

Article

Estimating Network Causal Effects with Misclassified Outcomes: Evidence from Karnataka

Yaqin Liao ^{1,*}  and Ming Lin ^{1,2}¹ Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen 361005, China; linming50@xmu.edu.cn² Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen 361005, China

* Correspondence: liaoyaqin@stu.xmu.edu.cn

Abstract

Misclassification of binary outcomes in network settings may bias the estimates of causal effects, including spillover effects that arise from social interactions, and may generate spurious causal effects. To address this issue, we develop a parametric framework that jointly estimates misclassification probabilities and causal effect parameters within a binary choice model with neighborhood exposure mappings. Monte Carlo simulations show that ignoring outcome misclassification or network-related variables leads to substantial bias, whereas the proposed method achieves a smaller bias and RMSE. By applying the method to microfinance and social network data from Karnataka, we find that under binary exposure, ignoring outcome misclassification yields statistically significant spillover and overall effects, whereas these effects become statistically insignificant once outcome misclassification is corrected for. Furthermore, omitting network-related variables overstates the direct effect. These results underscore the importance of jointly correcting for outcome misclassification and accounting for network-related variables to obtain credible causal inference.

Keywords: causal inference; network interference; misclassified outcomes; exposure mapping; binary choice models

MSC: 62D20

1. Introduction

Accumulating evidence suggests that social, spatial, and organizational linkages give rise to non-negligible inter-individual interactions. These linkages violate the Stable Unit Treatment Value Assumption and induce network spillovers, whereby one unit's treatment affects others' potential outcomes [1]. For example, Tchetgen Tchetgen et al. [2] studied the spillovers of prior incarceration for HIV, STIs, and hepatitis C using network ties based on recent sexual or injection partnerships. Hence, credible policy evaluation requires rigorous identification of such spillovers.

In the presence of spillovers, defining each individual's potential outcome as a function of the entire treatment vector leads to severe dimensionality and identifiability problems. Hence, the literature commonly assumes that an individual's potential outcome depends only on their own treatment and the treatments of others under a given interference structure. Two canonical frameworks arise: clustered interference, where spillovers are confined within pre-specified groups, as in cluster-randomized trials [3–5], and network neighborhood interference, where ties are encoded by an adjacency matrix and exposure mappings



Academic Editor: Anatoliy Swishchuk

Received: 10 March 2026

Revised: 5 April 2026

Accepted: 6 April 2026

Published: 8 April 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

(e.g., the number or share of treated neighbors) summarize neighborhood treatments into a scalar, enabling the definition and estimation of causal effects [6–11].

Building on these frameworks, a large body of research studied causal inference under interference by addressing a range of distinct challenges through various modeling approaches, including approximate neighborhood interference [10], heterogeneous spillover effects across network-driven subpopulations [12], and nonparametric estimation of stochastic policy effects under clustered interference [13]. Despite addressing different problems, this literature largely shares a common focus on identifying and estimating direct, spillover, and overall effects, typically assuming that binary outcomes are measured without error. However, this assumption may be unrealistic in many empirical contexts. Binary outcomes—such as microfinance loan take-up [12] and self-reported injection risk behaviors [14]—are often self-reported or constructed from administrative records and are therefore prone to misclassification. This concern is particularly relevant in our empirical application, which uses microfinance and social network data from [15], drawn from a sample of villages in Karnataka, India. The data combine a household census with a partial individual census, providing detailed information on household characteristics, economic behavior, and social interactions. As the data rely on self-reported borrowing and multiple sources, informal borrowing is likely subject to measurement error due to reporting errors or recall bias. Measurement errors of this kind are pervasive in applied research and have long been recognized as an important source of bias in econometric analysis. The literature on causal inference without interference likewise emphasizes the importance of correcting for such measurement errors, as neglecting them can lead to biased estimates. For example, Zeng et al. [16] considered outcome-dependent sampling, Shu and Yi [17] studied joint covariate–outcome correction in inverse probability weighting, and Wei et al. [18] developed validation-based efficient estimation.

In this context, identifying causal effects in the presence of both outcome misclassification and network interference remains largely unexplored. To address this gap, we develop a novel framework for estimating average causal effects under such settings with misclassified outcomes. We specify a parametric misclassification mechanism and develop a framework that jointly identifies misclassification probabilities and causal effect parameters within a binary choice model with neighborhood exposure mappings. Monte Carlo simulations evaluate the finite-sample performance of the proposed method and highlight the bias arising from ignoring outcome misclassification or network-related variables. An application to data on rural microfinance and social networks from Karnataka further demonstrated the empirical relevance of these issues, showing that under binary exposure, failure to correct for outcome misclassification can generate spurious spillover and overall effects. In particular, once outcome misclassification is corrected for, we find no evidence of positive spillover or overall effects, whereas an estimator that ignores outcome misclassification suggests statistically significant positive effects. Moreover, omitting network-related variables distorts the decomposition of causal effects and overstates the direct effect. Finally, we assess the robustness of our results to the choice of exposure mapping and to the network unconfoundedness assumption.

The remainder of this paper is organized as follows. Section 2 introduces the framework for network causal effects; Section 3 presents identification and estimation; Section 4 reports the simulations; Section 5 provides an empirical application; and Section 6 concludes the paper. Additional proofs and simulation results are provided in Appendix A.

2. Model Framework

We considered a population of a size N , indexed by $i \in \mathcal{N} = \{1, \dots, N\}$. For each individual i , we observed a binary treatment $D_i \in \mathcal{D}_i := \{0, 1\}$, covariates $X_i^{\text{ind}} \in \mathbb{R}^p$,

and a binary outcome $Y_i \in \{0, 1\}$. Let Y_i^* denote the latent (true) outcome, with Y_i observed while subject to measurement error. Let $d_{1:N} = [d_1, \dots, d_N]^\top \in \{0, 1\}^N$ be the treatment assignment vector for the population, and let us denote the potential outcome of i under assignment d by $Y_i^*(d_{1:N})$.

To make the potential outcomes framework tractable, we adopted the neighborhood interference structure proposed in Forastiere et al. [8]. Consider a population connected through a known network with an adjacency matrix $A = \{A_{ij}\}_{i,j=1}^N$. For each unit i , let $\mathcal{N}_i = \{j : A_{ij} = 1\}$ denote its neighbor set and $n_i := |\mathcal{N}_i|$ denote its cardinality. Let $d_{\mathcal{N}_i}$ and $d_{-\mathcal{N}_i}$ denote the treatment vectors for the neighbors and all other units, respectively, and let us define $g_i : \{0, 1\}^{n_i} \rightarrow \mathcal{Z}_i$ as the associated exposure mapping. We assumed that interference operated only through this exposure mapping.

Assumption 1 (Neighborhood interference). *If the exposure mapping yields the same value for two neighbor-treatment assignments, i.e., $g_i(d_{\mathcal{N}_i}) = g_i(d'_{\mathcal{N}_i})$, then the corresponding potential outcomes are equal:*

$$Y_i^*(d_i, d_{\mathcal{N}_i}, d_{-\mathcal{N}_i}) = Y_i^*(d_i, d'_{\mathcal{N}_i}, d'_{-\mathcal{N}_i}). \quad (1)$$

We define the exposure received by unit i as $Z_i = g_i(D_{\mathcal{N}_i})$, where $D_{\mathcal{N}_i} := (D_j)_{j \in \mathcal{N}_i}$ denotes the vector of neighbor treatments. Specifically, we define

$$Z_i = 1 \left(\sum_{j \in \mathcal{N}_i} D_j > 0 \right), \quad (2)$$

where Z_i is a binary indicator equal to one if individual i has at least one treated neighbor and is zero otherwise.

Under Assumption 1, potential outcomes depend only on an individual's own treatment and the corresponding induced neighborhood exposure:

$$Y_i^*(d_{1:N}) \equiv Y_i^*(d, z), \quad z = g_i(d_{\mathcal{N}_i}), \quad (3)$$

where $d \in \{0, 1\}$ denotes the individual's own treatment and $z \in \{0, 1\}$ denotes the exposure level induced by neighbors' treatment assignments.

In addition to network exposure, we allowed the covariate vector to include both individual-level characteristics and those of an individual's neighbors. Specifically, for each unit i , we observe

$$X_i = [X_i^{\text{ind}\top}, X_i^{\text{neigh}\top}]^\top, \quad (4)$$

where X_i^{ind} denotes individual-level characteristics and X_i^{neigh} summarizes the characteristics of i 's neighbors. We constructed the neighborhood covariates as the average of the neighbors' individual covariates

$$X_i^{\text{neigh}} = \begin{cases} \frac{1}{n_i} \sum_{j \in \mathcal{N}_i} X_j^{\text{ind}}, & n_i > 0, \\ 0, & n_i = 0, \end{cases} \quad (5)$$

so that neighboring characteristics entered the model as observed controls.

Given the model for potential outcomes and the observed data, we now introduce the identifying assumptions, establish the identification results, and develop estimation procedures for the causal parameters of interest.

3. Identification and Estimation

The parameter of interest is the average dose–response function (ADRF)

$$\mu_0^*(d, z) = E[Y_i^*(d, z)], \quad (6)$$

which characterizes the average potential outcome under each combination of its own treatment d and neighborhood exposure z . We first define the average direct effect (ADE) as

$$\text{ADE} = \sum_{z \in \mathcal{Z}_i} [\mu_0^*(1, z) - \mu_0^*(0, z)] P(Z_i = z), \quad (7)$$

which captures the average effect of its own treatment averaged over the distribution of neighborhood exposure. Next, the average spillover effect (ASE) for the exposure level z is defined as

$$\text{ASE}(z) = \sum_{d \in \mathcal{D}_i} [\mu_0^*(d, z) - \mu_0^*(d, 0)] P(D_i = d), \quad (8)$$

which quantifies the average change in potential outcomes induced solely by shifting the neighborhood exposure from zero to z while marginalizing over the distribution of an individual's own treatment status. Finally, the average overall effect (AOE) is defined as

$$\text{AOE} = \sum_{z \in \mathcal{Z}_i} [\mu_0^*(1, z) - \mu_0^*(0, 0)] P(Z_i = z), \quad (9)$$

which represents the total effect of moving from no treatment and no exposure to treatment with exposure.

Since potential outcomes are unobserved, we next introduce identifying assumptions, drawing on the framework of Forastiere et al. [8].

Assumption 2 (Causal framework assumptions).

(A1) *Consistency*: $Y_i^* = Y_i^*(D_i, Z_i)$.

(A2) *Network unconfoundedness*: $Y_i^*(d, z) \perp\!\!\!\perp (D_i, Z_i) \mid X_i$ for all $d \in \{0, 1\}$ and $z \in \mathcal{Z}_i$.

(A3) *Positivity*: $0 < P(D_i = d, Z_i = z \mid X_i) < 1$ for all $d \in \{0, 1\}$ and $z \in \mathcal{Z}_i$.

Together, Assumption 2 (A1–A3) ensures identification of a well-defined spillover effect in our setting. Assumption 2 (A1) links the observed outcomes to the relevant potential outcomes. Assumption 2 (A2) requires that, depending on the covariates, neither an individual's treatment nor the treatment status of their neighbors is driven by unobserved factors that also affect the potential outcomes. Assumption 2 (A3) guaranties sufficient overlap so that all treatment–exposure combinations occur with positive probability in the data.

Under Assumption 2 (A1–A3), by letting $p^*(d, z, x) = E[Y_i^* \mid D_i = d, Z_i = z, X_i = x]$, the population ADRF can be written as follows:

$$\mu_0^*(d, z) = \int p^*(d, z, x) f_X(x) dx. \quad (10)$$

Identification of the ADRF relies on accurate measurement of the outcome variable Y_i . When Y_i is subject to misclassification, conditional expectations based on the observed outcome may be systematically biased, thereby affecting the identification and estimation of $\mu_0^*(d, z)$. Let $\tilde{X}_i = [D_i, Z_i, X_i^\top]^\top \in \mathbb{R}^{2p+2}$. We assume that the latent outcome Y_i^* follows a binary choice model:

$$P(Y_i^* = 1 \mid \tilde{X}_i) = G(\tilde{X}_i^\top \beta), \quad (11)$$

where $G(\cdot) : \mathbb{R} \rightarrow (0, 1)$ is a known and strictly increasing cumulative distribution function (CDF), such as $G(t) = \Phi(t)$ in the probit model, where Φ denotes the standard normal CDF.

Let $\pi_{01} = P(Y_i = 1 \mid Y_i^* = 0)$ and $\pi_{10} = P(Y_i = 0 \mid Y_i^* = 1)$ denote the false positive rate and false negative rate, respectively. The parameter π_{01} represents the probability that a positive outcome is recorded when the true outcome is zero, while π_{10} represents the probability that a true positive outcome is recorded as zero. To characterize the dependence structure of the latent outcomes and the measurement process, we further imposed the following assumptions.

Assumption 3 (Conditional independence). *Conditional on $(\tilde{X}_i)_{i=1}^N$, the latent outcomes $(Y_i^*)_{i=1}^N$ are independent across individuals. In addition, conditional on $(Y_i^*)_{i=1}^N$, the misclassification process generating Y_i from Y_i^* is independent across individuals.*

Assumption 4 (Nondifferential misclassification). *Conditional on Y_i^* , the observed outcome is independent of the treatment, exposure, and covariates:*

$$Y_i \perp\!\!\!\perp (D_i, Z_i, X_i) \mid Y_i^*. \quad (12)$$

Under Assumption 4, the conditional probability of the observed outcome can be written as follows. The derivation is provided in Appendix A.1:

$$P(Y_i = 1 \mid \tilde{X}_i) = \pi_{01} + [1 - \pi_{10} - \pi_{01}]P(Y_i^* = 1 \mid \tilde{X}_i). \quad (13)$$

Let

$$f(Y_i \mid \tilde{X}_i; \theta) = \{P(Y_i = 1 \mid \tilde{X}_i; \theta)\}^{Y_i} \{1 - P(Y_i = 1 \mid \tilde{X}_i; \theta)\}^{1-Y_i}.$$

Under Assumption 3, the likelihood function is given by

$$\mathcal{L}(\theta) = \prod_{i=1}^N f(Y_i \mid \tilde{X}_i; \theta), \quad (14)$$

where $\theta = (\beta^\top, \pi_{10}, \pi_{01})^\top \in \mathbb{R}^{2p+4}$ collects all the model parameters.

To ensure identification, we imposed the following condition on the misclassification probabilities. Together with standard regularity conditions for the covariates, this condition yields identification of the parameter vector, as formalized in Theorem 1.

Assumption 5 (Monotonicity condition). $\pi_{10} + \pi_{01} < 1$.

Theorem 1. *Suppose that $G = \Phi$ is the standard normal CDF. Assume that $E[\|\tilde{X}_i\|^2] < \infty$, the support of $\tilde{X}_i$ contains a nonempty open subset of $\mathbb{R}^{2p+2}$, and Assumptions 1–5 hold. Then, the parameter vector θ is identified, and the expected log-likelihood $E[\log f(Y_i \mid \tilde{X}_i; \theta)]$ is uniquely maximized at the true parameter θ_0 .*

The result builds on arguments similar to those in Hausman et al. [19]. For completeness, we provide a proof of Theorem 1 in Appendix A.2, explicitly stating the support conditions required for identification. In particular, we assumed that the support of the covariates contained a nonempty open set, which ensured sufficient variation so that the conditional mean function was identified. Although Theorem 1 is stated under the probit model, the identification argument extends to other strictly increasing and continuously differentiable link functions (such as the logistic link) under additional regularity conditions.

We restricted the parameter space to

$$\Theta = \left\{ \theta : 0 \leq \pi_{10} < 1, 0 \leq \pi_{01} < 1, \pi_{10} + \pi_{01} < 1 \right\}, \quad (15)$$

and estimated θ by maximizing the log-likelihood:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \log \mathcal{L}(\theta). \quad (16)$$

We refer to this estimator as the network-based misclassification-corrected (Net-MC) estimator. Based on the estimated parameters, the estimated average dose–response function (ADRF) is constructed as follows:

$$\hat{\mu}_0^*(d, z) = \frac{1}{N} \sum_{i=1}^N G(\tilde{X}_i(d, z)^\top \hat{\beta}), \quad (17)$$

where $\tilde{X}_i(d, z) = [d, z, X_i^\top]^\top$. The average direct effect is estimated as follows:

$$\widehat{\text{ADE}} = \sum_{z \in \mathcal{Z}} [\hat{\mu}_0^*(1, z) - \hat{\mu}_0^*(0, z)] \hat{P}_N(Z = z), \quad (18)$$

where $\hat{P}_N(Z = z) = N^{-1} \sum_{i=1}^N 1(Z_i = z)$ denotes the empirical distribution of the exposure level. The average spillover effect is then computed as follows:

$$\widehat{\text{ASE}}(z) = \sum_{d \in \mathcal{D}} [\hat{\mu}_0^*(d, z) - \hat{\mu}_0^*(d, 0)] \hat{P}_N(D = d), \quad (19)$$

where $\hat{P}_N(D = d) = N^{-1} \sum_{i=1}^N 1(D_i = d)$ denotes the empirical distribution of the treatment. Finally, we estimated the average overall effect as follows:

$$\widehat{\text{AOE}} = \sum_{z \in \mathcal{Z}} [\hat{\mu}_0^*(1, z) - \hat{\mu}_0^*(0, 0)] \hat{P}_N(Z = z). \quad (20)$$

Given the dependence across individuals induced by the network structure, we conducted inference using bootstrap methods tailored to the data configuration. When the network could be partitioned into independent connected components (e.g., villages or schools), we implemented a cluster bootstrap that resampled components with replacement and recomputed the estimator in each replication (see Forastiere et al. [8]). When dependence is primarily local rather than cluster-based, a block bootstrap can instead be used to accommodate network dependence (see Kojevnikov [20]). Standard errors and confidence intervals were obtained from the bootstrap.

4. Monte Carlo Simulations

We evaluated the finite-sample performance of the proposed method through a series of Monte Carlo simulations. The experiments examined performance across varying network structures and investigated the consequences of ignoring outcome misclassification in the presence of network interference.

We considered undirected networks of a size N with an adjacency matrix A . To examine how network topology affected the performance of the estimators, we generated networks from two canonical models that differed in their link formation mechanisms and the degree distributions they induced. These designs allowed us to compare settings with relatively homogeneous degree distributions to those with substantial degree heterogeneity:

1. ER networks: As a benchmark design, the network was generated from an Erdős–Rényi random graph [21], in which each pair of distinct nodes $i \neq j$ was independently connected with a probability of $q = 6/N$. This yielded a sparse network with an expected average degree approximately equal to 6. The Erdős–Rényi model produces relatively homogeneous connectivity, with node degrees concentrated around their mean and limited variation across individuals.

2. BA networks: To introduce substantial degree heterogeneity, we generated networks using the Barabási–Albert model [22]. Starting from a small connected core, each new node attached to $m = 3$ existing nodes with a probability proportional to their degrees. This produced networks with an average degree approximately equal to 6. Unlike the Erdős–Rényi model, the Barabási–Albert process generates a heavy-tailed degree distribution characterized by a few highly connected hub nodes and many sparsely connected ones.

Given the realized network, individual-level covariates were generated as $X_i^{\text{ind}} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, I_p)$. Let $n_i = |\mathcal{N}_i|$ denote the number of neighbors of unit i . We defined the neighborhood covariate as the average of the neighbors' covariates:

$$X_i^{\text{neigh}} = \begin{cases} \frac{1}{n_i} \sum_{j \in \mathcal{N}_i} X_j^{\text{ind}}, & n_i > 0, \\ 0, & n_i = 0. \end{cases} \quad (21)$$

and defined the full covariate vector as $X_i = [X_i^{\text{ind}\top}, X_i^{\text{neigh}\top}]^\top$.

We next specify the treatment assignment mechanism. Treatment was generated according to a logistic model

$$\text{logit}\{P(D_i = 1 \mid X_i)\} = -1.5 + \sum_{j=1}^{2p} X_{ij}. \quad (22)$$

so that treatment assignment depended on the observed covariates. Neighborhood exposure is defined as

$$Z_i = 1\left(\sum_{j \in \mathcal{N}_i} D_j > 0\right), \quad (23)$$

indicating whether at least one neighbor is treated. Conditional on the treatment and exposure, the latent outcome was generated as follows:

$$Y_i^* = 1\left(D_i + Z_i + \sum_{j=1}^{2p} X_{ij} + \varepsilon_i \geq 0\right), \quad \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1). \quad (24)$$

Finally, we introduce outcome misclassification. Let

$$\pi_{01} = P(Y_i = 1 \mid Y_i^* = 0), \quad \pi_{10} = P(Y_i = 0 \mid Y_i^* = 1), \quad (25)$$

denote the false positive rate and false negative rate, respectively. These parameters represent the probabilities that the observed outcome differs from the true outcome due to misclassification. Conditional on the true outcome Y_i^* , the misclassification indicator satisfies

$$R_i \mid Y_i^* \stackrel{\text{i.i.d.}}{\sim} \begin{cases} \text{Bernoulli}(\pi_{10}), & Y_i^* = 1, \\ \text{Bernoulli}(\pi_{01}), & Y_i^* = 0. \end{cases} \quad (26)$$

The observed outcome is then constructed as follows:

$$Y_i = (1 - R_i)Y_i^* + R_i(1 - Y_i^*). \quad (27)$$

We conducted 1000 Monte Carlo replications with sample sizes $N \in \{2000, 4000, 8000\}$ and the covariate dimension $p = 20$. Although the two network designs were calibrated to have a similar average degree, they differed in their higher-order structural properties.

Table 1 reports a summary of the statistics of the simulated networks used in the Monte Carlo experiments. For the ER networks, the mean degree was approximately six for all sample sizes, with relatively small dispersion and moderate maximum degrees. In contrast, the BA networks exhibited substantially greater heterogeneity in the degree distribution, as reflected in the much larger standard deviations and maximum degrees. This difference reflects the presence of highly connected hub nodes characteristic of scale-free networks.

Table 1. Summary statistics of simulated networks.

Networks	N	Mean	SD	Max
ER	2000	5.9987	2.4441	16.0030
	4000	5.9991	2.4464	16.6360
	8000	5.9985	2.4473	17.2880
BA	2000	5.9910	7.5321	138.0380
	4000	5.9955	8.0531	195.3080
	8000	5.9977	8.5512	276.6300

Notes: The table reports a summary of the statistics of the node degrees in the simulated networks. The columns report the mean, standard deviation (SD), and maximum degree.

Table 2 reports the median absolute error (MAE) of the estimated misclassification probabilities (π_{01}, π_{10}) for the proposed Net-MC estimator across different misclassification rates and sample sizes. For both the ER and BA networks, the MAE decreased steadily as N increased from 2000 to 8000, indicating improved estimation accuracy with larger samples. The estimation performance was also quite similar across the two network designs, suggesting that the proposed method is robust to differences in network structure. As expected, the MAE tended to be slightly larger when the underlying misclassification probabilities were higher; however, the estimation error remained small and declined consistently with the sample size. Overall, these results indicate that the proposed method yields accurate and stable estimation of the misclassification probabilities.

Table 2. MAE of the estimated misclassification probabilities across network designs.

Networks	(π_{01}, π_{10})	2000	4000	8000
ER	(0.01, 0.01)	0.0032	0.0022	0.0015
	(0.01, 0.10)	0.0059	0.0037	0.0027
	(0.05, 0.10)	0.0089	0.0052	0.0038
	(0.10, 0.01)	0.0068	0.0044	0.0030
	(0.10, 0.05)	0.0095	0.0056	0.0039
BA	(0.01, 0.01)	0.0031	0.0021	0.0015
	(0.01, 0.10)	0.0060	0.0038	0.0027
	(0.05, 0.10)	0.0096	0.0052	0.0038
	(0.10, 0.01)	0.0068	0.0043	0.0028
	(0.10, 0.05)	0.0100	0.0055	0.0037

Notes: The table reports the median across Monte Carlo replications of $\frac{1}{2}(|\hat{\pi}_{01} - \pi_{01}| + |\hat{\pi}_{10} - \pi_{10}|)$, which corresponds to the median absolute error (MAE) of the estimated misclassification probabilities. Columns correspond to different sample sizes N .

Table 3 reports the median absolute error (MAE) of the estimates of β and compares the finite-sample performance of three estimators corresponding to different estimation methods:

1. Oracle: the infeasible probit estimator based on the latent outcome Y_i^* , which served as a benchmark and was invariant to misclassification probabilities.

2. Naive: the standard probit estimator applied to the observed outcome Y_i , ignoring outcome misclassification.
3. Net-MC: the proposed misclassification-corrected probit estimator that corrects for outcome misclassification and incorporates network-related variables.

As expected, the Oracle estimator achieved the smallest MAE across all settings. In contrast, the Naive estimator exhibited substantially larger errors in the presence of outcome misclassification, indicating that failure to correct for misclassification leads to substantial estimation errors. The proposed Net-MC estimator substantially outperformed the Naive estimator across all misclassification probabilities, with MAEs that were markedly smaller and approached those of the Oracle estimator as the sample size increased. Overall, these results demonstrate that explicitly correcting for outcome misclassification, together with incorporating network-related variables, led to more accurate estimation of the coefficient vector β .

Table 3. MAE of β estimates under different misclassification probabilities in ER networks.

(π_{01}, π_{10})	N	Oracle	Naive	Net-MC
(0.01, 0.01)	2000	0.1359	0.4394	0.1561
	4000	0.0801	0.4538	0.0888
	8000	0.0514	0.4616	0.0570
(0.01, 0.10)	2000	0.1359	0.7139	0.2254
	4000	0.0801	0.7194	0.1105
	8000	0.0514	0.7216	0.0698
(0.05, 0.10)	2000	0.1359	0.7530	0.3246
	4000	0.0801	0.7575	0.1305
	8000	0.0514	0.7587	0.0790
(0.10, 0.01)	2000	0.1359	0.6743	0.2272
	4000	0.0801	0.6792	0.1085
	8000	0.0514	0.6826	0.0681
(0.10, 0.05)	2000	0.1359	0.7405	0.3363
	4000	0.0801	0.7446	0.1296
	8000	0.0514	0.7456	0.0778

Notes: The table reports the median absolute error (MAE) of the estimates for the coefficient vector β across the Monte Carlo replications. Oracle denotes the infeasible probit estimator based on the latent outcome Y_i^* , which is invariant to misclassification probabilities. Naive denotes the standard probit estimator applied to the observed outcome Y_i , ignoring misclassification. Net-MC denotes the proposed misclassification-corrected probit estimator that corrects for outcome misclassification and incorporates network-related variables.

Table 4 reports the simulation results for the ADE estimates under ER networks, evaluated in terms of bias and root mean squared error (RMSE). In addition to the three estimators considered above, we considered a benchmark specification, namely the non-network misclassification-corrected (Non-Net-MC) estimator, which applies the misclassification-corrected probit estimator while omitting neighbor exposure and network-related covariates. The results show that this benchmark exhibited substantial bias, underscoring the importance of accounting for network spillovers in causal inference. We report the results for Non-Net-MC estimator only for the ADE analysis. By construction, it does not incorporate variation in neighbors' variables and therefore precludes the identification of network-related effects. Consequently, neither the average spillover effect (ASE) nor the average overall effect (AOE) was well defined under this specification. Among the feasible estimators, the proposed Net-MC estimator yielded ADE estimates with the smallest bias and RMSE, indicating that jointly accounting for outcome misclassification and network spillovers is crucial for accurate estimation of the ADE.

The results for the ASE and AOE under ER networks were qualitatively similar to those under BA networks, and they are reported in Tables A1–A6 in Appendix A.3.

Overall, the results were consistent across all settings. The proposed Net-MC estimator yielded ASE and AOE estimates with a uniformly smaller bias and RMSE than the Naive estimator, and its performance approached that of the Oracle estimator as the sample size increased. In contrast, estimators that failed to correct for outcome misclassification or to account for network-related variables produced substantially biased estimates. These findings underscore the importance of jointly correcting for outcome misclassification and accounting for network-related variables to obtain credible and reliable causal inference.

Table 4. Simulation results for ADE estimates under ER networks.

(π_{01}, π_{10})	Metric	N	Oracle	Naive	Non-Net-MC	Net-MC
(0.01, 0.01)	Bias	2000	0.0030	−0.0115	0.2954	0.0055
		4000	0.0010	−0.0117	0.2770	0.0013
		8000	0.0006	−0.0121	0.2686	0.0007
	RMSE	2000	0.0251	0.0277	0.3054	0.0331
		4000	0.0152	0.0212	0.2828	0.0171
		8000	0.0107	0.0173	0.2721	0.0117
(0.01, 0.10)	Bias	2000	0.0030	−0.0262	0.2994	0.0100
		4000	0.0010	−0.0259	0.2784	0.0011
		8000	0.0006	−0.0256	0.2743	0.0012
	RMSE	2000	0.0251	0.0379	0.3079	0.0499
		4000	0.0152	0.0319	0.2825	0.0216
		8000	0.0107	0.0292	0.2774	0.0149
(0.05, 0.10)	Bias	2000	0.0030	0.0046	0.2991	0.0097
		4000	0.0010	0.0051	0.2797	0.0010
		8000	0.0006	0.0053	0.2769	0.0014
	RMSE	2000	0.0251	0.0303	0.3086	0.0516
		4000	0.0152	0.0209	0.2839	0.0226
		8000	0.0107	0.0162	0.2794	0.0156
(0.10, 0.01)	Bias	2000	0.0030	0.0971	0.2837	0.0078
		4000	0.0010	0.0968	0.2730	0.0018
		8000	0.0006	0.0958	0.2703	0.0008
	RMSE	2000	0.0251	0.1026	0.2906	0.0396
		4000	0.0152	0.0994	0.2745	0.0186
		8000	0.0107	0.0971	0.2709	0.0128
(0.10, 0.05)	Bias	2000	0.0030	0.0475	0.2917	0.0097
		4000	0.0010	0.0473	0.2844	0.0017
		8000	0.0006	0.0467	0.2744	0.0009
	RMSE	2000	0.0251	0.0569	0.2995	0.0470
		4000	0.0152	0.0516	0.2891	0.0219
		8000	0.0107	0.0491	0.2802	0.0149

Notes: The table reports the bias and root mean squared error (RMSE) of the average direct effect (ADE) estimates. Oracle denotes the infeasible probit estimator based on the latent outcome Y_i^* and is thus invariant to misclassification probabilities. Naive denotes the standard probit estimator applied to the observed outcome Y_i , ignoring misclassification. Non-Net-MC refers to a benchmark specification that employs the misclassification-corrected probit estimator but omits network-related covariates. In contrast, Net-MC denotes the proposed misclassification-corrected probit estimator that corrects for outcome misclassification and incorporates network-related variables.

5. Empirical Analysis

5.1. Data Description

Our empirical analysis used village-level microfinance and social network data compiled by Banerjee et al. [15]. The dataset covers 75 villages in Karnataka, India, of which the microfinance institution Bharatha Swamukti Samsthe (BSS) operated in 43 villages. It combines a household census with a partial individual census, providing detailed information on household characteristics, economic behavior, and social interactions. This dataset

is particularly well suited for our analysis, as our research question required data that jointly capture treatment, network relationships, covariates, and binary outcomes, variables that are rarely jointly observed in empirical settings. It provides detailed information on both the social network structure and household-level borrowing, thereby meeting the key data requirements for studying network spillover effects in the presence of outcome misclassification. The dataset has also been widely used in the literature on social networks and economic behavior (e.g., Breza et al. [23], Lubold et al. [24], Lambotte [25]), supporting its reliability and empirical relevance. In addition, microfinance participation and informal borrowing behavior are inherently shaped by social interactions, as information diffusion and peer effects play an important role in household financial decision making.

The network adjacency matrix A_{ij} was constructed by following Banerjee et al. [15]. In particular, an undirected link between two households was defined as the union of reported visits, where a connection was formed if either household reported visiting the other. This definition is intended to capture economically meaningful relationships through which financial assistance and information may flow. Let D_i denote whether household i participates in the BSS program, and let Y_i indicate whether the household engages in informal borrowing, including loans from moneylenders, relatives, or friends. Following Banerjee et al. [15], we controlled for 11 baseline household characteristics that may affect both microfinance participation and borrowing behavior. These variables included the presence of an eligible female household member (age 18–57), general or OBC caste status, whether any household member served as a BSS leader, household size, the number of rooms and beds, access to a latrine and electricity, housing tenure status, and indicators for an RCC roof and a thatched roof. Together, these covariates capture key dimensions of household demographics, socioeconomic status, and housing quality that may influence both program participation and borrowing behavior.

The sample consisted of 8080 households across 43 BSS villages. There was considerable variation in village-level coverage, with complete-case sample sizes ranging from a minimum of 75 to a maximum of 305 households per village, with a mean of 187.91 and a median of 184.00. The village social networks exhibited substantial heterogeneity in connectivity. Household degrees ranged from 0 to 49, with an average of approximately nine social connections. This structural diversity indicates significant variation in the potential exposure to treated neighbors, which is essential for identifying spillover effects. Figure 1 presents the social network in a representative village. The network exhibited both densely connected clusters and households occupying more central positions. For visual clarity, the figure displays only the largest connected component of the village network.

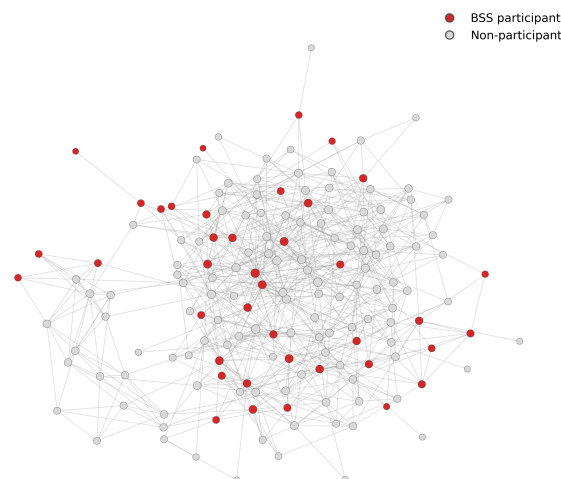


Figure 1. An illustrative social network from a representative village.

5.2. Empirical Results

Because the dataset combines multiple sources and relies on self-reported borrowing, the informal credit variable may be subject to measurement error due to reporting mistakes or recall bias. To address this issue, we allowed for potential misclassification of the binary outcome in estimating causal effects. We primarily report the estimates and 95% percentile confidence intervals (CIs) for the misclassification probabilities of informal borrowing, as well as the key causal effect parameters, under binary exposure. These confidence intervals were obtained from 1000 cluster bootstrap replications, in which villages were resampled with replacement.

Table 5 reports the estimated misclassification probabilities for the informal borrowing outcome obtained using the proposed Net-MC estimator. False positives (π_{01}), defined as recording borrowing when none occurred, were extremely rare (0.0085). In contrast, false negatives (π_{10}), defined as failing to report borrowing when it did occur, were substantially more common (0.1771). This asymmetry suggests that informal borrowing is more likely to be underreported than overreported in survey data.

Table 5. Estimated misclassification probabilities for informal borrowing.

Parameter	Estimate	95% Percentile CI
π_{01}	0.0085	[0.0052, 0.2499]
π_{10}	0.1771	[0.0001, 0.2002]

Notes: This table reports the estimated probabilities of outcome misclassification under binary exposure. π_{01} is the probability of reporting borrowing when none occurred, and π_{10} is the probability of failing to report actual borrowing. The 95% CI refers to the percentile bootstrap confidence interval, based on 1000 village-level bootstrap replications.

Table 6 reports the ADE, ASE, and AOE estimates based on alternative estimators. For the ADE, we reported estimates from three estimators: Naive, Non-Net-MC, and Net-MC. The Non-Net-MC estimator does not incorporate variation in network-related variables and therefore does not identify spillover effects; accordingly, the ASE and AOE were reported only for the Naive and Net-MC estimators.

For the ADE, the Non-Net-MC estimator yielded an estimate of 0.0003, which is slightly larger than the corresponding estimate from the Net-MC estimator. Once network-related variables were accounted for, the estimated ADE was substantially reduced, taking values of 0.0153 under the Naive estimator and -0.0005 under the Net-MC estimator. Although all three estimates were statistically insignificant, the comparison remains informative. The Non-Net-MC estimator tended to overstate the ADE when network-related variables were excluded. A similar pattern emerged for the spillover-related effects. Under the Naive estimator, the ASE and AOE were 0.0812 and 0.0968, respectively, both of which were statistically significant. In contrast, once outcome misclassification was corrected for, the spillover and overall effects became negligible and statistically insignificant, with the estimated ASE and AOE equal to -0.0085 and -0.0090 , respectively, under the Net-MC estimator.

Taken together, these results point to two distinct sources of bias in the empirical settings with network interactions. First, comparing the Non-Net-MC estimator with those that incorporate network-related variables shows that the omission of neighbor exposure and covariates can distort the decomposition of causal effects, attributing part of the spillover effect to the direct effect and thereby inflating the estimated ADE. Second, comparing the Naive and Net-MC estimators indicates that failure to correct for outcome misclassification can generate spurious evidence of spillover and overall effects, substantially inflating both their magnitude and statistical significance. Overall, these findings underscore the importance of jointly accounting for network-related variables and

correcting for outcome misclassification in order to obtain credible and unbiased estimates of causal effects.

Table 6. Estimates of ADE, ASE, and AOE using the Karnataka data under binary exposure.

Parameter	Estimator	Estimate	95% CI
ADE	Naive	0.0153	[−0.0103, 0.0448]
	Non-Net-MC	0.0003	[−0.0058, 0.0262]
	Net-MC	−0.0005	[−0.0025, 0.0013]
ASE	Naive	0.0812	[0.0501, 0.1147]
	Net-MC	−0.0085	[−0.0101, 0.0023]
AOE	Naive	0.0968	[0.0516, 0.1454]
	Net-MC	−0.0090	[−0.0116, 0.0037]

Notes: This table reports point estimates and 95% confidence intervals for the ADE, ASE, and AOE under binary exposure. The 95% CI refers to the percentile bootstrap confidence interval, based on 1000 village-level bootstrap replications.

5.3. Robustness Checks

In this subsection, we examine the robustness of our results to the modeling choices and identifying assumptions underlying the baseline analysis. We begin by replacing the binary exposure with a proportion-based exposure, defined as the fraction of treated neighbors, and examine whether our main conclusions continue to hold. Formally, the proportion-based exposure is defined as follows:

$$Z_i = \frac{1}{n_i} \sum_{j \in \mathcal{N}_i} D_j. \quad (28)$$

For isolated individuals with no social ties ($n_i = 0$), $Z_i = 0$ was set to zero.

In the analysis based on proportion-based exposure, we selected 0.1, 0.3, 0.5, and 0.7 as representative exposure levels according to several quantiles of the neighboring treatment proportion Z and estimated the corresponding causal effects at each level. The results are reported in Table 7. Compared with the results under the binary exposure specification (Table 6), the estimates were not fully identical across the two exposure definitions. In particular, under the Naive estimator, the spillover effect exhibited substantial differences; under binary exposure, the Naive estimate of the ASE was significantly positive, whereas under proportion-based exposure, the corresponding Naive estimates at different exposure levels were all close to zero and slightly negative. This suggests that the Naive estimator is sensitive to the definition of exposure.

It is worth noting that the Naive estimator is based directly on the observed outcome and does not correct for outcome misclassification. Therefore, the difference between the Naive and corrected estimates should not be interpreted solely as arising from the change in the exposure definition. By contrast, the Net-MC estimator exhibited a more stable pattern across the two exposure specifications, with the estimates of the ADE, ASE, and AOE all remaining statistically insignificant. Moreover, under proportion-based exposure, the Net-MC estimates of the ASE declined gradually from −0.0004 to −0.0027 as the exposure level increased from 0.1 to 0.7, indicating a mildly stronger negative spillover effect at higher exposure levels. However, the magnitude of this pattern remained small, and the statistical evidence was limited.

Overall, this robustness check indicates that although the Naive estimator varied across exposure definitions, the main conclusion based on the Net-MC estimator remained largely unchanged. The estimated direct, spillover, and overall effects were generally small and statistically insignificant. Therefore, our core empirical findings appear reasonably

robust to alternative exposure definitions, although this robustness is primarily reflected in the corrected estimates rather than in the uncorrected estimates.

Table 7. Estimates of ADE, ASE, and AOE using the Karnataka data under proportion-based exposure.

Parameter	Estimator	Estimate	95% CI
ADE	Naive	0.0343	[0.0105, 0.0588]
	Non-Net-MC	0.0003	[−0.0054, 0.0530]
	Net-MC	−0.0008	[−0.0026, 0.0020]
ASE(0.1)	Naive	−0.0020	[−0.0126, 0.0076]
	Net-MC	−0.0004	[−0.0006, 0.0000]
ASE(0.3)	Naive	−0.0061	[−0.0370, 0.0231]
	Net-MC	−0.0011	[−0.0018, 0.0000]
ASE(0.5)	Naive	−0.0101	[−0.0600, 0.0391]
	Net-MC	−0.0019	[−0.0030, 0.0001]
ASE(0.7)	Naive	−0.0141	[−0.0817, 0.0554]
	Net-MC	−0.0027	[−0.0040, 0.0003]
AOE	Naive	0.0298	[−0.0047, 0.0655]
	Net-MC	−0.0016	[−0.0036, 0.0018]

Notes: This table reports the point estimates and 95% percentile confidence intervals for the ADE, ASE, and AOE under proportion-based exposure. The 95% CI refers to the percentile bootstrap confidence interval, based on 1000 village-level bootstrap replications.

Beyond examining the robustness of our results to the exposure definition, we next turn to another potential concern. Because identification in our observational network setting relies on a network unconfoundedness assumption, our empirical findings may still be affected by omitted factors both observed and unobserved. In particular, the household treatment status and neighborhood exposure may be correlated with characteristics that are not fully accounted for by the available baseline covariates. To assess the sensitivity of our results to this concern, we conducted an additional robustness check. Although this issue cannot be fully ruled out empirically, this analysis allowed for a more cautious and ultimately more persuasive interpretation of the empirical findings.

As a diagnostic, we augmented the baseline specification with the village-level fixed effects. Specifically, we replaced the latent index model with

$$P(Y_i^* = 1 \mid \tilde{X}_i, V_i) = G\left(\tilde{X}_i^\top \beta + \sum_{k=2}^K \delta_k 1\{V_i = k\}\right), \quad (29)$$

where $1\{V_i = k\}$ denotes an indicator equal to one if household i belongs to village k and the first village is omitted as the reference category. With the village-level fixed effects, the comparison was made among households within the same village. This means that the estimates were not driven by average differences across villages but by differences among households living in the same village.

This analysis is particularly relevant in our context because both treatment participation and network exposure may be influenced by village-level characteristics. The results are reported in Table 8. Compared with the baseline Net-MC results under binary exposure, the estimates obtained after including the village-level fixed effects remained broadly similar in both sign and magnitude. Specifically, the estimated ADE, ASE, and AOE were all statistically insignificant. These findings suggest that our main results were not driven primarily by omitted village-level factors, and they provide further support for the conclusion that the direct, spillover, and overall effects are all statistically insignificant after correcting for outcome misclassification.

Table 8. Estimated ADE, ASE, and AOE with village-level fixed effects under binary exposure.

Parameter	Estimate	95% CI
ADE	−0.0021	[−0.0057, 0.0042]
ASE	−0.0059	[−0.0090, 0.0074]
AOE	−0.0080	[−0.0127, 0.0098]

Notes: This table reports estimates from the Net-MC estimator with village-level fixed effects. The 95% CI refers to the percentile bootstrap confidence interval, based on 1000 village-level bootstrap replications.

6. Conclusions

This paper studied how outcome misclassification affects the estimation of causal effects in network settings. We modeled interference through exposure mappings and adopted a parametric binary choice framework that allowed the misclassification probabilities and causal effect parameters to be jointly identified.

Monte Carlo simulations showed that estimators that ignore outcome misclassification or network-related variables can suffer from substantial bias, whereas the proposed method performed well with finite samples. The empirical analysis based on data from Karnataka further highlights, under binary exposure, two distinct sources of bias in network settings that help explain these differences. First, omitting network-related variables can distort the decomposition of causal effects and lead to an overstatement of the direct effect. Second, failure to correct for outcome misclassification can generate misleading evidence of spillover and overall effects. Once outcome misclassification was corrected for, we found no evidence of spillover or overall effects of the neighbors' microcredit participation on households' informal borrowing. Taken together, these results underscore that credible estimation of causal effects in network settings requires jointly accounting for network-related variables and correcting for outcome misclassification.

Finally, we assessed the robustness of our results to the choice of exposure mapping and the network unconfoundedness assumption. Although our empirical analysis delivered a clear and robust message, its causal interpretation remains conditional on the plausibility of several core assumptions, including the parametric link specification, the exposure mapping, and the network unconfoundedness assumption. In addition, our analysis assumed nondifferential misclassification, under which the misclassification error was independent of the observed covariates. Relaxing this assumption by allowing misclassification probabilities to vary with the observed characteristics would introduce additional identification challenges. Addressing these challenges is an important direction for future research and would further broaden the scope of causal inference with misclassified outcomes in network settings.

While our empirical application was based on a specific regional context, the framework we developed is not restricted to that setting. More generally, in empirical applications where network relationships, treatment, and outcome measures are available, the variables required for our approach can typically be constructed. Our framework therefore provides a general tool for evaluating causal effects in network environments, particularly when outcome misclassification is present.

Author Contributions: Conceptualization, Y.L. and M.L.; methodology, Y.L. and M.L.; software, Y.L.; formal analysis, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L. and M.L.; visualization, Y.L.; supervision, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available from Zenodo at <https://doi.org/10.5281/zenodo.7706650> (Banerjee et al. [15]).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. Derivation of Equation (13)

Proof. Let

$$\begin{aligned} p(d, z, x) &= P(Y_i = 1 \mid D_i = d, Z_i = z, X_i = x), \\ p^*(d, z, x) &= P(Y_i^* = 1 \mid D_i = d, Z_i = z, X_i = x). \end{aligned}$$

Under the law of total probability, we have

$$\begin{aligned} p(d, z, x) &= P(Y_i = 1 \mid Y_i^* = 1, D_i = d, Z_i = z, X_i = x)P(Y_i^* = 1 \mid D_i = d, Z_i = z, X_i = x) \\ &\quad + P(Y_i = 1 \mid Y_i^* = 0, D_i = d, Z_i = z, X_i = x)P(Y_i^* = 0 \mid D_i = d, Z_i = z, X_i = x). \end{aligned}$$

Under Assumption 4, the misclassification probabilities do not depend on (D_i, Z_i, X_i) . Hence, we have

$$\begin{aligned} P(Y_i = 1 \mid Y_i^* = 1, D_i = d, Z_i = z, X_i = x) &= P(Y_i = 1 \mid Y_i^* = 1) = 1 - \pi_{10}, \\ P(Y_i = 1 \mid Y_i^* = 0, D_i = d, Z_i = z, X_i = x) &= P(Y_i = 1 \mid Y_i^* = 0) = \pi_{01}. \end{aligned}$$

Substituting these expressions into the previous equation yields

$$p(d, z, x) = (1 - \pi_{10}) p^*(d, z, x) + \pi_{01}(1 - p^*(d, z, x)).$$

Rearranging the terms gives

$$p(d, z, x) = \pi_{01} + (1 - \pi_{10} - \pi_{01}) p^*(d, z, x),$$

This completes the proof. $\square$

Appendix A.2. Proof of Theorem 1

Proof. Note that

$$P(Y_i = 1 \mid \tilde{X}_i = x; \theta) = \pi_{01} + (1 - \pi_{10} - \pi_{01})\Phi(x^\top \beta),$$

where $\theta = (\beta^\top, \pi_{10}, \pi_{01})^\top \in \Theta$.

First, we explain why Assumption 5 is required for identification. Consider the probit link Φ , which satisfies the symmetry property

$$\Phi(-t) = 1 - \Phi(t).$$

For any

$$\theta = (\beta^\top, \pi_{10}, \pi_{01})^\top \in \Theta,$$

we define

$$\tilde{\theta} := (-\beta, 1 - \pi_{01}, 1 - \pi_{10}).$$

Then, we have

$$\begin{aligned} P(Y_i = 1 \mid \tilde{X}_i = x; \tilde{\theta}) &= (1 - \pi_{10}) + \left[1 - (1 - \pi_{01}) - (1 - \pi_{10})\right]\Phi(-x^\top \beta) \\ &= (1 - \pi_{10}) - (1 - \pi_{10} - \pi_{01})\Phi(-x^\top \beta) \\ &= (1 - \pi_{10}) - (1 - \pi_{10} - \pi_{01})\{1 - \Phi(x^\top \beta)\} \\ &= \pi_{01} + (1 - \pi_{10} - \pi_{01})\Phi(x^\top \beta) \\ &= P(Y_i = 1 \mid \tilde{X}_i = x; \theta). \end{aligned}$$

Thus, θ and $\tilde{\theta}$ generate the same conditional probability function. However, we have

$$(1 - \pi_{01}) + (1 - \pi_{10}) = 2 - (\pi_{10} + \pi_{01}) > 1,$$

and thus $\tilde{\theta}$ does not belong to the parameter space Θ . Therefore, without the restriction $\pi_{10} + \pi_{01} < 1$, the model would not be identified.

We next prove identification. Let

$$a := \pi_{01}, \quad b := 1 - \pi_{10} - \pi_{01},$$

Similarly, we define

$$a' := \pi'_{01}, \quad b' := 1 - \pi'_{10} - \pi'_{01}.$$

Under Assumption 5, $b > 0$, and $b' > 0$.

Suppose that

$$a + b\Phi(\tilde{X}_i^\top \beta) = a' + b'\Phi(\tilde{X}_i^\top \beta') \quad \text{a.s.}$$

Since both sides are continuously differentiable in x , and the support of $\tilde{X}_i$ contains a nonempty open subset, it follows that

$$a + b\Phi(x^\top \beta) = a' + b'\Phi(x^\top \beta') \quad (\text{A1})$$

for all x in some nonempty open subset $\mathcal{U}$.

Differentiating both sides of Equation (A1) twice with respect to x yields

$$b\phi(x^\top \beta)\beta = b'\phi(x^\top \beta')\beta' \quad \forall x \in \mathcal{U}.$$

Because $\phi(\cdot) > 0$, the two sides are nonzero scalar multiples of β and β' . Hence, β and β' must be collinear. There exists $\lambda \neq 0$ such that

$$\beta' = \lambda\beta.$$

Substituting this into the gradient identity gives

$$b\phi(x^\top \beta) = b'\lambda\phi(\lambda x^\top \beta) \quad \forall x \in \mathcal{U}.$$

Since $\mathcal{U}$ is open and $\beta \neq 0$, the scalar $x^\top \beta$ varies over a nondegenerate interval. Hence, we have

$$b\phi(t) = b'\lambda\phi(\lambda t) \quad \forall t \in I$$

for some nondegenerate interval I .

By using

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2},$$

this implies

$$e^{-t^2/2} = \frac{b'\lambda}{b} e^{-\lambda^2 t^2/2} \quad \forall t \in I.$$

Therefore, $\lambda^2 = 1$. Since $b > 0$, $b' > 0$, and $\phi > 0$, we must have $\lambda > 0$, and hence $\lambda = 1$. Thus, we have

$$\beta = \beta'.$$

By substituting $\beta = \beta'$ back into Equation (A1), we get

$$(a - a') + (b - b')\Phi(x^\top \beta) = 0 \quad \forall x \in \mathcal{U}.$$

Because $\mathcal{U}$ is open and $\beta \neq 0$, the function $\Phi(x^\top \beta)$ is nonconstant on $\mathcal{U}$. Hence, we have

$$a = a', \quad b = b'.$$

Therefore, we have

$$\pi_{01} = \pi'_{01}, \quad \pi_{10} = \pi'_{10},$$

and thus $\theta = \theta'$.

Note that

$$f(Y_i | \tilde{X}_i; \theta) = \{P(Y_i = 1 | \tilde{X}_i; \theta)\}^{Y_i} \{1 - P(Y_i = 1 | \tilde{X}_i; \theta)\}^{1-Y_i}.$$

We define

$$Q(\theta) := E[\log f(Y_i | \tilde{X}_i; \theta)].$$

Finally, we verify that the population log-likelihood $Q(\theta)$ is uniquely maximized at θ_0 . Let θ_0 denote the true parameter value. Through identification, if $\theta \neq \theta_0$, then

$$f(Y_i | \tilde{X}_i; \theta) \neq f(Y_i | \tilde{X}_i; \theta_0) \quad \text{with positive probability.}$$

If, in addition,

$$E[|\log f(Y_i | \tilde{X}_i; \theta)|] < \infty \quad \forall \theta \in \Theta, \quad (\text{A2})$$

then under Lemma 2.2 of Newey and McFadden [26], it follows that $Q(\theta)$ is uniquely maximized at θ_0 .

It remains to verify Equation (A2). Since

$$P(Y_i = 1 | \tilde{X}_i; \theta) = \pi_{01} + (1 - \pi_{10} - \pi_{01})\Phi(\tilde{X}_i^\top \beta),$$

and $1 - \pi_{10} - \pi_{01} > 0$, there exists a constant $c > 0$ such that

$$P(Y_i = 1 | \tilde{X}_i; \theta) \geq c \Phi(\tilde{X}_i^\top \beta), \quad 1 - P(Y_i = 1 | \tilde{X}_i; \theta) \geq c \Phi(-\tilde{X}_i^\top \beta).$$

Hence, for some constant $C > 0$, we have

$$|\log f(Y_i | \tilde{X}_i; \theta)| \leq C + |\log \Phi(\tilde{X}_i^\top \beta)| + |\log \Phi(-\tilde{X}_i^\top \beta)|.$$

Let $u := \tilde{X}_i^\top \beta$. It is well known that

$$\frac{d}{du} \log \Phi(u) = \frac{\phi(u)}{\Phi(u)} =: \lambda(u),$$

where $\phi(u)$ is the standard normal density. The function $\lambda(u)$ is continuous, convex, and satisfies

$$\frac{\lambda(u)}{-u} \rightarrow 1 \quad \text{as } u \rightarrow -\infty, \quad \lambda(u) \rightarrow 0 \quad \text{as } u \rightarrow \infty.$$

Hence, there exists a constant $C > 0$ such that

$$|\lambda(u)| \leq C(1 + |u|) \quad \forall u \in \mathbb{R}.$$

Under the mean value theorem, for some $\tilde{u}$ between 0 and u , we have

$$|\log \Phi(u)| = |\log \Phi(0) + \lambda(\tilde{u}) u| \leq |\log \Phi(0)| + |\lambda(\tilde{u})| |u|.$$

Using $|\tilde{u}| \leq |u|$ and the above bound on $\lambda(\cdot)$, we obtain

$$|\log \Phi(u)| \leq C(1 + (1 + |u|)|u|) \leq C(1 + u^2),$$

for some constant $C > 0$, which may vary from line to line. The same bound applies to $|\log \Phi(-u)|$. Therefore, we have

$$|\log f(Y_i | \tilde{X}_i; \theta)| \leq C(1 + u^2).$$

Since $u = \tilde{X}_i^\top \beta$, we have

$$u^2 \leq \|\beta\|^2 \|\tilde{X}_i\|^2,$$

and hence

$$|\log f(Y_i | \tilde{X}_i; \theta)| \leq C(1 + \|\tilde{X}_i\|^2).$$

Thus, under $E[\|\tilde{X}_i\|^2] < \infty$, we obtain

$$E[|\log f(Y_i | \tilde{X}_i; \theta)|] < \infty,$$

which verifies Equation (A2).

This completes the proof. $\square$

Appendix A.3. Additional Simulation Results

This subsection reports additional simulation results that complement the main analysis. Specifically, we present the results for the ADE and AOE estimates under the ER network design, as well as the results for the coefficient estimates and for the ADE, ASE, and AOE estimates under the BA network design.

Table A1. Simulation results for ASE estimates under ER networks.

(π_{01}, π_{10})	Metric	N	Oracle	Naive	Net-MC
(0.01, 0.01)	Bias	2000	0.0016	−0.0011	0.0016
		4000	−0.0002	−0.0018	−0.0001
		8000	0.0009	−0.0007	0.0008
	RMSE	2000	0.0204	0.0263	0.0217
		4000	0.0135	0.0181	0.0144
		8000	0.0097	0.0131	0.0101
(0.01, 0.10)	Bias	2000	0.0016	0.0203	0.0041
		4000	−0.0002	0.0205	0.0010
		8000	0.0009	0.0206	0.0012
	RMSE	2000	0.0204	0.0363	0.0265
		4000	0.0135	0.0295	0.0166
		8000	0.0097	0.0258	0.0117
(0.05, 0.10)	Bias	2000	0.0016	0.0035	0.0030
		4000	−0.0002	0.0035	0.0008
		8000	0.0009	0.0042	0.0009
	RMSE	2000	0.0204	0.0335	0.0306
		4000	0.0135	0.0238	0.0191
		8000	0.0097	0.0178	0.0129

Table A1. *Cont.*

(π_{01}, π_{10})	Metric	N	Oracle	Naive	Net-MC
(0.10, 0.01)	Bias	2000	0.0016	−0.0308	0.0006
		4000	−0.0002	−0.0316	−0.0008
		8000	0.0009	−0.0295	0.0004
	RMSE	2000	0.0204	0.0445	0.0269
		4000	0.0135	0.0386	0.0172
		8000	0.0097	0.0336	0.0118
(0.10, 0.05)	Bias	2000	0.0016	−0.0180	0.0032
		4000	−0.0002	−0.0187	−0.0004
		8000	0.0009	−0.0172	0.0004
	RMSE	2000	0.0204	0.0381	0.0313
		4000	0.0135	0.0296	0.0190
		8000	0.0097	0.0240	0.0129

Notes: The table reports the bias and root mean squared error (RMSE) of the ASE estimates across the Monte Carlo replications. Definitions of the estimation methods are provided in the notes to Table 4.

Table A2. Simulation results for AOE estimates under ER networks.

(π_{01}, π_{10})	Metric	N	Oracle	Naive	Net-MC
(0.01, 0.01)	Bias	2000	0.0047	−0.0129	0.0073
		4000	0.0008	−0.0138	0.0013
		8000	0.0016	−0.0130	0.0015
	RMSE	2000	0.0339	0.0381	0.0415
		4000	0.0214	0.0292	0.0236
		8000	0.0151	0.0223	0.0162
(0.01, 0.10)	Bias	2000	0.0047	−0.0069	0.0145
		4000	0.0008	−0.0064	0.0022
		8000	0.0016	−0.0061	0.0024
	RMSE	2000	0.0339	0.0428	0.0576
		4000	0.0214	0.0307	0.0288
		8000	0.0151	0.0229	0.0197
(0.05, 0.10)	Bias	2000	0.0047	0.0074	0.0130
		4000	0.0008	0.0079	0.0018
		8000	0.0016	0.0088	0.0024
	RMSE	2000	0.0339	0.0458	0.0601
		4000	0.0214	0.0332	0.0314
		8000	0.0151	0.0253	0.0212
(0.10, 0.01)	Bias	2000	0.0047	0.0679	0.0087
		4000	0.0008	0.0669	0.0010
		8000	0.0016	0.0682	0.0012
	RMSE	2000	0.0339	0.0817	0.0495
		4000	0.0214	0.0740	0.0262
		8000	0.0151	0.0720	0.0184
(0.10, 0.05)	Bias	2000	0.0047	0.0299	0.0133
		4000	0.0008	0.0291	0.0014
		8000	0.0016	0.0300	0.0013
	RMSE	2000	0.0339	0.0541	0.0590
		4000	0.0214	0.0430	0.0302
		8000	0.0151	0.0381	0.0204

Notes: The table reports the bias and root mean squared error (RMSE) of the AOE estimates across the Monte Carlo replications. Definitions of the estimation methods are provided in the notes to Table 4.

Table A3. MAEs of β estimates under different misclassification probabilities in BA networks.

(π_{01}, π_{10})	N	Oracle	Naive	Net-MC
(0.01, 0.01)	2000	0.1359	0.4442	0.1536
	4000	0.0783	0.4614	0.0859
	8000	0.0503	0.4701	0.0537
(0.01, 0.10)	2000	0.1359	0.7192	0.2315
	4000	0.0783	0.7243	0.1057
	8000	0.0503	0.7260	0.0659
(0.05, 0.10)	2000	0.1359	0.7573	0.3593
	4000	0.0783	0.7610	0.1264
	8000	0.0503	0.7628	0.0749
(0.10, 0.01)	2000	0.1359	0.6783	0.2395
	4000	0.0783	0.6847	0.1062
	8000	0.0503	0.6877	0.0657
(0.10, 0.05)	2000	0.1359	0.7452	0.3771
	4000	0.0783	0.7488	0.1215
	8000	0.0503	0.7499	0.0753

Notes: The table reports the MAE of the estimates of the coefficient vector β across the Monte Carlo replications. Definitions of the estimation methods are provided in the notes to Table 3.

Table A4. Simulation results for ADE estimates under BA networks.

(π_{01}, π_{10})	Metric	N	Oracle	Naive	Non-Net-MC	Net-MC
(0.01, 0.01)	Bias	2000	0.0034	−0.0078	0.3185	0.0044
		4000	0.0016	−0.0081	0.3074	0.0015
		8000	0.0003	−0.0082	0.3039	0.0005
	RMSE	2000	0.0220	0.0269	0.3338	0.0296
		4000	0.0156	0.0194	0.3142	0.0168
		8000	0.0104	0.0147	0.3066	0.0114
(0.01, 0.10)	Bias	2000	0.0034	−0.0200	0.3254	0.0055
		4000	0.0016	−0.0189	0.3115	0.0019
		8000	0.0003	−0.0196	0.3067	0.0007
	RMSE	2000	0.0220	0.0346	0.3325	0.0393
		4000	0.0156	0.0271	0.3159	0.0211
		8000	0.0104	0.0237	0.3116	0.0150
(0.05, 0.10)	Bias	2000	0.0034	0.0116	0.3248	0.0072
		4000	0.0016	0.0131	0.3130	0.0022
		8000	0.0003	0.0124	0.3117	0.0007
	RMSE	2000	0.0220	0.0335	0.3320	0.0461
		4000	0.0156	0.0248	0.3162	0.0225
		8000	0.0104	0.0191	0.3137	0.0158
(0.10, 0.01)	Bias	2000	0.0034	0.1034	0.3164	0.0070
		4000	0.0016	0.1037	0.3065	0.0021
		8000	0.0003	0.1033	0.3081	0.0005
	RMSE	2000	0.0220	0.1089	0.3216	0.0385
		4000	0.0156	0.1064	0.3093	0.0186
		8000	0.0104	0.1046	0.3111	0.0122
(0.10, 0.05)	Bias	2000	0.0034	0.0544	0.3224	0.0061
		4000	0.0016	0.0557	0.3131	0.0028
		8000	0.0003	0.0551	0.3084	0.0005
	RMSE	2000	0.0220	0.0627	0.3281	0.0413
		4000	0.0156	0.0598	0.3160	0.0221
		8000	0.0104	0.0572	0.3134	0.0142

Notes: The table reports the bias and root mean squared error (RMSE) of the ADE estimates across the Monte Carlo replications. Definitions of the estimation methods are provided in the notes to Table 4.

Table A5. Simulation results for ASE estimates under BA networks.

(π_{01}, π_{10})	Metric	N	Oracle	Naive	Net-MC
(0.01, 0.01)	Bias	2000	−0.0001	−0.0008	−0.0001
		4000	0.0006	−0.0001	0.0005
		8000	0.0002	0.0000	0.0002
	RMSE	2000	0.0176	0.0233	0.0189
		4000	0.0126	0.0170	0.0134
		8000	0.0081	0.0116	0.0087
(0.01, 0.10)	Bias	2000	−0.0001	0.0189	0.0008
		4000	0.0006	0.0199	0.0007
		8000	0.0002	0.0196	0.0004
	RMSE	2000	0.0176	0.0334	0.0228
		4000	0.0126	0.0283	0.0150
		8000	0.0081	0.0241	0.0100
(0.05, 0.10)	Bias	2000	−0.0001	0.0054	0.0012
		4000	0.0006	0.0072	0.0005
		8000	0.0002	0.0061	0.0003
	RMSE	2000	0.0176	0.0298	0.0258
		4000	0.0126	0.0219	0.0167
		8000	0.0081	0.0166	0.0108
(0.10, 0.01)	Bias	2000	−0.0001	−0.0239	0.0003
		4000	0.0006	−0.0228	0.0005
		8000	0.0002	−0.0233	0.0001
	RMSE	2000	0.0176	0.0370	0.0234
		4000	0.0126	0.0304	0.0157
		8000	0.0081	0.0273	0.0100
(0.10, 0.05)	Bias	2000	−0.0001	−0.0124	0.0018
		4000	0.0006	−0.0110	0.0009
		8000	0.0002	−0.0119	0.0003
	RMSE	2000	0.0176	0.0323	0.0274
		4000	0.0126	0.0239	0.0169
		8000	0.0081	0.0195	0.0110

Notes: The table reports the bias and root mean squared error (RMSE) of the ASE estimates across the Monte Carlo replications. Definitions of the estimation methods are provided in the notes to Table 4.

Table A6. Simulation results for AOE estimates under BA networks.

(π_{01}, π_{10})	Metric	N	Oracle	Naive	Net-MC
(0.01, 0.01)	Bias	2000	0.0034	−0.0087	0.0044
		4000	0.0022	−0.0083	0.0021
		8000	0.0005	−0.0084	0.0007
	RMSE	2000	0.0295	0.0364	0.0368
		4000	0.0215	0.0267	0.0229
		8000	0.0141	0.0194	0.0154
(0.01, 0.10)	Bias	2000	0.0034	−0.0019	0.0066
		4000	0.0022	0.0002	0.0026
		8000	0.0005	−0.0009	0.0011
	RMSE	2000	0.0295	0.0422	0.0479
		4000	0.0215	0.0300	0.0272
		8000	0.0141	0.0204	0.0187
(0.05, 0.10)	Bias	2000	0.0034	0.0165	0.0088
		4000	0.0022	0.0198	0.0028
		8000	0.0005	0.0180	0.0010
	RMSE	2000	0.0295	0.0480	0.0551
		4000	0.0215	0.0370	0.0294
		8000	0.0141	0.0281	0.0199

Table A6. Cont.

(π_{01}, π_{10})	Metric	N	Oracle	Naive	Net-MC
(0.10, 0.01)	Bias	2000	0.0034	0.0813	0.0076
		4000	0.0022	0.0829	0.0026
		8000	0.0005	0.0820	0.0006
	RMSE	2000	0.0295	0.0932	0.0475
		4000	0.0215	0.0888	0.0258
		8000	0.0141	0.0850	0.0168
(0.10, 0.05)	Bias	2000	0.0034	0.0426	0.0082
		4000	0.0022	0.0454	0.0037
		8000	0.0005	0.0439	0.0008
	RMSE	2000	0.0295	0.0613	0.0517
		4000	0.0215	0.0550	0.0295
		8000	0.0141	0.0491	0.0189

Notes: The table reports the bias and root mean squared error (RMSE) of the AOE estimates across the Monte Carlo replications. Definitions of the estimation methods are provided in the notes to Table 4.

References

- Halloran, M.E.; Struchiner, C.J. Causal inference in infectious diseases. *Epidemiology* **1995**, *6*, 142–151. [\[CrossRef\]](#)
- Tchetgen Tchetgen, E.J.; Fulcher, I.R.; Shpitser, I. Auto-g-computation of causal effects on a network. *J. Am. Stat. Assoc.* **2021**, *116*, 833–844. [\[CrossRef\]](#)
- Sobel, M.E. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *J. Am. Stat. Assoc.* **2006**, *101*, 1398–1407. [\[CrossRef\]](#)
- Tchetgen Tchetgen, E.J.; VanderWeele, T.J. On causal inference in the presence of interference. *Stat. Methods Med. Res.* **2012**, *21*, 55–75. [\[CrossRef\]](#)
- Qu, Z.; Xiong, R.; Liu, J.; Imbens, G. Semiparametric estimation of treatment effects in observational studies with heterogeneous partial interference. *arXiv* **2021**, arXiv:2107.12420. [\[CrossRef\]](#)
- Manski, C.F. Identification of treatment response with social interactions. *Econom. J.* **2013**, *16*, S1–S23. [\[CrossRef\]](#)
- Aronow, P.M.; Samii, C. Estimating average causal effects under general interference, with application to a social network experiment. *Ann. Appl. Stat.* **2017**, *11*, 1912–1947. [\[CrossRef\]](#)
- Forastiere, L.; Airolidi, E.M.; Mealli, F. Identification and estimation of treatment and interference effects in observational studies on networks. *J. Am. Stat. Assoc.* **2021**, *116*, 901–918. [\[CrossRef\]](#)
- Leung, M.P. Treatment and spillover effects under network interference. *Rev. Econ. Stat.* **2020**, *102*, 368–380. [\[CrossRef\]](#)
- Leung, M.P. Causal inference under approximate neighborhood interference. *Econometrica* **2022**, *90*, 267–293. [\[CrossRef\]](#)
- Bargagli-Stoffi, F.J.; Tortu, C.; Forastiere, L. Heterogeneous treatment and spillover effects under clustered network interference. *Ann. Appl. Stat.* **2025**, *19*, 28–55. [\[CrossRef\]](#)
- Bong, H.; Fogarty, C.B.; Levina, E.; Zhu, J. Heterogeneous treatment effects under network interference: A nonparametric approach based on node connectivity. *arXiv* **2024**, arXiv:2410.11797.
- Lee, C.; Zeng, D.; Hudgens, M.G. Efficient nonparametric estimation of stochastic policy effects with clustered interference. *J. Am. Stat. Assoc.* **2025**, *120*, 382–394. [\[CrossRef\]](#)
- Buchanan, A.L.; Hernández-Ramírez, R.U.; Lok, J.J.; Vermund, S.H.; Friedman, S.R.; Forastiere, L.; Spiegelman, D. Assessing direct and spillover effects of intervention packages in network-randomized studies. *Epidemiology* **2024**, *35*, 481–488. [\[CrossRef\]](#)
- Banerjee, A.; Breza, E.; Chandrasekhar, A.G.; Duflo, E.; Jackson, M.O.; Kinnan, C. Changes in social network structure in response to exposure to formal credit markets. *Rev. Econ. Stud.* **2024**, *91*, 1331–1372. [\[CrossRef\]](#)
- Zeng, M.; Jia, Z.; Sui, Z.; Xu, J.; Zhang, H. Causal inference with outcome dependent sampling and mismeasured outcome. *arXiv* **2023**, arXiv:2309.11764.
- Shu, D.; Yi, G.Y. Weighted causal inference methods with mismeasured covariates and misclassified outcomes. *Stat. Med.* **2019**, *38*, 1835–1854. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wei, S.; Zhang, C.; Geng, Z.; Luo, S. Identifiability and estimation for potential-outcome means with misclassified outcomes. *Mathematics* **2024**, *12*, 2801. [\[CrossRef\]](#)
- Hausman, J.A.; Abrevaya, J.; Scott-Morton, F.M. Misclassification of the dependent variable in a discrete-response setting. *J. Econom.* **1998**, *87*, 239–269. [\[CrossRef\]](#)
- Kojevnikov, D. The bootstrap for network dependent processes. *arXiv* **2021**, arXiv:2101.12312.
- Erdős, P.; Rényi, A. On random graphs I. *Publ. Math. Debr.* **1959**, *6*, 290–297. [\[CrossRef\]](#)

22. Barabási, A.L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [[CrossRef](#)]
23. Breza, E.; Chandrasekhar, A.G.; McCormick, T.H.; Pan, M. Using aggregated relational data to feasibly identify network structure without network data. *Am. Econ. Rev.* **2020**, *110*, 2454–2484. [[CrossRef](#)] [[PubMed](#)]
24. Lubold, S.; Chandrasekhar, A.G.; McCormick, T.H. Identifying the latent space geometry of network models through analysis of curvature. *J. R. Stat. Soc. Ser. B–Stat. Methodol.* **2023**, *85*, 240–292. [[CrossRef](#)]
25. Lambotte, M. Peer effects in binary outcomes: Strategic complementarity and taste for conformity with endogenous networks. *J. Appl. Econom.* **2025**, *40*, 608–626. [[CrossRef](#)]
26. Newey, W.K.; McFadden, D. Large sample estimation and hypothesis testing. *Handb. Econom.* **1994**, *4*, 2111–2245.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.