_{11.0523}$	$0.2852_{0.0557}$
			ATE	$\mathbf{0.076}_{0.0833}$	$0.1739_{0.0943}$	$0.8558_{0.0891}$	$0.3097_{0.1377}$	$0.9616_{0.325}$	$0.5847_{0.1165}$	$1.0077_{0.0836}$	$0.7259_{0.4971}$	$11.982_{16.2407}$	$\mathbf{0.4881}_{0.0460}$
	ε_{PEHE}	BlogCata	IME	$3.6311_{1.2841}$	$3.9781_{1.7998}$	$5.666_{2.2252}$	$7.2773_{3.4133}$	$4.7532_{2.3242}$	$5.5978_{4.2786}$	$3.1134_{0.1372}$	–	$140.5076_{127.1403}$	$\mathbf{2.9013}_{0.0460}$
			ISE	$2.9072_{0.0845}$	$2.8033_{0.0519}$	$2.7983_{0.0453}$	$2.8027_{0.0516}$	$2.7983_{0.0453}$	$2.8022_{0.0522}$	$3.0264_{0.124}$	–	$270.6221_{341.4499}$	$\mathbf{2.7966}_{0.0467}$
			ITE	$3.6261_{1.255}$	$3.9765_{1.7725}$	$5.6698_{2.2135}$	$7.2954_{3.3815}$	$4.7462_{2.3115}$	$5.5797_{4.2738}$	$3.0796_{0.1691}$	$3.2063_{0.3024}$	$12038.2843_{15704.8492}$	$\mathbf{2.8580}_{0.0460}$
		Flickr	IME	$9.3637_{5.1163}$	$3.573_{0.6824}$	$4.0064_{1.4430}$	$3.1551_{0.1926}$	$3.2343_{0.1757}$	$3.0985_{0.0172}$	$3.0916_{0.0363}$	–	$68.2053_{39.6463}$	$\mathbf{0.7660}_{0.2204}$
			ISE	$3.0457_{0.0475}$	$3.0295_{0.0976}$	$3.0457_{0.0475}$	$3.0566_{0.0409}$	$3.0457_{0.0475}$	$3.0314_{0.0964}$	$3.0338_{0.0290}$	–	$352.7498_{303.2289}$	$\mathbf{0.7214}_{0.2673}$
			ITE	$9.3395_{5.1269}$	$3.5895_{0.641}$	$4.0251_{1.4578}$	$3.1328_{0.1616}$	$3.242_{0.1979}$	$3.0533_{0.0709}$	$3.0834_{0.0812}$	$3.1153_{0.0732}$	$8165.584_{7729.401}$	$\mathbf{0.9543}_{0.3456}$
		DBLP	IME	$3.3294_{0.2079}$	$3.3316_{0.2094}$	$3.4027_{0.1876}$	$3.3357_{0.2167}$	$3.45_{0.1486}$	$3.3558_{0.2131}$	$3.4669_{0.2156}$	–	$6.3263_{11.4395}$	$\mathbf{0.5119}_{0.0155}$
			ISE	$3.4736_{0.1455}$	$3.483_{0.1517}$	$3.4736_{0.1455}$	$3.4745_{0.146}$	$3.4736_{0.1455}$	$3.4737_{0.1455}$	$3.508_{0.1563}$	–	$12.3642_{21.6718}$	$\mathbf{0.5175}_{0.0370}$
			ITE	$3.3944_{0.0849}$	$3.3985_{0.0839}$	$3.5013_{0.0936}$	$3.4116_{0.0815}$	$3.5403_{0.1502}$	$3.4462_{0.0883}$	$3.5871_{0.0966}$	$0.8227_{0.219}$	$13.5564_{15.9518}$	$\mathbf{0.7080}_{0.0249}$