ROBUST PROPAGATION-INTEGRATED INFERENCE FOR OBSERVATIONAL STUDIES WITH NETWORK INTERFERENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

In observational studies with network interference, the treatment-propensity error can affect exposure probabilities through the network, creating additional challenges for point estimation and uncertainty quantification. We propose Propagation-Integrated Inference (PIN), which constructs observable propagation directions through cross-fitting and uses the propagation directions for both point estimation and uncertainty quantification. PIN incorporates the propagation directions into a first-order estimator and then uses second-order propagation correction to adjust the remaining propensity–outcome product error. PIN further constructs a propagation-integrated standard error that simultaneously accounts for network dependence and the adjustment used in point estimation. We derive an upper bound on the approximation error of the propagation directions, prove that second-order propagation correction does not increase the upper bound of the remaining propensity–outcome product error, and establish asymptotic normality of the final PIN point estimator and conservative coverage of confidence intervals. Across three treatment-propensity shift simulations, PIN maintains 97–98% average coverage, with 5–12-point higher coverage than the state-of-the-art baseline and only 5–9% longer intervals, while reducing absolute bias by 32–78%. In a real-network study under high treatment-propensity shift, PIN achieves 97% average coverage and reduces absolute bias by approximately 85–89% relative to the baseline.

1 INTRODUCTION

Network interference means that a node's outcome depends on the node's own treatment and the treatment of neighboring nodes (Wu et al., 2025). Reliable point estimation and uncertainty quantification need to account for network interference (Cai et al., 2024). In randomized network experiments, the exposure probabilities of nodes to treatment are known (Gao and Ding, 2025). In observational studies, exposure probabilities must be obtained by estimating treatment propensity from the data. The error between the estimated treatment propensity and the unobserved true treatment propensity cannot be observed and can also propagate through the network to other nodes (Hays and Raghavan, 2025). We refer to this error as *the propagated exposure-probability error*.

We first identify three challenges to achieving robustness to network interference in observational studies. The *Direction* challenge is to identify observable information that can correct the propagated exposure-probability error (Adhikari et al., 2026). The *Bias* challenge arises because the propagated exposure-probability error may underestimate or overestimate the weights of some nodes, thereby compromising point estimation. The *Uncertainty* challenge arises because the propagated exposure-probability error affects point estimation variability, thereby leading to inaccurate uncertainty quantification (Chen et al., 2026).

To address these three challenges, we propose Propagation-Integrated Inference (PIN). For *Direction*, PIN uses cross-fitting to construct observable propagation directions from pretreatment features and treatment-prediction residuals. For *Bias*, PIN first constructs a first-order estimator to eliminate the error that arises when either the propagated exposure-probability error or the outcome-side error occurs alone. PIN then incorporates a second-order propagation correction to adjust for the remaining propensity–outcome product error that arises when both types of errors are present simultaneously. For *Uncertainty*, PIN incorporates the propagation adjustment used in point estimation into uncertainty

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

quantification and constructs a propagation-integrated standard error to simultaneously reflect the additional uncertainty induced by network dependence and treatment-propensity error propagation.

Theoretical results provide justification for PIN. We prove that observable propagation directions can approximate the correction direction of the propagated exposure-probability error. We further prove that second-order propagation correction can reduce the upper bound of the error induced by the propagated exposure-probability error. Under the corresponding estimation rate conditions, we prove that the final PIN point estimator has asymptotic normality. We also prove that confidence intervals constructed based on the propagation-integrated standard error have conservative coverage.

We evaluate propensity-scale, fixed-overlap nonlinear-rank, and network-aligned rank treatment-propensity shift simulations, together with a real-network treatment-propensity shift. Figure 1 summarizes the results across different simulations. As treatment-propensity shift increases, average coverage is approximately 98%, 97%, and 97%, compared with 93%, 89%, and 85% for WLS+, the state-of-the-art baseline, while our intervals are only 5–9% longer than baselines. The mechanism ladders in Figure 1 show that first-order estimator repair removes 41–68% of the bias of WLS+, and second-order estimator removes an additional 5–9%, leaving only 23–54% of the original bias.

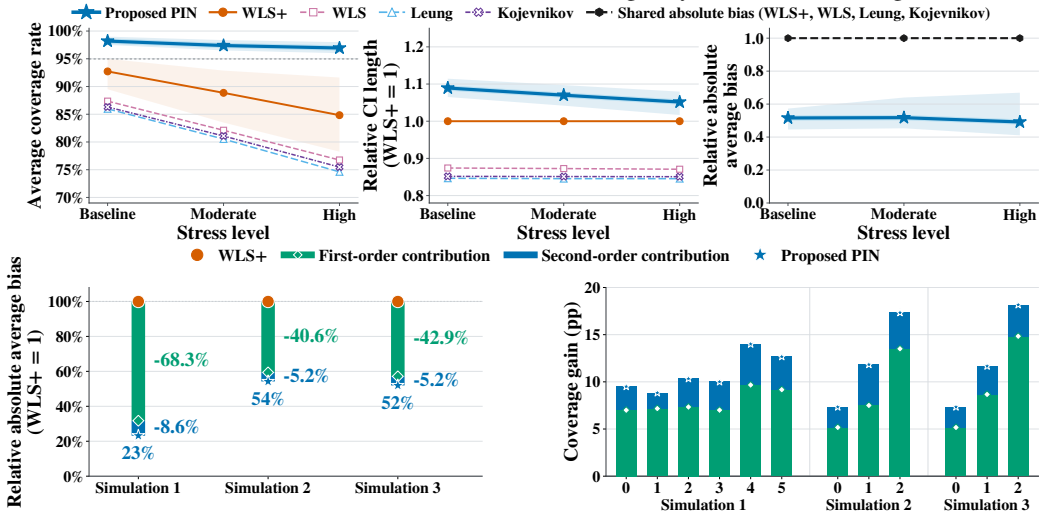


Figure 1: Top: coverage, relative CI length, and relative absolute average bias. Bottom: sequential first- and second-order contributions to bias reduction and coverage gains (Section 5.1).

2 Network Interference

Network interference means that unit i's outcome can depend on both its own treatment and neighbors' treatments. Let $\mathcal{N}_n = \{1, \ldots, n\}$ index units and $A = (A_{ij})$ denote the observed network, where $A_{ij} = 1$ indicates a link between units i and j. Let $D = (D_1, \ldots, D_n)^\top \in \{0,1\}^n$ denote the random network-wide treatment assignment, where D_i indicates whether unit i is treated, and let $Y_i(\cdot)$ denote unit i's potential-outcome function over treatment assignments. The observed outcome is $Y_i(D)$. The exposure mapping $T_i = T(i, D, A) \in \mathcal{T}$ summarizes unit i's treatment-network configuration. For example, $T_i = 1$ can indicate at least one treated neighbor even when $D_i = 0$. Define the exposure probability $\pi_i(t) = \Pr(T_i = t)$, which is known in randomized experiments.

For two exposure states $t, t' \in \mathcal{T}$, define $\mu_i(t) = \frac{\mathbb{E}_D[Y_i(D)\mathbf{1}\{T_i=t\}]}{\pi_i(t)}$, the mean outcome of unit i when its exposure T_i equals t. Averaging across units gives $\mu(t) = n^{-1}\sum_{i=1}^n \mu_i(t)$. Our estimand $\tau(t,t') = \mu(t) - \mu(t') = n^{-1}\sum_{i=1}^n \mu_i(t) - \mu_i(t')$ compares the population mean outcomes when $T_i = t$ versus $T_i = t'$ (Zhu et al., 2025; Zhang and Wang, 2025; Huang and Tian, 2026).

In randomized network experiments, A and the potential-outcome functions $Y_i(\cdot)$ are treated as fixed, and the known treatment-assignment mechanism determines the exposure probabilities $\pi_i(t)$ (Aronow and Samii, 2017; Leung, 2022; Gao and Ding, 2025). For each unit i, however, we observe only its realized exposure T_i and outcome $Y_i(D)$. Even if $T_i = t$, this provides only one realized outcome under exposure t, not the exposure-specific mean $\mu_i(t) = \frac{\mathbb{E}_D[Y_i(D)\mathbf{1}\{T_i=t\}]}{\pi_i(t)}$, and no outcome is observed for that unit under exposure t'. Thus, $\mu_i(t)$ and $\mu_i(t')$, and hence $\tau(t,t') = \mu(t) - \mu(t')$,

remain unknown and must be estimated. Under interference, the same treatment D_j can affect node i's exposure and outcomes, making their contributions to $\widehat{\tau}(t,t')$ dependent. Thus, inference requires a point estimate $\widehat{\tau}(t,t')$ and an estimate of $\mathrm{Var}\{\widehat{\tau}(t,t')\}$ that accounts for network dependence.

For point estimation, inverse-exposure weighting accounts for unequal probabilities of observing each exposure. Regressing Y_i on exposure indicators with weights $1/\pi_i(T_i)$ gives the Hájek mean $\widehat{\mu}(t) = \{\sum_i \mathbf{1}(T_i = t)Y_i/\pi_i(t)\}/\{\sum_i \mathbf{1}(T_i = t)/\pi_i(t)\}$ and the contrast $\widehat{\tau}(t,t') = \widehat{\mu}(t) - \widehat{\mu}(t')$.

For uncertainty quantification, network interference requires accounting for dependence across nodes. Let $\widehat{r}_i = Y_i - \widehat{\mu}(T_i)$ and define the centered node-level score for node i as

$$\widehat{s}_i(t,t') = \frac{\mathbf{1}\{T_i = t\}\widehat{r}_i/\pi_i(t)}{\sum_j \mathbf{1}\{T_j = t\}/\pi_j(t)} - \frac{\mathbf{1}\{T_i = t'\}\widehat{r}_i/\pi_i(t')}{\sum_j \mathbf{1}\{T_j = t'\}/\pi_j(t')}.$$

To account for dependence between nearby nodes, set $K_{ij} = \mathbf{1}\{\ell_A(i,j) \le b_n\}$, where $\ell_A(i,j)$ is the shortest-path distance in A and b_n is the distance cutoff. Replacing the negative eigenvalues of K with zero gives K^+ and the nonnegative variance estimate $\widehat{\mathrm{Var}}^+\{\widehat{\tau}(t,t')\} = \sum_{i=1}^n \sum_{j=1}^n K^+_{ij}\widehat{s}_i(t,t')\widehat{s}_j(t,t')$. The weighted least-squares point estimator and the corrected network variance estimator together constitute WLS+, a state-of-the-art design-based method for network interference (Gao and Ding, 2025).

3 Observational Network Interference: Challenges

Unlike randomized experiments, the exposure probabilities $\pi_i(t) = \Pr(T_i = t)$ are unknown in observational studies because the treatment-assignment mechanism is unknown. We therefore estimate the node-level treatment propensity $e_i = \Pr(D_i = 1 \mid H_i)$ from pretreatment features H_i, obtaining $\widehat{e}^{\text{plug-in}} = (\widehat{e}_1^{\text{plug-in}}, \ldots, \widehat{e}_n^{\text{plug-in}})^\top$. The observed network A and exposure mapping then transform $\widehat{e}^{\text{plug-in}}$ into the estimated exposure probability

$$\widehat{\pi}_i^{\text{plug-in}}(t) = \pi_i(t; \widehat{e}^{\text{plug-in}}, A) = \begin{cases} \displaystyle\prod_{j:A_{ij}=1} (1 - \widehat{e}_j^{\text{plug-in}}), & t = 0, \\ \displaystyle 1 - \prod_{j:A_{ij}=1} (1 - \widehat{e}_j^{\text{plug-in}}), & t = 1. \end{cases}$$

WLS+ makes the consequences of replacing $\pi_i(t)$ by $\widehat{\pi}_i^{\text{plug-in}}(t)$ explicit: $\pi_i(t)$ enters both the inverse-exposure weights for point estimation and the centered node-level score used in $\widehat{\mathrm{Var}}^+\{\widehat{\tau}(t,t')\}$ for uncertainty quantification. In observational studies, errors in $\widehat{e}^{\text{plug-in}}$ enter both $\widehat{\tau}(t,t') = \widehat{\mu}(t) - \widehat{\mu}(t')$ and $\widehat{\mathrm{Var}}^+\{\widehat{\tau}(t,t')\}$ through $\widehat{\pi}_i^{\text{plug-in}}(t) = \pi_i(t; \widehat{e}^{\text{plug-in}}, A)$. We refer to $\widehat{e}^{\text{plug-in}} - e$ as the *treatment-propensity error*, where $e = (e_1, \ldots, e_n)^\top$ is the vector of unknown true treatment propensities.

Existing observational methods include generalized-propensity-score adjustment (Forastiere et al., 2021), semiparametric influence function estimation for a single observational network (Ogburn et al., 2024), nonparametric doubly robust estimation (Liu et al., 2023), and machine-learning-based AIPW with network-aware cross-fitting (Emmenegger et al., 2025). The cited methods address confounding, flexible nuisance estimation, double robustness, or network dependence, and are included as baselines in Section 5. To our knowledge, the cited formulations do not jointly track the treatment-propensity error $\widehat{e}^{\text{plug-in}} - e$ through A into both point estimation and uncertainty quantification.

Directly substituting the plug-in estimate $\widehat{e}^{\text{plug-in}}$ for the treatment propensities e in both the point estimator and the variance estimator is not innocuous. Figure 2 identifies three resulting challenges.

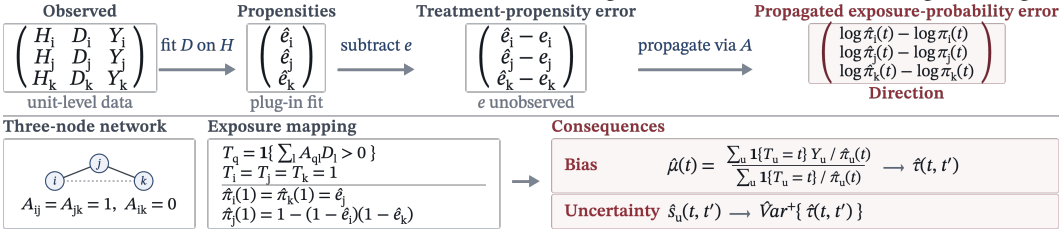


Figure 2: Three-node illustration of the propensity challenges in observational studies.

Direction. Through the network A, the treatment-propensity error $\widehat{e}^{\text{plug-in}} - e$ becomes a vector of values of the *propagated exposure-probability error* across nodes, $\{\log \widehat{\pi}_i^{\text{plug-in}}(t) - \log \pi_i(t)\}_{i=1}^n$.

For example, in Figure 2, $\widehat{\pi}_i^{\text{plug-in}}(1) = \widehat{\pi}_k^{\text{plug-in}}(1) = \widehat{e}_j^{\text{plug-in}}$. If $\widehat{e}_j^{\text{plug-in}} > e_j$, then both $\widehat{\pi}_i^{\text{plug-in}}(1)$ and $\widehat{\pi}_k^{\text{plug-in}}(1)$ are overestimated. The plug-in probabilities $\{\widehat{\pi}_i^{\text{plug-in}}(t)\}_{i=1}^n$ alone do not reveal which nodes have overestimated or underestimated exposure probabilities, because the true treatment propensities e are *unobservable*. The challenge is therefore to find *observable* directions that reflect how treatment-propensity errors related to pretreatment features H_i can propagate through the network A and generate systematic patterns in $\{\log \widehat{\pi}_i^{\text{plug-in}}(t) - \log \pi_i(t)\}_{i=1}^n$.

Bias. The first downstream branch in Figure 2 uses $\widehat{\pi}_i^{\text{plug-in}}(t)$ in the normalized inverse-exposure weights $\mathbf{1}\{T_i = t\}/\{\widehat{\pi}_i^{\text{plug-in}}(t)\widehat{h}(t)\}$. In observational studies, the propagated exposure-probability error is often not randomly distributed across nodes. The pretreatment features H_i associated with the propagated exposure-probability error may also predict the outcome $\mu_i(t)$ (Brookhart et al., 2006). For example, nodes with certain values of H_i may have $\widehat{\pi}_i^{\text{plug-in}}(t) > \pi_i(t)$, while also having larger exposure-specific mean outcomes, $\mu_i(t) > \mu(t)$, so the normalized inverse-exposure weights underweight nodes with larger $\mu_i(t)$. This alignment between the propagated exposure-probability error and heterogeneity in $\mu_i(t)$ can shift $\widehat{\mu}(t)$, and consequently $\widehat{\tau}(t,t')$, away from $\mu(t)$ and $\tau(t,t')$.

Uncertainty. $\widehat{\text{Var}}^{+}\widehat{\tau}(t,t')$ depends on the centered node-level score $\widehat{s}_i(t,t')$, and the construction of $\widehat{s}_i(t,t')$ requires the correct exposure probabilities $\pi_i(t)$. In observational studies, the true $\pi_i(t)$ is unobserved, so the centered node-level score is instead evaluated using the plug-in exposure probabilities, $\widehat{s}_i(t,t';\widehat{\pi}^{\text{plug-in}})$. Because only $\widehat{e}^{\text{plug-in}}$ and $\widehat{\pi}^{\text{plug-in}}$ are observable, the uncertainty challenge is to construct a new uncertainty quantification procedure whose discrepancy from the ideal uncertainty quantification based on the true π remains sufficiently small.

4 Propagation-Integrated Inference

4.1 Step 1: Construct the Propagation Directions

The *Direction* challenge arises because $\widehat{e}^{\text{plug-in}} - e$ are unobserved. Even when $\widehat{e}^{\text{plug-in}}$ contains errors, $\widehat{e}^{\text{plug-in}}$ alone does not reveal the directions along which $\widehat{e}^{\text{plug-in}} - e$ should be corrected. PIN therefore constructs a *score-completed path* $\widehat{e}^{\text{complete}}$. Using $\widehat{e}^{\text{plug-in}}$ and the observed pretreatment features H_i, PIN estimates a propensity correction $\widehat{\phi}$ that captures $\widehat{e}^{\text{plug-in}} - e$ related to H_i. The network A then maps the correction $\widehat{\phi}$ for H_i into the score-completed exposure probabilities $\widehat{\pi}_i^{\text{complete}}(t)$. We also provides theoretical justification that $\{\log \widehat{\pi}_i^{\text{complete}}(t) - \log \widehat{\pi}_i^{\text{plug-in}}(t)\}_{i=1}^n$ can provide observable propagation directions for $\{\log \widehat{\pi}_i^{\text{plug-in}}(t) - \log \pi_i(t)\}_{i=1}^n$.

Let $\mathcal{I}_1, \ldots, \mathcal{I}_L$ partition the analysis nodes $\mathcal{N}_n$. Fold $\mathcal{I}_\ell$ is held out while the propensity model is fitted on $\mathcal{J}_\ell = \mathcal{N}_n \setminus \mathcal{I}_\ell$. Neither D nor Y enters the fold construction. Cross-fitting over $\ell = 1, \ldots, L$ and assembling the held-out predictions yields a propensity prediction for every node in $\mathcal{N}_n$.

Plug-in Path. For each outer fold $\mathcal{I}_\ell$, the propensity model is fitted using $\{(H_k, D_k) : k \in \mathcal{J}_\ell\}$. For each held-out node $i \in \mathcal{I}_\ell$, the plug-in treatment propensity is

$$\widehat{e}_i^{\text{plug-in}} = \widehat{\Pr}(D_i = 1 \mid H_i; \{(H_k, D_k) : k \in \mathcal{J}_\ell\}), \qquad i \in \mathcal{I}_\ell.$$

Cycling through the outer folds $\ell = 1, \ldots, L$ and assembling the predictions yields $\widehat{e}^{\text{plug-in}} = \left(\widehat{e}_1^{\text{plug-in}}, \ldots, \widehat{e}_n^{\text{plug-in}}\right)^\top$. The plug-in exposure probabilities are $\widehat{\pi}_i^{\text{plug-in}}(t) = \pi_i(t; \widehat{e}^{\text{plug-in}}, A)$.

Score-Completed Path. Within the outer-training set $\mathcal{J}_\ell$ associated with fold $\mathcal{I}_\ell$, PIN partitions $\mathcal{J}_\ell$ into inner folds $\mathcal{J}_{\ell,1}, \ldots, \mathcal{J}_{\ell,Q}$. For each $k \in \mathcal{J}_{\ell,q}$, the propensity model is fitted using $\{(H_r, D_r) : r \in \mathcal{J}_\ell \setminus \mathcal{J}_{\ell,q}\}$, yielding the inner out-of-fold treatment propensity

$$\widetilde{e}_k = \widehat{\Pr}(D_k = 1 \mid H_k; \{(H_r, D_r) : r \in \mathcal{J}_\ell \setminus \mathcal{J}_{\ell,q}\}), \qquad k \in \mathcal{J}_{\ell,q}.$$

Cycling through $q = 1, \ldots, Q$ and assembling the predictions yields $\widetilde{e}_k \in \mathcal{J}_\ell$ for the entire outer-training set $\mathcal{J}_\ell$. Because $\widetilde{e}_k$ is the predicted treatment probability for node k, while D_k is the observed treatment indicator, the difference $D_k - \widetilde{e}_k$ gives the treatment-prediction residual. $|\mathcal{J}_\ell|^{-1} \sum_{k \in \mathcal{J}_\ell} H_k(D_k - \widetilde{e}_k)$ therefore measures treatment-prediction residuals associated with the pretreatment features H_k. PIN then converts $|\mathcal{J}_\ell|^{-1} \sum_{k \in \mathcal{J}_\ell} H_k(D_k - \widetilde{e}_k)$ into the fold-specific correction coefficient $\widehat{\phi}_\ell$ by the one-step Newton update given in Appendix A.1 and applies

$$\operatorname{logit} \widehat{e}_i^{\text{complete}} = \operatorname{logit} \widehat{e}_i^{\text{plug-in}} + H_i^\top \widehat{\phi}_\ell, \qquad i \in \mathcal{I}_\ell.$$

The treatment labels and all outcomes in $\mathcal{I}_\ell$ are excluded from the computation of $\widehat{\phi}_\ell$. Cycling through the outer folds $\ell = 1, \ldots, L$ and assembling the predictions yields $\widehat{e}^{\text{complete}} = (\widehat{e}_1^{\text{complete}}, \ldots, \widehat{e}_n^{\text{complete}})^\top$. Finally, PIN computes

$$\widehat{\pi}_i^{\text{complete}}(t) = \pi_i(t; \widehat{e}^{\text{complete}}, A).$$

Remark 1. Even if the overall $\widehat{e}^{\text{plug-in}}$ is accurate, there may still be some values of H_k for which D_k is often higher or lower than $\widetilde{e}_k$ (Zhao, 2019). The score-completed path uses inner cross-fitting so that D_k is not used to predict its own $\widetilde{e}_k$, thereby avoiding in-sample fitting; $\widehat{\phi}_\ell$ summarizes the systematic relationship between $D_k - \widetilde{e}_k$ and H_k over $\mathcal{J}_\ell$ into a coefficient vector, and uses $H_i^\top \widehat{\phi}_\ell$ to make the correction to $\widehat{e}_i^{\text{plug-in}}$ for each held-out node $i \in \mathcal{I}_\ell$.

For each exposure $t \in \mathcal{T}$, define the *propagation directions*:

$$d_i^{\text{complete}}(t) = \log \widehat{\pi}_i^{\text{complete}}(t) - \log \widehat{\pi}_i^{\text{plug-in}}(t). \tag{1}$$

Proposition 4.1. *Under Assumptions A.1–A.3. Let $a_n > 0$ and $\rho_n > 0$ be deterministic sequences satisfying $a_n \to 0$ and $\rho_n \to 0$, and $\Delta_n = \max_{1 \le i \le n} \sum_{j=1}^n A_{ij}$*

$$\max_{1 \le i \le n} \left| \operatorname{logit} \widehat{e}_i^{\text{complete}} - \operatorname{logit} e_i \right| = O_p(a_n^2 + \rho_n + n^{-1/2}). \tag{2}$$

$$\max_{1 \le i \le n,\, t \in \{0,1\}} \left| d_i^{\text{complete}}(t) - \left\{ \log \pi_i(t) - \log \widehat{\pi}_i^{\text{plug-in}}(t) \right\} \right| = O_p\left\{ \Delta_n (a_n^2 + \rho_n + n^{-1/2}) \right\}. \tag{3}$$

Proposition 4.1 provides theoretical justification for using $d_i^{\text{complete}}(t)$ as a propagation direction. To use propagation directions in the subsequent bias correction and uncertainty quantification, PIN collects them into a node-level vector $\widehat{U}_i$. Fix an ordering $\mathcal{T} = \{t_1, \ldots, t_{|\mathcal{T}|}\}$:

$$\widehat{U}_i = \left(d_i^{\text{complete}}(t_1),\ \ldots,\ d_i^{\text{complete}}(t_{|\mathcal{T}|}) \right)^\top \in \mathbb{R}^r, \qquad r = |\mathcal{T}|. \tag{4}$$

4.2 Step 2: the Propagation-Integrated Estimator

The *Bias* challenge arises because the pretreatment features H_i can be associated with both the *propagated exposure-probability error* $\log \widehat{\pi}_i^{\text{plug-in}}(t) - \log \pi_i(t)$ and the *outcome-side error* in predicting $\mu_i(t)$, represented below by $\widehat{m}_i(t) - \mu_i(t)$. When the propagated exposure-probability error and the outcome-side error are present, the interaction between the propagated exposure-probability error and the outcome-side error generates *the remaining propensity–outcome product error* (Hirshberg and Wager, 2021). Step 2.1 constructs $\widehat{\tau}^{(1)}(t, t')$ to cancel the contribution of either the propagated exposure-probability error or the outcome-side error acting alone. Step 2.2 then adds $\widehat{\Delta}^{(2)}_{\text{prop}}(t, t')$ to address the remaining propensity–outcome product error through $\widehat{U}_i$.

4.2.1 Step 2.1: From Propagation Directions to the First-Order Estimator.

PIN uses the same outer folds and splitting as Step 1. For each outer fold ℓ, PIN fits the exposure-specific outcome model on the training nodes,

$$Y_k \approx \widehat{\alpha}_{t,-\ell} + \widehat{\gamma}_{t,-\ell}^\top \widehat{U}_k, \qquad k \in \mathcal{J}_\ell, \quad T_k = t.$$

The fitted coefficients are then applied to each held-out node $i \in \mathcal{I}_\ell$: $\widehat{m}_i(t) = \widehat{\alpha}_{t,-\ell} + \widehat{\gamma}_{t,-\ell}^\top \widehat{U}_i, i \in \mathcal{I}_\ell$. The coefficient γ_t measures outcome variation along $\widehat{U}_i$ under exposure t. The residual $Y_i - \widehat{m}_i(t)$ corrects the outcome model and supports uncertainty estimation.

Write $\widehat{h}(t) = n^{-1} \sum_{j=1}^n \mathbf{1}\{T_j = t\} / \widehat{\pi}_j^{\text{plug-in}}(t)$ for the Hájek normalization at exposure t. PIN estimates the mean response under exposure t by

$$\widehat{\mu}^{(1)}(t) = \frac{1}{n} \sum_{i=1}^n \widehat{m}_i(t) + \frac{1}{n\widehat{h}(t)} \sum_{i=1}^n \frac{\mathbf{1}\{T_i = t\}}{\widehat{\pi}_i^{\text{plug-in}}(t)} \{Y_i - \widehat{m}_i(t)\}, \quad \widehat{\tau}^{(1)}(t, t') = \widehat{\mu}^{(1)}(t) - \widehat{\mu}^{(1)}(t').$$

The superscript (1) marks this as the first-order estimator. If $\widehat{\pi}_i^{\text{plug-in}}(t) = \pi_i(t)$ but $\widehat{m}_i(t) - \mu_i(t) \neq 0$, the outcome-side error exists. $\widehat{\mu}^{(1)}(t)$ uses the weighted term $Y_i - \widehat{m}_i(t)$ to cancel the effect of the outcome-side error. Because the propagated exposure-probability error is zero in this case, the remaining propensity–outcome product error is also zero. If $\widehat{m}_i(t) = \mu_i(t)$ but $\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t) \neq 0$, the propagated exposure-probability error still exists. Because the outcome-side error is zero, the weighted term $Y_i - \widehat{m}_i(t)$ has expectation zero under $\mathbb{E}_D$, and the remaining propensity–outcome product error is also zero. Therefore, no error remains from either the outcome-side error or the remaining propensity–outcome product error in these cases.

When both predictions are inaccurate, the first-order adjustment removes the separate contribution of either the propagated exposure-probability error or the outcome-side error acting alone, leaving the remaining propensity–outcome product error. The resulting $\widehat{\mu}^{(1)}(t) - \mu(t)$ is

$$-\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\{\widehat{m}_i(t) - \mu_i(t)\}. \tag{5}$$

4.2.2 Step 2.2: Second-Order Propagation Correction and Final Estimator.

Step 2.1 leaves the remaining propensity–outcome product error. Because the remaining propensity–outcome product error is generated by the interaction of the propagated exposure-probability error and the outcome-side error, Step 2.2 constructs a correction that combines information associated with the propagated exposure-probability error and the outcome-side error simultaneously.

Let $\widehat{B}_i(t)$ place $\widehat{U}_i$ in the block corresponding to exposure t and zero in all other exposure blocks. For $\mathcal{T} = \{0, 1\}$, $\widehat{B}_i(0) = (\widehat{U}_i^\top, 0)^\top$, $\widehat{B}_i(1) = (0, \widehat{U}_i^\top)^\top$. The node-level contribution to the propensity-side moment $\widehat{q}_i$, which is associated with the propagated exposure-probability error, the node-level contribution to the outcome-side moment $\widehat{p}_i$, which is associated with the outcome-side error, and the corresponding second-moment matrix $\widehat{\Omega}$ are

$$\widehat{q}_i = \sum_{t\in\mathcal{T}} \widehat{B}_i(t) - \frac{\widehat{B}_i(T_i)}{\widehat{\pi}_i^{\text{plug-in}}(T_i)\widehat{h}(T_i)}, \qquad \widehat{p}_j = \widehat{B}_j(T_j)c_{t,t'}(T_j)\left\{Y_j - \widehat{m}_j(T_j)\right\}, \qquad \widehat{\Omega} = \frac{1}{n}\sum_{j=1}^{n}\widehat{B}_j(T_j)\widehat{B}_j(T_j)^\top.$$

Averaging $\widehat{q}_i$ across nodes produces a moment that measures weighting discrepancies along the propagation directions $\widehat{U}_i$. The function $c_{t,t'}(T_j)$ equals 1, -1, or 0 when $T_j = t$, $T_j = t'$, or another exposure, respectively. To express the outcome-side moment in the same propagation directions as the propensity-side moment, the matrix $\widehat{\Omega}^\dagger$ converts the outcome-side moment into coefficients along the propagation directions $\widehat{U}_i$. Thus, the bilinear product $\widehat{q}_i^\top\widehat{\Omega}^\dagger\widehat{p}_j$ combines the propagated exposure-probability error from node i with the outcome-side error from node j along the same propagation directions. Averaging this bilinear product over distinct ordered pairs gives the second-order propagation correction:

$$\widehat{\Delta}^{(2)}_{\text{prop}}(t, t') = \frac{1}{n(n-1)}\sum_{i\neq j}\widehat{q}_i^\top\widehat{\Omega}^\dagger\widehat{p}_j.$$

The correction targets the component of the remaining propensity–outcome product error represented by $\widehat{U}_i$. Excluding $i = j$ removes the diagonal self-product that would otherwise combine both empirical components from the same node. This construction corresponds to a second-order higher-order influence function correction (Robins et al., 2017).

Appendix B shows that, for a constant $C > 0$ independent of n, the remaining propensity–outcome product error after $\widehat{\tau}^{(1)}(t, t')$ is bounded by $C\max_i|\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)|[n^{-1}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t)\}^2]^{1/2}$. After adding $\widehat{\Delta}^{(2)}_{\text{prop}}(t, t')$, the remaining propensity–outcome product error bound becomes $C\max_i|\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)|\inf_{v\in\mathbb{R}^r}[n^{-1}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2]^{1/2}$.

Proposition 4.2. *For every $t \in \mathcal{T}$ with $\min_i \pi_i(t) > 0$, where v indexes coefficients along $\widehat{U}_i$,*

$$\inf_{v\in\mathbb{R}^r}\left[\frac{1}{n}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2\right]^{1/2} \leq \left[\frac{1}{n}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t)\}^2\right]^{1/2}, \tag{6}$$

so multiplying both sides by $\max_i|\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)|$ gives the above bounds.

Proposition 4.2 provides theoretical justification for adding $\widehat{\Delta}^{(2)}_{\text{prop}}(t, t')$ by showing that subtracting $\widehat{U}_i^\top v$ from $\widehat{m}_i(t) - \mu_i(t)$ cannot increase the bound. The proof is given in Appendix B.5.

The final PIN point estimator adds the correction to the first-order estimator,

$$\widehat{\tau}_{\text{PIN}}(t, t') = \widehat{\tau}^{(1)}(t, t') + \widehat{\Delta}^{(2)}_{\text{prop}}(t, t').$$

4.3 Step 3: Uncertainty and the Propagation-Integrated Standard Error

The *Uncertainty* challenge requires the centered node-level score $\widehat{s}_i^{\text{PIN}}(t, t'; \widehat{\pi}^{\text{plug-in}})$ for the final estimator $\widehat{\tau}_{\text{PIN}}(t, t')$, rather than the centered node-level score of the uncorrected plug-in estimator.

Node-Level Variation under the Plug-in Path. Let $\widehat{s}_i^{\mathrm{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}})$ denote the centered node-level score in Appendix C.1.1. The centered node-level score is computed using the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$, the propagation directions $\widehat{U}_i$, and the fitted outcome predictions $\widehat{m}_i(t)$ from Steps 1 and 2. The network variance term is

$$\widehat{\mathrm{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\mathrm{PIN}}(t,t')\} = \sum_{i=1}^{n}\sum_{j=1}^{n} K_{ij}^{+}\widehat{s}_i^{\mathrm{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}})\widehat{s}_j^{\mathrm{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}}).$$

The kernel K^+ is the same kernel used by WLS+.

The Remaining Pairwise Component. The second-order propagation correction $\widehat{\Delta}^{(2)}_{\mathrm{prop}}(t,t')$ is constructed from contributions indexed by distinct node pairs $i \neq j$. The centered node-level score captures the part of the second-order propagation correction attributable to individual nodes, while $\widehat{R}(t,t')$ captures the remaining pairwise component. Appendix C.1.2 gives the entries of $\widehat{R}(t,t')$, with $\widehat{R}_{ii}(t,t') = 0$. The pairwise term is

$$\frac{2}{\{n(n-1)\}^2}\,\mathrm{tr}\{\widehat{R}(t,t')K^{+}\widehat{R}(t,t')K^{+}\}.$$

Propagation-Integrated Standard Error. The reported standard error is

$$\widehat{\mathrm{SE}}_{\mathrm{PIN}}(t,t') = \left[\widehat{\mathrm{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\mathrm{PIN}}(t,t')\}\right]^{1/2} + \left[\frac{2}{\{n(n-1)\}^2}\,\mathrm{tr}\{\widehat{R}(t,t')K^{+}\widehat{R}(t,t')K^{+}\}\right]^{1/2}. \tag{7}$$

Proposition 4.3. *Under Assumptions B.1–B.4, C.1, and C.2. For any fixed contrast t,t' and $\alpha \in (0,1)$,*

$$\liminf_{n\to\infty} \Pr\left\{|\widehat{\tau}_{\mathrm{PIN}}(t,t') - \tau(t,t')| \le \Phi^{-1}(1-\alpha/2)\widehat{\mathrm{SE}}_{\mathrm{PIN}}(t,t')\right\} \ge 1-\alpha, \tag{8}$$

where Φ is the standard normal distribution function.

Proposition 4.3 establishes conservative coverage for confidence intervals. The result follows from Theorem C.2, proved in Appendix C.4.4.

5 Numerical Results

5.1 Simulation

Following Gao and Ding (2025), in each monte carlo replication, we use a synthetic network of (N=1456) nodes and regenerate treatment, exposure, and outcomes while holding the network and pretreatment covariates fixed. Each node has a scalar covariate X_i, latent network positions (ρ_{i1},ρ_{i2}), network degree $\deg_i$, with pretreatment features $H_i = (X_i,\rho_{i1},\rho_{i2},\deg_i)^\top$ used for treatment-propensity estimation. We consider two outcome models: a continuous linear-in-means (LIM) outcome and a binary complex-contagion (CC) outcome. Details are provided in Appendix D.1.

We consider three treatment-propensity shifts. Simulation 1 increases the differences in propensity across nodes, making nodes with originally higher propensity more likely to receive treatment. Simulation 2 keeps the overall distribution of propensity unchanged, constructs a nonlinear score based on each node's own and neighboring H_i, and redistributes high and low propensity values accordingly. Simulation 3 also keeps the overall distribution of propensity unchanged, but redistributes high and low propensity values according to the residualized average of neighboring H_i. Across all settings, the target is $\tau(1,0) = \mu(1) - \mu(0)$, comparing exposure to at least one treated neighbor with exposure to no treated neighbors. Details are provided in Appendix D.3.

The state-of-art baselines are considered to compare with the proposed PIN: WLS+/WLS (Gao and Ding, 2025), EHW (White, 1980), Leung (Leung, 2022), and Kojevnikov (Kojevnikov et al., 2021), Forastiere (Forastiere et al., 2021), NDR (Liu et al., 2023), and netAIPW (Emmenegger et al., 2025).

Simulation 1: Treatment-propensity logit-scale shift.

At the high shift, PIN coverage remains near the 0.95 target: 0.95 for LIM and 0.98 for CC, while other baselines range from 0.38 to 0.87 for LIM and from 0.77 to 0.90 for CC. Even the best competing baseline trails PIN by 8.3 percentage points for LIM and 8.0 points for CC. PIN achieves this coverage with CI lengths 3.82 for LIM and 0.21 for CC, only 0.32 and 0.04 above the longest baseline intervals. Absolute bias falls from 0.84 to 0.21 for LIM, a 75.0% reduction, and from 0.03 to

0.01 for CC, a 66.7% reduction. At the moderate shift, the LIM bias path is 0.791 → 0.268 → 0.085: first-order repair contributes 74.2% of the total repair and second-order integration contributes 25.8%, yielding an 89.2% total reduction and coverage gains of 5.0 and 3.8 percentage points. The CC bias path is 0.029 → 0.011 → 0.007: first- and second-order repair contribute 80.4% and 19.6%, yielding a 76.2% total reduction and coverage gains of 2.8 and 2.5 points (Table 1).

Table 1: Treatment-propensity logit-scale shift (Simulation 1).

Difficulty	Method	Panel A: Linear-in-Means					Panel B: Complex-Contagion				
		Coverage			Average		Coverage			Average	
		Unadj	Add	Sat	|Bias|	CI len.	Unadj	Add	Sat	|Bias|	CI len.
Moderate multipliers 1.3	**Proposed**	**0.96**	**0.95**	**0.95**	**0.18**	**3.66**	**0.98**	**0.98**	**0.98**	**0.02**	**0.19**
	WLS+	0.91	0.83	0.83	0.85	3.46	0.96	0.86	0.86	0.03	0.17
	WLS	0.88	0.78	0.78	0.85	3.22	0.92	0.77	0.75	0.03	0.14
	EHW	0.49	0.34	0.33	0.85	1.48	0.86	0.70	0.70	0.03	0.13
	Leung	0.86	0.75	0.74	0.85	3.02	0.93	0.77	0.76	0.03	0.14
	Kojevnikov	0.86	0.74	0.73	0.85	3.03	0.93	0.78	0.76	0.03	0.14
High multipliers 1.5	**Proposed**	**0.96**	**0.96**	**0.96**	**0.06**	**3.60**	**0.97**	**0.98**	**0.97**	**0.02**	**0.19**
	WLS+	0.90	0.81	0.81	0.92	3.44	0.95	0.81	0.81	0.04	0.17
	WLS	0.88	0.76	0.76	0.92	3.20	0.92	0.67	0.67	0.04	0.14
	EHW	0.51	0.28	0.28	0.92	1.47	0.86	0.62	0.62	0.04	0.12
	Leung	0.85	0.71	0.71	0.92	3.01	0.91	0.69	0.69	0.04	0.14
	Kojevnikov	0.86	0.72	0.72	0.92	3.01	0.91	0.70	0.70	0.04	0.14

Notes: 200 paired draws with 95% confidence intervals. Mean main-MC wall time: 0.03 s/rep Apple M2 chip, 8 workers.

Simulation 2: Fixed-overlap nonlinear rank shift.

At the high shift, PIN coverage is 0.96 for LIM and 0.97 for CC, while other baselines range from 0.27 to 0.74 for LIM and from 0.74 to 0.87 for CC. Even the best competing baseline trails PIN by 22.0 percentage points for LIM and 10.0 points for CC. PIN CI lengths exceed the longest baseline intervals by only 0.06 for LIM and 0.02 for CC, or 1.7% and 11.8%. Bias repair follows the same two-stage pattern as Simulation 1: at the moderate shift, LIM bias decreases from 1.032 to 0.582 and then 0.471, while CC bias decreases from 0.037 to 0.027 and then 0.022. These correspond to total bias reductions of 54.4% and 42.4%, with second-order integration contributing 19.8% and 32.4% of the total repair. Notably, for CC, second-order integration contributes 4.8 percentage points of coverage recovery, exceeding the 2.8 points from first-order repair (Figure 3 and Table 2).

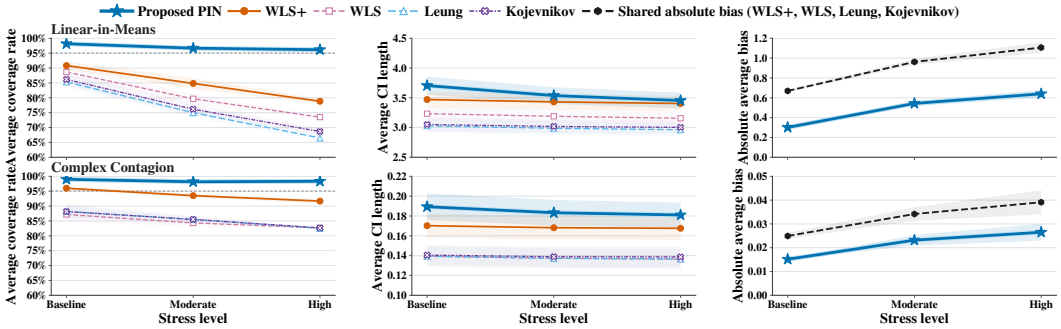


Figure 3: Average coverage rate, confidence interval length, and absolute bias for fixed-overlap nonlinear rank shift (Simulation 2).

Table 2: Nonlinear treatment-propensity rank shift (Simulation 2).

Difficulty	Method	Panel A: Linear-in-Means					Panel B: Complex-Contagion				
		Coverage			Average		Coverage			Average	
		Unadj	Add	Sat	|Bias|	CI len.	Unadj	Add	Sat	|Bias|	CI len.
Moderate rank weights 0.25	**Proposed**	**0.97**	**0.96**	**0.96**	**0.54**	**3.54**	**0.99**	**0.98**	**0.97**	**0.02**	**0.18**
	WLS+	0.90	0.83	0.81	0.96	3.43	0.95	0.93	0.93	0.03	0.17
	WLS	0.85	0.77	0.77	0.96	3.19	0.89	0.82	0.82	0.03	0.14
	EHW	0.42	0.31	0.30	0.96	1.46	0.81	0.81	0.81	0.03	0.13
	Leung	0.80	0.73	0.72	0.96	2.99	0.88	0.84	0.84	0.03	0.14
	Kojevnikov	0.82	0.73	0.73	0.96	3.02	0.89	0.84	0.83	0.03	0.14
High rank weights 0.40	**Proposed**	**0.97**	**0.95**	**0.95**	**0.64**	**3.45**	**0.99**	**0.98**	**0.98**	**0.03**	**0.18**
	WLS+	0.81	0.78	0.78	1.11	3.40	0.94	0.91	0.91	0.04	0.17
	WLS	0.74	0.73	0.72	1.11	3.16	0.83	0.82	0.82	0.04	0.14
	EHW	0.30	0.23	0.23	1.11	1.45	0.72	0.80	0.81	0.04	0.12
	Leung	0.69	0.66	0.65	1.11	2.96	0.82	0.83	0.82	0.04	0.14
	Kojevnikov	0.71	0.68	0.67	1.11	3.00	0.82	0.83	0.82	0.04	0.14

Notes: 200 paired draws with 95% confidence intervals. Mean main-MC wall time: 0.03 s/rep Apple M2 chip, 8 workers.

Simulation 3: Fixed-overlap network-aligned rank shift.

At the high shift, PIN coverage is 0.95 for LIM and 0.98 for CC, while other baselines range from 0.26 to 0.76 for LIM and from 0.73 to 0.89 for CC. The best baseline trails PIN by 19.0 percentage points for LIM and 9.0 points for CC. PIN CI lengths exceed the longest baseline intervals by 0.03 for both outcomes, or 0.9% for LIM and 17.6% for CC. At the moderate shift, LIM bias decreases from 0.934 to 0.405 and then 0.333, while CC bias decreases from 0.039 to 0.026 and then 0.022. Total bias reductions are 64.4% and 44.2%, with second-order integration contributing 12.1% and

23.4% of the total repair. The two stages recover 8.3 and 2.5 percentage points of LIM coverage and 5.7 and 2.7 points of CC coverage, giving total gains of 10.8 and 8.3 points (Table 3).

Table 3: Edge-aligned treatment-propensity rank shift (Simulation 3).

Difficulty	Method	Panel A: Linear-in-Means					Panel B: Complex-Contagion				
		Coverage			Average		Coverage			Average	
		Unadj	Add	Sat	Bias	CI len.	Unadj	Add	Sat	Bias	CI len.
Moderate edge weights 0.15	**Proposed**	**0.98**	**0.96**	**0.96**	**0.48**	**3.56**	**0.99**	**0.98**	**0.98**	**0.02**	**0.19**
	WLS+	0.89	0.83	0.83	0.90	3.42	0.96	0.93	0.93	0.03	0.17
	WLS	0.85	0.78	0.78	0.90	3.17	0.91	0.84	0.83	0.03	0.14
	EHW	0.48	0.35	0.35	0.90	1.46	0.83	0.81	0.80	0.03	0.13
	Leung	0.81	0.73	0.73	0.90	2.98	0.90	0.85	0.85	0.03	0.14
	Kojevnikov	0.83	0.75	0.75	0.90	3.00	0.92	0.85	0.85	0.03	0.14
High edge weights 0.30	**Proposed**	**0.96**	**0.95**	**0.95**	**0.61**	**3.37**	**0.98**	**0.98**	**0.98**	**0.03**	**0.18**
	WLS+	0.77	0.77	0.77	1.12	3.38	0.93	0.92	0.92	0.04	0.17
	WLS	0.70	0.69	0.69	1.12	3.13	0.79	0.81	0.81	0.04	0.14
	EHW	0.30	0.25	0.25	1.12	1.44	0.72	0.80	0.81	0.04	0.12
	Leung	0.67	0.64	0.62	1.12	2.95	0.79	0.82	0.83	0.04	0.14
	Kojevnikov	0.68	0.64	0.63	1.12	2.97	0.81	0.83	0.83	0.04	0.14

Notes: 200 paired draws with 95% confidence intervals. Mean main-MC wall time: 0.05 s/rep Apple M2 chip, 8 workers.

Additional observational-study baselines.

Table F.1 uses the aligned marginal direct effect in Appendix F. Across shifts 0, 1.0, and 1.5, PIN coverage is 1.00, 0.98, and 1.00, while CI length grows from 2.46 to 2.76, or 12.2%. At the high shift, design-based coverage is 0.43–0.82 with bias 0.74, versus PIN coverage 1.00 and bias 0.54. Forastiere attains 1.00 coverage, but length 1215.20 is about 440 times PIN and bias is 1.22. NDR and netAIPW have lengths 1.34 and 1.20 and biases 0.42 and 0.39, but coverage is 0.75 and 0.72. Only PIN combines coverage at least 0.95 with length below 3.

5.2 Real-Network Semi-Synthetic Study

Following Gao and Ding (2025), we use the observed school network from the five largest treated schools in Paluck et al. (2016), containing $N = 3306$ students. In each Monte Carlo replication, we regenerate treatment, exposure, and synthetic outcomes while holding the observed network and pretreatment features fixed. Each student has an outcome-adjustment vector X_i containing gender and grade, and a pretreatment feature vector H_i containing gender, grade, school membership, randomization-block membership, and network degree. We consider both one-dimensional and their two-dimensional exposure mapping. We consider a treatment-propensity shift that uses a standardized score S_i to increasingly differentiate treatment propensities across students while preserving the average treatment propensity among eligible students, so that students with larger S_i receive larger treatment propensities and students with smaller S_i receive smaller treatment propensities as the shift parameter increases. Details are provided in Appendix E.

At the high shift, PIN averages 0.97 coverage under both mappings. WLS+ averages 0.94 under the one-dimensional mapping and 0.92 under the two-dimensional mapping (Figure 4). Absolute bias is 0.01 versus 0.03 and 0.02 versus 0.03, while CI length is 0.56 versus 0.41 and 0.64 versus 0.37.

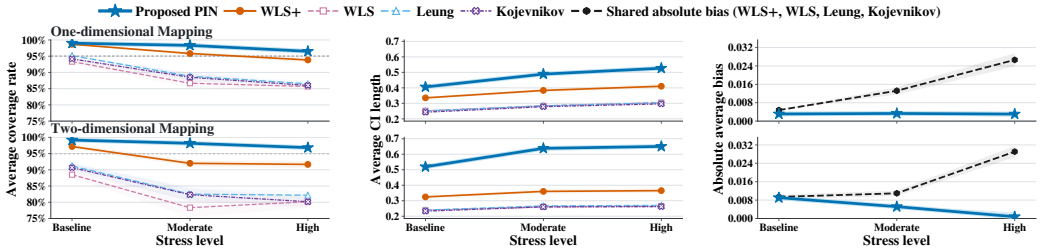


Figure 4: Direct effects by exposure mapping (real-network semi-synthetic study).

6 Discussion

We introduced Propagation-Integrated Inference (PIN) to address the bias and uncertainty quantification caused by the propagation of the treatment-propensity error through the network in observational studies. PIN integrates the adjustment for point estimation with uncertainty quantification, rather than adjusting standard errors alone. Our theory shows how propagation directions provide an observable basis for adjusting point estimation and how the propagation-integrated standard error supports conservative coverage under the corresponding conditions. Future work may investigate how to select pretreatment features to construct more useful propagation directions, extend PIN to broader forms of network dependence, and evaluate the applicability of PIN in real-world observational studies.

AI USE STATEMENT

In this work, a generative AI assistant was used to check compliance with the ICLR 2027 template, edit and condense the manuscript language, and refine the LaTeX formatting, including the presentation of cited baseline methods and reported results.

REPRODUCIBILITY STATEMENT

The PIN implementation is available in our anonymous code repository.

ETHICS STATEMENT

The simulations use synthetic treatments and outcomes. The real-network semi-synthetic study reuses an existing school-network structure and pretreatment covariates, introduces no new participant interaction, and does not analyze the observed study outcome.

REFERENCES

Qianyi Chen, Anpeng Wu, Bo Li, Lu Deng, and Yong Wang. Journey to the centre of cluster: Harnessing interior nodes for A/B testing under network interference. In *The Fourteenth International Conference on Learning Representations*, 2026.

Shishir Adhikari, Sourav Medya, and Elena Zheleva. Learning exposure mapping functions for inferring heterogeneous peer effects. In *The Fourteenth International Conference on Learning Representations*, 2026.

Jin Zhu, Jingyi Li, Hongyi Zhou, Yinan Lin, Zhenhua Lin, and Chengchun Shi. Balancing interference and correlation in spatial experimental designs: A causal graph cut approach. In *Proceedings of the 42nd International Conference on Machine Learning*, pages 79918–79937, 2025.

Mohsen Bayati, Yuwei Luo, William Overman, Sadegh Shirani, and Ruoxuan Xiong. Higher-order causal message passing for experimentation with complex interference. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Zhiheng Zhang and Zichen Wang. Online experimental design with estimation-regret trade-off under network interference. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Qiang Huang and Jin Tian. Modeling interference for treatment effect estimation in network dynamic environment. In *The Fourteenth International Conference on Learning Representations*, 2026.

M. Alan Brookhart, Sebastian Schneeweiss, Kenneth J. Rothman, Robert J. Glynn, Jerry Avorn, and Til Stürmer. Variable selection for propensity score models. *American Journal of Epidemiology*, 163(12):1149–1156, 2006. doi:10.1093/aje/kwj149.

Peter M. Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017. doi:10.1214/16-AOAS1005.

Michael P. Leung. Causal inference under approximate neighborhood interference. *Econometrica*, 90(1):267–293, 2022. doi:10.3982/ECTA17841.

Mengsi Gao and Peng Ding. Causal inference in network experiments: regression-based analysis and design-based properties. *Journal of Econometrics*, 252:106119, 2025. doi:10.1016/j.jeconom.2025.106119.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980. doi:10.2307/1912934.

Qingyuan Zhao. Covariate balancing propensity score by tailored loss functions. *The Annals of Statistics*, 47(2):965–993, 2019. doi:10.1214/18-AOS1698.

Denis Kojevnikov, Vadim Marmer, and Kyungchul Song. Limit theorems for network dependent random variables. *Journal of Econometrics*, 222(2):882–908, 2021. doi:10.1016/j.jeconom.2020.05.019.

Laura Forastiere, Edoardo M. Airoldi, and Fabrizia Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534):901–918, 2021. doi:10.1080/01621459.2020.1768100.

Elizabeth L. Ogburn, Oleg Sofrygin, Iván Díaz, and Mark J. van der Laan. Causal inference for social network data. *Journal of the American Statistical Association*, 119(545):597–611, 2024. doi:10.1080/01621459.2022.2131557.

Jie Liu, Fangjuan Ye, and Yang Yang. Nonparametric doubly robust estimation of causal effect on networks in observational studies. *Stat*, 12(1):e549, 2023. doi:10.1002/sta4.549.

Corinne Emmenegger, Meta-Lina Spohn, Timon Elmer, and Peter Bühlmann. Treatment effect estimation with observational network data using machine learning. *Journal of Causal Inference*, 13(1):20230082, 2025. doi:10.1515/jci-2023-0082.

Elizabeth Levy Paluck, Hana Shepherd, and Peter M. Aronow. Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3):566–571, 2016. doi:10.1073/pnas.1514483113.

Elizabeth Levy Paluck, Hana R. Shepherd, and Peter Aronow. Changing climates of conflict: a social network experiment in 56 schools, New Jersey, 2012–2013. Inter-university Consortium for Political and Social Research [distributor], Version V2, 2020. doi:10.3886/ICPSR37070.v2.

James M. Robins, Lingling Li, Rajarshi Mukherjee, Eric Tchetgen Tchetgen, and Aad W. van der Vaart. Higher order estimating equations for high-dimensional models. *The Annals of Statistics*, 45(5):1951–1987, 2017. doi:10.1214/16-AOS1515.

David A. Hirshberg and Stefan Wager. Augmented minimax linear estimation. *The Annals of Statistics*, 49(6):3206–3227, 2021. doi:10.1214/21-AOS2080.

Chencheng Cai, Xu Zhang, and Edoardo M. Airoldi. Independent-set design of experiments for estimating treatment and spillover effects under network interference. In *The Twelfth International Conference on Learning Representations*, 2024. https://openreview.net/forum?id=w50MQ9Vfty.

Anpeng Wu, Haiyi Qiu, Zhengming Chen, Zijian Li, Ruoxuan Xiong, Fei Wu, and Kun Zhang. Causal graph transformer for treatment effect estimation under unknown interference. In *The Thirteenth International Conference on Learning Representations*, 2025. https://openreview.net/forum?id=foQ4AeEGG7.

Chris Hays and Manish Raghavan. Double machine learning for causal inference under shared-state interference. In *Proceedings of the 42nd International Conference on Machine Learning*, pages 22318–22350, 2025.

A Construct the Propagation Directions: Derivation, Theory and Proof

A.1 Derivation

Fix an outer fold $\mathcal{I}_\ell$. The inner out-of-fold treatment propensities $\widetilde{e}_k$, $k \in \mathcal{J}_\ell$, are constructed without using D_k to fit $\widetilde{e}_k$. The plug-in treatment propensities $\widehat{e}_i^{\text{plug-in}}$, $i \in \mathcal{I}_\ell$, come from the propensity model fitted on the outer-training set $\mathcal{J}_\ell$. Assume that $\widetilde{e}_k$ and $\widehat{e}_i^{\text{plug-in}}$ lie in $(0, 1)$, so $\operatorname{logit} \widetilde{e}_k$ and $\operatorname{logit} \widehat{e}_i^{\text{plug-in}}$ are finite. The goal is to estimate the fold-specific correction coefficient $\widehat{\phi}_\ell$ from the treatment-prediction residuals $D_k - \widetilde{e}_k$ and apply $H_i^\top \widehat{\phi}_\ell$ to $\operatorname{logit} \widehat{e}_i^{\text{plug-in}}$.

The average

$$\frac{1}{|\mathcal{J}_\ell|} \sum_{k \in \mathcal{J}_\ell} H_k (D_k - \widetilde{e}_k)$$

compares the treatment indicators D_k with the inner out-of-fold treatment propensities $\widetilde{e}_k$ after weighting by each component of the pretreatment features H_k. PIN uses the one-step Newton update

to match the averages of $H_k D_k$ and $H_k \operatorname{logit}^{-1}\{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi\}$ over $\mathcal{J}_\ell$, rather than matching D_k at every node. Adding $H_k^\top \phi$ to $\operatorname{logit} \widetilde{e}_k$ gives the treatment propensity $\operatorname{logit}^{-1}\{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi\}$.

Holding $\{D_k, \widetilde{e}_k, H_k : k \in \mathcal{J}_\ell\}$ fixed, consider the average Bernoulli log-likelihood

$$\frac{1}{|\mathcal{J}_\ell|} \sum_{k \in \mathcal{J}_\ell} \Big[D_k \{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi\} - \log\{1 + \exp(\operatorname{logit} \widetilde{e}_k + H_k^\top \phi)\} \Big]. \tag{A.1}$$

Differentiating (A.1) with respect to ϕ and setting the derivative to zero gives

$$\frac{1}{|\mathcal{J}_\ell|} \sum_{k \in \mathcal{J}_\ell} H_k \left[D_k - \operatorname{logit}^{-1}\big\{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi\big\} \right] = 0. \tag{A.2}$$

The score equation (A.2) expresses the desired agreement between the averages of $H_k D_k$ and $H_k \operatorname{logit}^{-1}\{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi\}$ over $\mathcal{J}_\ell$. At $\phi = 0$, the derivative of (A.1) is $|\mathcal{J}_\ell|^{-1} \sum_{k \in \mathcal{J}_\ell} H_k (D_k - \widetilde{e}_k)$, the average formed from the treatment-prediction residuals $D_k - \widetilde{e}_k$ in Step 1.

PIN applies the one-step Newton update at $\phi = 0$ rather than solving the nonlinear score equation (A.2) to convergence. The derivative of $\operatorname{logit}^{-1}\{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi\}$ at $\phi = 0$ is

$$\frac{\partial}{\partial \phi} \operatorname{logit}^{-1}\big\{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi\big\}\Big|_{\phi=0} = \widetilde{e}_k(1 - \widetilde{e}_k) H_k.$$

Consequently, expanding the score equation around $\phi = 0$ gives

$$\frac{1}{|\mathcal{J}_\ell|} \sum_{k \in \mathcal{J}_\ell} H_k \left[D_k - \operatorname{logit}^{-1}\big\{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi\big\} \right] \approx \frac{1}{|\mathcal{J}_\ell|} \sum_{k \in \mathcal{J}_\ell} H_k (D_k - \widetilde{e}_k) - \left[\frac{1}{|\mathcal{J}_\ell|} \sum_{k \in \mathcal{J}_\ell} H_k H_k^\top \widetilde{e}_k (1 - \widetilde{e}_k) \right] \phi. \tag{A.3}$$

The matrix multiplying ϕ describes how the average of $H_k[D_k - \operatorname{logit}^{-1}\{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi\}]$ over $\mathcal{J}_\ell$ changes with ϕ. Setting (A.3) to zero gives the fold-specific correction coefficient $\widehat{\phi}_\ell$:

$$\widehat{\phi}_\ell = \left[\frac{1}{|\mathcal{J}_\ell|} \sum_{k \in \mathcal{J}_\ell} H_k H_k^\top \widetilde{e}_k (1 - \widetilde{e}_k) \right]^\dagger \frac{1}{|\mathcal{J}_\ell|} \sum_{k \in \mathcal{J}_\ell} H_k (D_k - \widetilde{e}_k). \tag{A.4}$$

The symbol † denotes the Moore–Penrose inverse and coincides with the ordinary inverse for a nonsingular matrix. The fold-specific correction coefficient $\widehat{\phi}_\ell$ sets the linearized score in (A.3) to zero.

For each held-out node $i \in \mathcal{I}_\ell$, PIN applies the fold-specific correction coefficient $\widehat{\phi}_\ell$ through

$$\operatorname{logit} \widehat{e}_i^{\,\text{complete}} = \operatorname{logit} \widehat{e}_i^{\,\text{plug-in}} + H_i^\top \widehat{\phi}_\ell, \qquad i \in \mathcal{I}_\ell. \tag{A.5}$$

The inner out-of-fold treatment propensities $\widetilde{e}_k$ are used to estimate the fold-specific correction coefficient $\widehat{\phi}_\ell$ within the outer-training set $\mathcal{J}_\ell$. The plug-in treatment propensities $\widehat{e}_i^{\,\text{plug-in}}$ receive $H_i^\top \widehat{\phi}_\ell$ on the logit scale. The treatment indicators $\{D_i : i \in \mathcal{I}_\ell\}$ and all outcomes are excluded from the computation of $\widehat{\phi}_\ell$.

Applying $\operatorname{logit}^{-1}$ in (A.5) and assembling $\widehat{e}_i^{\,\text{complete}}$ across the outer folds yields the score-completed path $\widehat{e}^{\,\text{complete}}$. The exposure mapping then gives the score-completed exposure probabilities

$$\widehat{\pi}_i^{\,\text{complete}}(t) = \pi_i(t; \widehat{e}^{\,\text{complete}}, A).$$

Comparing the score-completed exposure probabilities $\widehat{\pi}_i^{\,\text{complete}}(t)$ and the plug-in exposure probabilities $\widehat{\pi}_i^{\,\text{plug-in}}(t)$ gives the propagation directions $d_i^{\,\text{complete}}(t)$ in Step 1:

$$d_i^{\,\text{complete}}(t) = \log \widehat{\pi}_i^{\,\text{complete}}(t) - \log \widehat{\pi}_i^{\,\text{plug-in}}(t).$$

A.2 Conditions

The conditions below concern the plug-in treatment propensities $\widehat{e}_i^{\,\text{plug-in}}$, the inner out-of-fold treatment propensities $\widetilde{e}_k$, and the propagation directions $d_i^{\,\text{complete}}(t) = \log \widehat{\pi}_i^{\,\text{complete}}(t) - \log \widehat{\pi}_i^{\,\text{plug-in}}(t)$ in Step 1. For each outer fold $\mathcal{I}_\ell$, $\widetilde{e}_k$ denotes the inner out-of-fold treatment propensity constructed within the outer-training set $\mathcal{J}_\ell$. The dependence of $\widetilde{e}_k$ on ℓ is suppressed, as in Step 1.

Assumption A.1. Conditional on the fixed network A and pretreatment features $\{H_i\}_{i=1}^n$, the treatment indicators $D_1, \ldots, D_n$ are independent Bernoulli variables with true treatment propensities $e_i \in (0, 1)$. The outer folds $\mathcal{I}_1, \ldots, \mathcal{I}_L$ partition the analysis nodes $\mathcal{N}_n$, and the outer-training set is $\mathcal{J}_\ell = \mathcal{N}_n \setminus \mathcal{I}_\ell$. The fold construction is independent of D and uses no outcomes. The number of

outer folds L and the dimension of H_i are fixed. For constants $c_0, C > 0$ independent of n,

$$\min_{1\le\ell\le L} |\mathcal{J}_\ell| \ge c_0 n, \qquad \max_{1\le i\le n} \|H_i\| \le C.$$

The inner out-of-fold treatment propensities $\widetilde{e}_k$ and the plug-in treatment propensities $\widehat{e}_i^{\text{plug-in}}$ lie in $(0,1)$. The exposure mapping is $T_i = \mathbf{1}\{\sum_j A_{ij} D_j > 0\}$, with $\mathcal{T} = \{0,1\}$, and every analyzed node has at least one neighbor. Write

$$\Delta_n = \max_{1\le i\le n} \sum_{j=1}^{n} A_{ij} \ge 1.$$

Here Δ_n is the maximum degree of the network A, that is, the largest number of neighbors of any node.

Assumption A.2. There exist deterministic sequences $a_n > 0$ and $\rho_n > 0$ tending to zero and possibly random coefficient vectors $\phi_\ell^\star$ such that

$$\max_{1\le\ell\le L} \|\phi_\ell^\star\| = O_p(a_n). \tag{A.6}$$

For each outer fold $\mathcal{I}_\ell$, the inner out-of-fold treatment propensities $\widetilde{e}_k$ satisfy

$$\begin{aligned} \operatorname{logit} e_k - \operatorname{logit} \widetilde{e}_k &= H_k^\top \phi_\ell^\star + r_{k\ell}^{\text{in}}, \qquad k \in \mathcal{J}_\ell, \\ \max_{1\le\ell\le L} \max_{k\in\mathcal{J}_\ell} |r_{k\ell}^{\text{in}}| &= O_p(\rho_n). \end{aligned} \tag{A.7}$$

The plug-in treatment propensities $\widehat{e}_i^{\text{plug-in}}$ satisfy

$$\begin{aligned} \operatorname{logit} e_i - \operatorname{logit} \widehat{e}_i^{\text{plug-in}} &= H_i^\top \phi_\ell^\star + r_{i\ell}^{\text{out}}, \qquad i \in \mathcal{I}_\ell, \\ \max_{1\le\ell\le L} \max_{i\in\mathcal{I}_\ell} |r_{i\ell}^{\text{out}}| &= O_p(\rho_n). \end{aligned} \tag{A.8}$$

The same coefficient vector $\phi_\ell^\star$ appears in $H_k^\top \phi_\ell^\star$ for $k \in \mathcal{J}_\ell$ and $H_i^\top \phi_\ell^\star$ for $i \in \mathcal{I}_\ell$.

The remainders $r_{k\ell}^{\text{in}}$ and $r_{i\ell}^{\text{out}}$ contain the parts of $\operatorname{logit} e_k - \operatorname{logit} \widetilde{e}_k$ and $\operatorname{logit} e_i - \operatorname{logit} \widehat{e}_i^{\text{plug-in}}$ not represented by $H_k^\top \phi_\ell^\star$ and $H_i^\top \phi_\ell^\star$, respectively.

Assumption A.3. Define

$$\widehat{G}_\ell = \frac{1}{|\mathcal{J}_\ell|} \sum_{k\in\mathcal{J}_\ell} H_k H_k^\top \widetilde{e}_k(1-\widetilde{e}_k). \tag{A.9}$$

For a constant $c_1 > 0$ independent of n,

$$\Pr\left\{ \min_{1\le\ell\le L} \lambda_{\min}(\widehat{G}_\ell) \ge c_1 \right\} \longrightarrow 1.$$

Here $\lambda_{\min}(\widehat{G}_\ell)$ denotes the smallest eigenvalue of $\widehat{G}_\ell$. A positive lower bound on $\lambda_{\min}(\widehat{G}_\ell)$ ensures that $\phi_\ell^\star$ in $H_i^\top \phi_\ell^\star$ is sufficiently identified in every direction and prevents amplification by $\widehat{G}_\ell^\dagger$ in the one-step Newton update for the fold-specific correction coefficient $\widehat{\phi}_\ell$.

A.3 Theoretical Result

Recall the fold-specific correction coefficient $\widehat{\phi}_\ell$ obtained by the one-step Newton update and the score-completed path $\widehat{e}^{\text{complete}}$ from Step 1:

$$\widehat{\phi}_\ell = \widehat{G}_\ell^\dagger \frac{1}{|\mathcal{J}_\ell|} \sum_{k\in\mathcal{J}_\ell} H_k (D_k - \widetilde{e}_k), \tag{A.10}$$

$$\operatorname{logit} \widehat{e}_i^{\text{complete}} = \operatorname{logit} \widehat{e}_i^{\text{plug-in}} + H_i^\top \widehat{\phi}_\ell, \qquad i \in \mathcal{I}_\ell. \tag{A.11}$$

The plug-in path $\widehat{e}^{\text{plug-in}}$ and the score-completed path $\widehat{e}^{\text{complete}}$ assemble $\widehat{e}_i^{\text{plug-in}}$ and $\widehat{e}_i^{\text{complete}}$ across the outer folds. The exposure mapping gives the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t) = \pi_i(t; \widehat{e}^{\text{plug-in}}, A)$ and the score-completed exposure probabilities $\widehat{\pi}_i^{\text{complete}}(t) = \pi_i(t; \widehat{e}^{\text{complete}}, A)$. The propagation directions $d_i^{\text{complete}}(t) = \log \widehat{\pi}_i^{\text{complete}}(t) - \log \widehat{\pi}_i^{\text{plug-in}}(t)$ are defined in Step 1.

Proposition A.1. *Under Assumptions A.1–A.3, the fold-specific correction coefficient $\widehat{\phi}_\ell$ satisfies*

$$\max_{1\le\ell\le L} \|\widehat{\phi}_\ell - \phi_\ell^\star\| = O_p(a_n^2 + \rho_n + n^{-1/2}), \tag{A.12}$$

and the score-completed path $\widehat{e}^{\text{complete}}$ satisfies

$$\max_{1\le i\le n} \left| \operatorname{logit} \widehat{e}_i^{\text{complete}} - \operatorname{logit} e_i \right| = O_p(a_n^2 + \rho_n + n^{-1/2}). \tag{A.13}$$

Consequently, the propagation directions $d_i^{\text{complete}}(t)$ *satisfy*

$$\max_{1\le i\le n,\, t\in\{0,1\}} \left| d_i^{\text{complete}}(t) + \log \widehat{\pi}_i^{\text{plug-in}}(t) - \log \pi_i(t) \right| = O_p\Big\{ \Delta_n \left(a_n^2 + \rho_n + n^{-1/2} \right) \Big\}. \tag{A.14}$$

If $\rho_n + n^{-1/2} = o(a_n)$*, the left-hand side of* (A.14) *is* $o_p(\Delta_n a_n)$.

A.4 Proof

Proof of Proposition A.1. The bound for $\widehat{\phi}_\ell - \phi_\ell^\star$. For every $k \in \mathcal{J}_\ell$,

$$D_k - \widetilde{e}_k = (D_k - e_k) + (e_k - \widetilde{e}_k). \tag{A.15}$$

Conditional on A, $\{H_i\}_{i=1}^n$, and $\{\mathcal{I}_\ell\}_{\ell=1}^L$, Assumption A.1 gives independence of $D_1, \dots, D_n$, with mean equal to the treatment propensity e_k and variance $e_k(1-e_k)$ for D_k. Taking the expectation conditional on A, $\{H_i\}_{i=1}^n$, and $\{\mathcal{I}_\ell\}_{\ell=1}^L$ therefore gives

$$\mathbb{E}_D\left[\left\| \frac{1}{|\mathcal{J}_\ell|} \sum_{k\in\mathcal{J}_\ell} H_k(D_k - e_k) \right\|^2 \right] = \frac{1}{|\mathcal{J}_\ell|^2} \sum_{k\in\mathcal{J}_\ell} \|H_k\|^2 e_k(1-e_k) \le \frac{C^2}{4|\mathcal{J}_\ell|} \le \frac{C^2}{4c_0 n} = O(n^{-1}).$$

The assignment expectation $\mathbb{E}_D$ does not condition on the inner out-of-fold treatment propensities $\{\widetilde{e}_k : k \in \mathcal{J}_\ell\}$ or the plug-in path $\widehat{e}^{\text{plug-in}}$. Since L is fixed and $\min_{1\le\ell\le L} |\mathcal{J}_\ell| \ge c_0 n$, applying Markov's inequality to $\| |\mathcal{J}_\ell|^{-1} \sum_{k\in\mathcal{J}_\ell} H_k(D_k - e_k) \|^2$ and summing the probability bounds over $\ell = 1, \dots, L$ gives

$$\max_{1\le\ell\le L} \left\| \frac{1}{|\mathcal{J}_\ell|} \sum_{k\in\mathcal{J}_\ell} H_k(D_k - e_k) \right\| = O_p(n^{-1/2}). \tag{A.16}$$

Assumption A.2 gives

$$\operatorname{logit} e_k - \operatorname{logit} \widetilde{e}_k = H_k^\top \phi_\ell^\star + r_{k\ell}^{\text{in}}, \qquad k \in \mathcal{J}_\ell.$$

The first derivative of $\operatorname{logit}^{-1}$ is bounded by $1/4$, and the absolute value of the second derivative is bounded by $1/4$. Expanding $\operatorname{logit}^{-1}\{\operatorname{logit} \widetilde{e}_k + H_k^\top \phi_\ell^\star + r_{k\ell}^{\text{in}}\}$ at $\operatorname{logit} \widetilde{e}_k$ therefore gives

$$\left| e_k - \widetilde{e}_k - \widetilde{e}_k(1-\widetilde{e}_k)\left(H_k^\top \phi_\ell^\star + r_{k\ell}^{\text{in}} \right) \right| \le \frac{1}{8} \left(H_k^\top \phi_\ell^\star + r_{k\ell}^{\text{in}} \right)^2.$$

Since $\widetilde{e}_k(1-\widetilde{e}_k) \le 1/4$, adding $\widetilde{e}_k(1-\widetilde{e}_k) r_{k\ell}^{\text{in}}$ inside the absolute value and applying the triangle inequality gives

$$\left| e_k - \widetilde{e}_k - \widetilde{e}_k(1-\widetilde{e}_k) H_k^\top \phi_\ell^\star \right| \le \frac{1}{8} \left(H_k^\top \phi_\ell^\star + r_{k\ell}^{\text{in}} \right)^2 + \frac{1}{4} |r_{k\ell}^{\text{in}}|.$$

Assumption A.2 provides $\max_{1\le\ell\le L} \|\phi_\ell^\star\| = O_p(a_n)$ and $\max_{1\le\ell\le L} \max_{k\in\mathcal{J}_\ell} |r_{k\ell}^{\text{in}}| = O_p(\rho_n)$. Together with $\max_{1\le k\le n} \|H_k\| \le C$, the bounds on $\|\phi_\ell^\star\|$ and $|r_{k\ell}^{\text{in}}|$ imply

$$\max_{1\le\ell\le L} \max_{k\in\mathcal{J}_\ell} |\operatorname{logit} e_k - \operatorname{logit} \widetilde{e}_k| = \max_{1\le\ell\le L} \max_{k\in\mathcal{J}_\ell} \left| H_k^\top \phi_\ell^\star + r_{k\ell}^{\text{in}} \right| = O_p(a_n + \rho_n).$$

Taking the maximum over ℓ and k in the preceding inequality, together with

$$\max_{1\le\ell\le L} \max_{k\in\mathcal{J}_\ell} \left| H_k^\top \phi_\ell^\star + r_{k\ell}^{\text{in}} \right| = O_p(a_n + \rho_n)$$

and

$$\max_{1\le\ell\le L} \max_{k\in\mathcal{J}_\ell} |r_{k\ell}^{\text{in}}| = O_p(\rho_n),$$

gives

$$\max_{1\le\ell\le L} \max_{k\in\mathcal{J}_\ell} \left| e_k - \widetilde{e}_k - \widetilde{e}_k(1-\widetilde{e}_k) H_k^\top \phi_\ell^\star \right| = O_p\left((a_n + \rho_n)^2 + \rho_n \right).$$

Because $a_n \to 0$ and $\rho_n \to 0$, for sufficiently large n,

$$(a_n + \rho_n)^2 + \rho_n = a_n^2 + 2a_n\rho_n + \rho_n^2 + \rho_n \le a_n^2 + 4\rho_n = O(a_n^2 + \rho_n).$$

Consequently,

$$\max_{1\le\ell\le L} \max_{k\in\mathcal{J}_\ell} \left| e_k - \widetilde{e}_k - \widetilde{e}_k(1-\widetilde{e}_k) H_k^\top \phi_\ell^\star \right| = O_p(a_n^2 + \rho_n). \tag{A.17}$$

Multiplying (A.15) by H_k, averaging over $\mathcal{J}_\ell$, and using (A.17) with $\max_{1\le k\le n} \|H_k\| \le C$ gives

$$\frac{1}{|\mathcal{J}_\ell|} \sum_{k\in\mathcal{J}_\ell} H_k(D_k - \widetilde{e}_k) = \widehat{G}_\ell \phi_\ell^\star + \frac{1}{|\mathcal{J}_\ell|} \sum_{k\in\mathcal{J}_\ell} H_k(D_k - e_k) + O_p(a_n^2 + \rho_n), \tag{A.18}$$

uniformly over $\ell = 1, \dots, L$. On the event $\min_{1\le\ell\le L} \lambda_{\min}(\widehat{G}_\ell) \ge c_1$, $\widehat{G}_\ell^\dagger = \widehat{G}_\ell^{-1}$ and $\|\widehat{G}_\ell^{-1}\|_{\text{op}} \le c_1^{-1}$ for every ℓ. Substituting (A.18) into (A.10) therefore gives

$$\widehat{\phi}_\ell - \phi_\ell^\star = \widehat{G}_\ell^{-1} \frac{1}{|\mathcal{J}_\ell|} \sum_{k\in\mathcal{J}_\ell} H_k(D_k - e_k) + O_p(a_n^2 + \rho_n),$$

uniformly over $\ell = 1, \dots, L$ on $\{\min_{1\le\ell\le L} \lambda_{\min}(\widehat{G}_\ell) \ge c_1\}$. Assumption A.3 ensures that $\Pr\{\min_{1\le\ell\le L} \lambda_{\min}(\widehat{G}_\ell) \ge c_1\} \to 1$. Combining $\|\widehat{G}_\ell^{-1}\|_{\mathrm{op}} \le c_1^{-1}$ with (A.16) thus gives

$$\max_{1\le\ell\le L} \|\widehat{\phi}_\ell - \phi_\ell^\star\| = O_p(a_n^2 + \rho_n + n^{-1/2}),$$

proving (A.12).

The bound for $\operatorname{logit} \widehat{e}_i^{\,\mathrm{complete}} - \operatorname{logit} e_i$**.** For $i \in \mathcal{I}_\ell$, Assumption A.2 gives

$$\operatorname{logit} e_i - \operatorname{logit} \widehat{e}_i^{\,\text{plug-in}} = H_i^\top \phi_\ell^\star + r_{i\ell}^{\mathrm{out}}.$$

Substituting $\operatorname{logit} \widehat{e}_i^{\,\text{plug-in}} - \operatorname{logit} e_i = -H_i^\top \phi_\ell^\star - r_{i\ell}^{\mathrm{out}}$ into (A.11) gives

$$\operatorname{logit} \widehat{e}_i^{\,\mathrm{complete}} - \operatorname{logit} e_i = H_i^\top (\widehat{\phi}_\ell - \phi_\ell^\star) - r_{i\ell}^{\mathrm{out}}, \qquad i \in \mathcal{I}_\ell. \tag{A.19}$$

Therefore,

$$\left|\operatorname{logit} \widehat{e}_i^{\,\mathrm{complete}} - \operatorname{logit} e_i\right| \le \|H_i\| \|\widehat{\phi}_\ell - \phi_\ell^\star\| + |r_{i\ell}^{\mathrm{out}}|.$$

Using $\max_{1\le i\le n} \|H_i\| \le C$, $\max_{1\le\ell\le L} \max_{i\in\mathcal{I}_\ell} |r_{i\ell}^{\mathrm{out}}| = O_p(\rho_n)$, and (A.12) gives

$$\max_{1\le i\le n} \left|\operatorname{logit} \widehat{e}_i^{\,\mathrm{complete}} - \operatorname{logit} e_i\right| = O_p(a_n^2 + \rho_n + n^{-1/2}),$$

proving (A.13).

The bounds for $\log \widehat{\pi}_i^{\,\mathrm{complete}}(t) - \log \pi_i(t)$**.** Under $T_i = \mathbf{1}\{\sum_j A_{ij} D_j > 0\}$, Assumption A.1 gives $\pi_i(0; e, A) = \prod_{j:A_{ij}=1} (1 - e_j)$. Consequently,

$$\log \pi_i(0; e, A) = -\sum_{j:A_{ij}=1} \log\{1 + \exp(\operatorname{logit} e_j)\}.$$

Differentiating $\log \pi_i(0; e, A)$ with respect to $\operatorname{logit} e_j$ gives $-A_{ij} e_j$, so

$$\sum_{j=1}^n \left|\frac{\partial \log \pi_i(0; e, A)}{\partial(\operatorname{logit} e_j)}\right| = \sum_{j:A_{ij}=1} e_j \le \sum_{j=1}^n A_{ij}.$$

The inequality $\sum_{j:A_{ij}=1} e_j \le \sum_{j=1}^n A_{ij}$ holds for every $e \in (0,1)^n$, and $\sum_{j=1}^n A_{ij}$ does not depend on e. Integrating with respect to the logits along the line segment from $(\operatorname{logit} e_j)_{j=1}^n$ to $(\operatorname{logit} \widehat{e}_j^{\,\mathrm{complete}})_{j=1}^n$ therefore gives

$$\left|\log \widehat{\pi}_i^{\,\mathrm{complete}}(0) - \log \pi_i(0)\right| \le \left(\sum_{j=1}^n A_{ij}\right) \max_{1\le j\le n} \left|\operatorname{logit} \widehat{e}_j^{\,\mathrm{complete}} - \operatorname{logit} e_j\right|. \tag{A.20}$$

For $t = 1$,

$$\pi_i(1; e, A) = 1 - \prod_{j:A_{ij}=1} (1 - e_j) = 1 - \exp\left\{\sum_{j:A_{ij}=1} \log(1 - e_j)\right\}.$$

Every analyzed node has at least one neighbor, and $e_j \in (0,1)$ implies $-\sum_{j:A_{ij}=1} \log(1 - e_j) > 0$. Differentiating $\log \pi_i(1; e, A)$ with respect to $\operatorname{logit} e_j$ and summing the absolute derivatives gives

$$\sum_{j=1}^n \left|\frac{\partial \log \pi_i(1; e, A)}{\partial(\operatorname{logit} e_j)}\right| = \frac{\sum_{j:A_{ij}=1} e_j}{\exp\{-\sum_{j:A_{ij}=1} \log(1 - e_j)\} - 1} \le \frac{-\sum_{j:A_{ij}=1} \log(1 - e_j)}{\exp\{-\sum_{j:A_{ij}=1} \log(1 - e_j)\} - 1} \le 1.$$

The first inequality follows from $e_j \le -\log(1 - e_j)$, and the final inequality follows from

$$-\sum_{j:A_{ij}=1} \log(1 - e_j) \le \exp\left\{-\sum_{j:A_{ij}=1} \log(1 - e_j)\right\} - 1.$$

The bound $\sum_{j=1}^n |\partial \log \pi_i(1; e, A)/\partial(\operatorname{logit} e_j)| \le 1$ holds for every $e \in (0,1)^n$. Integrating with respect to the logits along the line segment from $(\operatorname{logit} e_j)_{j=1}^n$ to $(\operatorname{logit} \widehat{e}_j^{\,\mathrm{complete}})_{j=1}^n$ gives

$$\left|\log \widehat{\pi}_i^{\,\mathrm{complete}}(1) - \log \pi_i(1)\right| \le \max_{1\le j\le n} \left|\operatorname{logit} \widehat{e}_j^{\,\mathrm{complete}} - \operatorname{logit} e_j\right|. \tag{A.21}$$

The definition of the propagation directions $d_i^{\,\mathrm{complete}}(t) = \log \widehat{\pi}_i^{\,\mathrm{complete}}(t) - \log \widehat{\pi}_i^{\,\text{plug-in}}(t)$ gives

$$d_i^{\,\mathrm{complete}}(t) + \log \widehat{\pi}_i^{\,\text{plug-in}}(t) - \log \pi_i(t) = \log \widehat{\pi}_i^{\,\mathrm{complete}}(t) - \log \pi_i(t). \tag{A.22}$$

Since $\sum_{j=1}^n A_{ij} \le \Delta_n$ and $\Delta_n \ge 1$, combining (A.13), (A.20), (A.21), and (A.22) gives

$$\max_{1\le i\le n,\, t\in\{0,1\}} \left|d_i^{\,\mathrm{complete}}(t) + \log \widehat{\pi}_i^{\,\text{plug-in}}(t) - \log \pi_i(t)\right| = O_p\left\{\Delta_n(a_n^2 + \rho_n + n^{-1/2})\right\},$$

proving (A.14).

If $\rho_n + n^{-1/2} = o(a_n)$, then $a_n \to 0$ implies $a_n^2 + \rho_n + n^{-1/2} = o(a_n)$. Consequently,

$$\max_{1\le i\le n,\, t\in\{0,1\}} \left| d_i^{\text{complete}}(t) + \log \widehat{\pi}_i^{\text{plug-in}}(t) - \log \pi_i(t) \right| = o_p(\Delta_n a_n).$$

When $\widehat{e}_i^{\text{complete}}$ is restricted to $[\epsilon, 1-\epsilon]$. Suppose additionally that $e_i \in [\epsilon, 1-\epsilon]$ for every i, where $0 < \epsilon < 1/2$. Replacing $\widehat{e}_i^{\text{complete}}$ by $\min\{1-\epsilon, \max\{\epsilon, \widehat{e}_i^{\text{complete}}\}\}$ restricts $\operatorname{logit} \widehat{e}_i^{\text{complete}}$ to $[\operatorname{logit}\epsilon, \operatorname{logit}(1-\epsilon)]$. Since $\operatorname{logit} e_i$ belongs to $[\operatorname{logit}\epsilon, \operatorname{logit}(1-\epsilon)]$,

$$\left| \operatorname{logit}\Big(\min\{1-\epsilon, \max\{\epsilon, \widehat{e}_i^{\text{complete}}\}\}\Big) - \operatorname{logit} e_i \right| \le \left| \operatorname{logit} \widehat{e}_i^{\text{complete}} - \operatorname{logit} e_i \right|.$$

Therefore, (A.13) remains valid after replacing $\widehat{e}_i^{\text{complete}}$ by $\min\{1-\epsilon, \max\{\epsilon, \widehat{e}_i^{\text{complete}}\}\}$. The fold-specific correction coefficient $\widehat{\phi}_\ell$ in (A.10) is unchanged because (A.10) uses only $\{H_k, D_k, \widetilde{e}_k : k \in \mathcal{J}_\ell\}$. Applying $\pi_i(t; \cdot, A)$ to $(\min\{1-\epsilon, \max\{\epsilon, \widehat{e}_j^{\text{complete}}\}\})_{j=1}^n$ gives the score-completed exposure probabilities $\widehat{\pi}_i^{\text{complete}}(t)$ after restricting $\widehat{e}_j^{\text{complete}}$ to $[\epsilon, 1-\epsilon]$. Equations (A.20) and (A.21) apply to $(\min\{1-\epsilon, \max\{\epsilon, \widehat{e}_j^{\text{complete}}\}\})_{j=1}^n$, and (A.22) continues to hold when the propagation directions $d_i^{\text{complete}}(t)$ are computed from the score-completed exposure probabilities $\widehat{\pi}_i^{\text{complete}}(t)$. Hence, (A.14) and the conclusion under $\rho_n + n^{-1/2} = o(a_n)$ remain valid. □

B The Propagation-Integrated Estimator: Derivation, Theory and Proof

Step 1 collects the propagation directions $d_i^{\text{complete}}(t)$ into the node-level vector $\widehat{U}_i$. Step 2 uses the propagation directions $\widehat{U}_i$ in the fitted outcome predictions $\widehat{m}_i(t)$ and the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t, t')$. The derivation below identifies the component of the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$ represented along $\widehat{U}_i$ and the corresponding component of the remaining propensity–outcome product error targeted by $\widehat{\Delta}^{(2)}_{\text{prop}}(t, t')$.

Throughout the analysis, the network A, the pretreatment features, and the potential-outcome functions $Y_i(\cdot)$ are fixed. The assignment expectation $\mathbb{E}_D$ averages over the network-wide treatment assignment D. The exposure-specific means $\mu_i(t)$, the population mean $\mu(t)$, and the estimand $\tau(t, t')$ are unchanged:

$$\mu_i(t) = \frac{\mathbb{E}_D[Y_i(D)\mathbf{1}\{T_i = t\}]}{\pi_i(t)}, \qquad \mu(t) = \frac{1}{n}\sum_{i=1}^n \mu_i(t), \qquad \tau(t, t') = \mu(t) - \mu(t'). \tag{B.1}$$

The function $c_{t,t'}(T_i)$ equals 1 when $T_i = t$, -1 when $T_i = t'$, and 0 otherwise. No assignment expectation $\mathbb{E}_D$ below conditions on the full collection of plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$ and fitted outcome predictions $\widehat{m}_i(t)$.

B.1 Derivation

B.1.1 The first-order estimator

The average $n^{-1}\sum_i \widehat{m}_i(t)$ estimates $\mu(t)$ from the fitted outcome predictions $\widehat{m}_i(t)$. The normalized inverse-exposure average of the residuals $Y_i - \widehat{m}_i(t)$ adjusts $n^{-1}\sum_i \widehat{m}_i(t)$ using the observed outcomes $Y_i = Y_i(D)$. Combining the two averages gives the first-order estimator $\widehat{\mu}^{(1)}(t)$, with the Hájek normalization $\widehat{h}(t)$ from Step 2.1:

$$\widehat{h}(t) = \frac{1}{n}\sum_{i=1}^n \frac{\mathbf{1}\{T_i = t\}}{\widehat{\pi}_i^{\text{plug-in}}(t)}, \qquad \widehat{\mu}^{(1)}(t) = \frac{1}{n}\sum_{i=1}^n \widehat{m}_i(t) + \frac{1}{n\widehat{h}(t)}\sum_{i=1}^n \frac{\mathbf{1}\{T_i = t\}\{Y_i - \widehat{m}_i(t)\}}{\widehat{\pi}_i^{\text{plug-in}}(t)}. \tag{B.2}$$

Hold the numerical values of the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$ and the fitted outcome predictions $\widehat{m}_i(t)$ fixed. Replacing the empirical numerator and denominator averages in $\widehat{\mu}^{(1)}(t) - \mu(t)$ by their assignment expectations $\mathbb{E}_D$, with $\mathbb{E}_D[\mathbf{1}\{T_i = t\}Y_i(D)] = \pi_i(t)\mu_i(t)$, gives the remaining propensity–outcome product error:

$$-\frac{1}{n}\sum_{i=1}^n \left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_{j=1}^n \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1 \right] \{\widehat{m}_i(t) - \mu_i(t)\}. \tag{B.3}$$

The remaining propensity–outcome product error in (B.3) is zero when $\widehat{\pi}_i^{\text{plug-in}}(t) = \pi_i(t)$ for every i, or when $\widehat{m}_i(t) = \mu_i(t)$ for every i. Otherwise, (B.3) contains the interaction associated with the propagated exposure-probability error $\log \widehat{\pi}_i^{\text{plug-in}}(t) - \log \pi_i(t)$ and the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$. Replacing the numerator and denominator averages by their assignment expectations $\mathbb{E}_D$ is not the same as taking $\mathbb{E}_D\{\widehat{\mu}^{(1)}(t)\}$. The proof controls the difference between the empirical averages and their assignment expectations.

B.1.2 The second-order propagation correction

The exposure-t block of $\widehat{B}_i(t)$ contains the propagation directions $\widehat{U}_i$, with zeros in every other exposure block. For node j, the node-level contribution to the outcome-side moment is $\widehat{p}_j = \widehat{B}_j(T_j)c_{t,t'}(T_j)\{Y_j - \widehat{m}_j(T_j)\}$. The node-level contribution to the propensity-side moment $\widehat{q}_i$ and the second-moment matrix $\widehat{\Omega}$ are defined in Step 2.2 by

$$\widehat{q}_i = \sum_{t\in\mathcal{T}} \widehat{B}_i(t) - \frac{\widehat{B}_i(T_i)}{\widehat{\pi}_i^{\text{plug-in}}(T_i)\widehat{h}(T_i)}, \qquad \widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} \widehat{B}_i(T_i)\widehat{B}_i(T_i)^\top. \tag{B.4}$$

Replacing the empirical averages in the propensity-side moment $n^{-1}\sum_i \widehat{q}_i$ by their assignment expectations $\mathbb{E}_D$, at fixed numerical values of $\widehat{\pi}_i^{\text{plug-in}}(t)$ and $\widehat{U}_i$, gives the exposure-t block

$$\frac{1}{n}\sum_{i=1}^{n} \widehat{U}_i \left[1 - \frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_{j=1}^{n} \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)}\right]. \tag{B.5}$$

The exposure-t block in (B.5) is zero when the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$ equal the exposure probabilities $\pi_i(t)$. Replacing the empirical averages in the outcome-side moment $n^{-1}\sum_j \widehat{p}_j$ by their assignment expectations $\mathbb{E}_D$, at fixed numerical values of $\widehat{m}_j(t)$ and $\widehat{U}_j$, gives the exposure blocks

$$\frac{1}{n}\sum_{j=1}^{n} \pi_j(t)\widehat{U}_j\{\mu_j(t) - \widehat{m}_j(t)\} \quad \text{at exposure } t,$$

$$-\frac{1}{n}\sum_{j=1}^{n} \pi_j(t')\widehat{U}_j\{\mu_j(t') - \widehat{m}_j(t')\} \quad \text{at exposure } t'.$$

Every other exposure block is zero. The exposure-t block of the second-moment matrix $\widehat{\Omega}$ is $n^{-1}\sum_j \mathbf{1}\{T_j = t\}\widehat{U}_j\widehat{U}_j^\top$. The matrix $\widehat{\Omega}^\dagger$ converts the outcome-side moment $n^{-1}\sum_j \widehat{p}_j$ into coefficients along the propagation directions $\widehat{U}_i$. Replacing the empirical average in the exposure-t block of $\widehat{\Omega}$ by its assignment expectation $\mathbb{E}_D$ gives $n^{-1}\sum_j \pi_j(t)\widehat{U}_j\widehat{U}_j^\top$, not $n^{-1}\sum_j \widehat{U}_j\widehat{U}_j^\top$. The weights $\pi_j(t)$ in the theoretical result therefore follow from the second-moment matrix $\widehat{\Omega}$.

The bilinear product of the propensity-side moment $n^{-1}\sum_i \widehat{q}_i$ and the outcome-side moment $n^{-1}\sum_j \widehat{p}_j$, with $\widehat{\Omega}^\dagger$ between the two moments, estimates the component of the remaining propensity–outcome product error in (B.3) represented by the propagation directions $\widehat{U}_i$, with the opposite sign. Step 2.2 excludes the diagonal self-product with $i = j$ and retains $n(n-1)$ in the denominator of the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$:

$$\widehat{\Delta}^{(2)}_{\text{prop}}(t,t') = \frac{1}{n(n-1)}\sum_{i\neq j} \widehat{q}_i^\top \widehat{\Omega}^\dagger \widehat{B}_j(T_j)c_{t,t'}(T_j)\{Y_j - \widehat{m}_j(T_j)\}. \tag{B.6}$$

The final PIN point estimator $\widehat{\tau}_{\text{PIN}}(t,t')$ remains

$$\widehat{\tau}_{\text{PIN}}(t,t') = \widehat{\mu}^{(1)}(t) - \widehat{\mu}^{(1)}(t') + \widehat{\Delta}^{(2)}_{\text{prop}}(t,t'). \tag{B.7}$$

The second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$ uses the second-order higher-order influence function construction of Robins et al. (2017). Distinct network nodes need not be independent. The proof uses the exact factorization of $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$ and bounds for averages over nodes, rather than independence of nodes i and j.

B.1.3 Exact factorization of the second-order propagation correction

Expanding the bilinear product of the propensity-side moment $n^{-1}\sum_i \widehat{q}_i$ and the outcome-side moment $n^{-1}\sum_j \widehat{p}_j$, then subtracting the diagonal self-products $\widehat{q}_i^\top \widehat{\Omega}^\dagger \widehat{p}_i$, gives

$$\widehat{\Delta}^{(2)}_{\mathrm{prop}}(t,t') = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n \widehat{q}_i\right)^\top \widehat{\Omega}^\dagger \left[\frac{1}{n}\sum_{j=1}^n \widehat{B}_j(T_j)c_{t,t'}(T_j)\{Y_j - \widehat{m}_j(T_j)\}\right] - \frac{1}{n(n-1)}\sum_{i=1}^n \widehat{q}_i^\top \widehat{\Omega}^\dagger \widehat{B}_i(T_i)c_{t,t'}(T_i)\{Y_i - \widehat{m}_i(T_i)\}. \tag{B.8}$$

B.2 Conditions

The following conditions concern the network-wide treatment assignment D, the treatment indicators entering the exposure mapping T_i and the observed outcome $Y_i(D)$, and the arrays of plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$, fitted outcome predictions $\widehat{m}_i(t)$, and propagation directions $\widehat{U}_i$.

Assumption B.1. Conditional on the network A, the pretreatment features, and the fixed potential-outcome functions $Y_i(\cdot)$, the treatment indicators $D_1,\dots,D_n$ are independent Bernoulli variables with treatment propensities $e_1,\dots,e_n$. The treatment propensities e_i are represented by the propensity model based on the pretreatment features H_i. The exposure set $\mathcal{T}$ is fixed. For constants $\epsilon, C > 0$ independent of n,

$$\min_{i,\,t\in\mathcal{T}} \pi_i(t) \geq \epsilon, \qquad \max_{i,\mathbf{d}\in\{0,1\}^n} |Y_i(\mathbf{d})| \leq C.$$

The plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$ satisfy $\min_{i,\,t\in\mathcal{T}} \widehat{\pi}_i^{\text{plug-in}}(t) \geq \epsilon/2$ with probability tending to one.

The lower bound on the exposure probabilities $\pi_i(t)$ is not a lower bound on the treatment propensities e_i. Clipping the plug-in treatment propensities $\widehat{e}_i^{\text{plug-in}}$ does not establish the lower bound on $\pi_i(t)$.

Assumption B.2. The shortest-path distance ℓ_A is symmetric. For a fixed integer $R \geq 1$, $T(i,\mathbf{d},A)$ and $Y_i(\mathbf{d})$ depend only on $\{d_j : \ell_A(i,j) \leq R\}$. The distance cutoff b_n is chosen without D or Y and satisfies $b_n \geq 2R$. For a fixed $\eta \in [0, 1/4)$,

$$\max_i \left|\{j : \ell_A(i,j) \leq b_n\}\right| \leq Cn^\eta. \tag{B.9}$$

The kernel K in Step 3 has entries $K_{ij} = \mathbf{1}\{\ell_A(i,j) \leq b_n\}$. Replacing the negative eigenvalues of K by zero gives the kernel K^+.

The outcome $Y_i(\mathbf{d})$ need not be determined by $T(i,\mathbf{d},A)$. For a directed network, the chosen symmetric shortest-path distance ℓ_A must include every treatment variable entering T_i and $Y_i(D)$ within distance R.

Assumption B.3. The number of outer folds L and the dimension r of the propagation directions $\widehat{U}_i$ are bounded independently of n. For $r > 0$, $n^{-1}\sum_i \widehat{U}_i\widehat{U}_i^\top$ is nonsingular with probability tending to one. After one common nonsingular linear transformation of the propagation directions $\widehat{U}_i$ that preserves every fitted outcome prediction $\widehat{m}_i(t)$,

$$\frac{1}{n}\sum_{i=1}^n \widehat{U}_i\widehat{U}_i^\top = I_r, \qquad \max_i \|\widehat{U}_i\| \leq C, \qquad \max_{i,\,t\in\mathcal{T}} |\widehat{m}_i(t)| \leq C \tag{B.10}$$

with probability tending to one. The normalization of the propagation directions $\widehat{U}_i$ is only for analysis and does not change the propagation directions $\widehat{U}_i$ used in the implementation.

Assumption B.4. For fixed $a, b > 0$,

$$\max_{i,\,t\in\mathcal{T}} |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)| = O_p(n^{-a}), \qquad \max_{i,\,t\in\mathcal{T}} |\widehat{m}_i(t) - \mu_i(t)| = O_p(n^{-b}). \tag{B.11}$$

The first rate in (B.11) is stated for $\max_{i,t} |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)|$, using the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$ and the exposure probabilities $\pi_i(t)$ associated with the propagated exposure-probability error $\log \widehat{\pi}_i^{\text{plug-in}}(t) - \log \pi_i(t)$. The second rate is stated for the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$. Assumption B.4 places no additional rate condition on $\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v$ after subtracting the best $\pi_i(t)$-weighted fit $\widehat{U}_i^\top v$. For the binary exposure mapping in Appendix A, Assumptions A.1–A.2

and the derivative bounds in Appendix A also give

$$\max_{i,\, t\in\mathcal{T}} |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)| = O_p\{\Delta_n(a_n + \rho_n)\}.$$

Thus, $\Delta_n(a_n + \rho_n) = O(n^{-a})$ is one sufficient way to obtain the first rate in (B.11). Appendix A does not establish the rate for the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$ or the restrictions on $\widehat{\pi}_i^{\text{plug-in}}(t)$, $\widehat{m}_i(t)$, and $\widehat{U}_i$ in Assumption B.3.

B.3 Decomposing the remaining propensity–outcome product error after the first-order estimator

The calculations below isolate the remaining propensity–outcome product error in (B.3). Empirical averages are replaced by their assignment expectations $\mathbb{E}_D$ at fixed numerical values of the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$, the fitted outcome predictions $\widehat{m}_i(t)$, and the propagation directions $\widehat{U}_i$, not by conditioning on the fitted arrays. Appendix C.4.3 connects the empirical averages and their assignment expectations to the final PIN point estimator $\widehat{\tau}_{\text{PIN}}(t, t')$.

B.3.1 What the first-order adjustment removes

Adding and subtracting $\mu_i(t)$ in (B.2) gives the exact identity

$$\begin{aligned}\widehat{\mu}^{(1)}(t) - \mu(t) = \frac{1}{n}\sum_{i=1}^{n}\{\widehat{m}_i(t) - \mu_i(t)\} - \frac{1}{n\widehat{h}(t)}\sum_{i=1}^{n}\frac{\mathbf{1}\{T_i = t\}\{\widehat{m}_i(t) - \mu_i(t)\}}{\widehat{\pi}_i^{\text{plug-in}}(t)} \\ + \frac{1}{n\widehat{h}(t)}\sum_{i=1}^{n}\frac{\mathbf{1}\{T_i = t\}\{Y_i - \mu_i(t)\}}{\widehat{\pi}_i^{\text{plug-in}}(t)}.\end{aligned} \tag{B.12}$$

The average of the fitted outcome predictions $n^{-1}\sum_i \widehat{m}_i(t)$ contributes $n^{-1}\sum_i\{\widehat{m}_i(t) - \mu_i(t)\}$. The normalized inverse-exposure average of the residuals $Y_i - \widehat{m}_i(t)$ subtracts a normalized inverse-exposure average of the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$. When $\widehat{\pi}_i^{\text{plug-in}}(t) = \pi_i(t)$, replacing the empirical numerator and denominator averages by their assignment expectations $\mathbb{E}_D$ makes the two averages of the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$ cancel. When $\widehat{m}_i(t) = \mu_i(t)$, both averages are zero. The assignment expectation $\mathbb{E}_D$ of the numerator involving $Y_i - \mu_i(t)$ is zero because

$$\mathbb{E}_D[\mathbf{1}\{T_i = t\}\{Y_i(D) - \mu_i(t)\}] = 0. \tag{B.13}$$

After replacing empirical averages by their assignment expectations $\mathbb{E}_D$, the first-order adjustment in $\widehat{\mu}^{(1)}(t)$ removes the contribution of either the propagated exposure-probability error $\log\widehat{\pi}_i^{\text{plug-in}}(t) - \log\pi_i(t)$ or the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$ acting alone. The observed average involving $Y_i - \mu_i(t)$ remains random and is not the remaining propensity–outcome product error in (B.3).

When the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$ and the fitted outcome predictions $\widehat{m}_i(t)$ are inaccurate, replacing the empirical averages in (B.12) by their assignment expectations $\mathbb{E}_D$ leaves the remaining propensity–outcome product error in (B.3). The bracketed factor in (B.3) can be written using $\pi_i(t) - \widehat{\pi}_i^{\text{plug-in}}(t)$ as

$$\begin{aligned}&\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1 \\ &\quad= \frac{\{\pi_i(t) - \widehat{\pi}_i^{\text{plug-in}}(t)\}/\widehat{\pi}_i^{\text{plug-in}}(t) - n^{-1}\sum_j\{\pi_j(t) - \widehat{\pi}_j^{\text{plug-in}}(t)\}/\widehat{\pi}_j^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)}.\end{aligned} \tag{B.14}$$

The bracketed factor in (B.3) is associated with the propagated exposure-probability error $\log\widehat{\pi}_i^{\text{plug-in}}(t) - \log\pi_i(t)$, while $\widehat{m}_i(t) - \mu_i(t)$ is the outcome-side error. Under the lower bounds on the exposure probabilities $\pi_i(t)$ and plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$ in Assumption B.1, weighted Cauchy–Schwarz gives, for a constant C independent of n,

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\{\widehat{m}_i(t) - \mu_i(t)\}\right| \\ &\qquad\le C\max_i|\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)|\left[\frac{1}{n}\sum_{i=1}^{n}\pi_i(t)\{\widehat{m}_i(t) - \mu_i(t)\}^2\right]^{1/2}.\end{aligned} \tag{B.15}$$

The bound in (B.15) uses $\max_i |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)|$ and the $\pi_i(t)$-weighted norm of the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$. Equation (B.15) bounds the remaining propensity–outcome product error before adding the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t, t')$.

B.3.2 What the second-order propagation correction removes from the remaining propensity–outcome product error

For any coefficient vector $v \in \mathbb{R}^r$ along the propagation directions $\widehat{U}_i$, the remaining propensity–outcome product error separates exactly as

$$\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\{\widehat{m}_i(t) - \mu_i(t)\} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\widehat{U}_i^\top v \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}.
\end{aligned} \tag{B.16}$$

At exposure t, the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t, t')$ is constructed to offset

$$-\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\widehat{U}_i^\top v$$

in (B.16). The coefficient vector v along the propagation directions $\widehat{U}_i$ is not arbitrary when relating (B.16) to the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t, t')$. For $r > 0$, suppose $\sum_j \pi_j(t)\widehat{U}_j\widehat{U}_j^\top$ is nonsingular. Use $v \in \mathbb{R}^r$ satisfying

$$\left[\frac{1}{n}\sum_{j=1}^{n}\pi_j(t)\widehat{U}_j\widehat{U}_j^\top\right]v = \frac{1}{n}\sum_{j=1}^{n}\pi_j(t)\widehat{U}_j\{\widehat{m}_j(t) - \mu_j(t)\}. \tag{B.17}$$

To see why (B.17) determines the relevant v, hold the numerical values of $\widehat{\pi}_i^{\text{plug-in}}(t)$, $\widehat{m}_i(t)$, and $\widehat{U}_i$ fixed. In the product of averages in (B.8), replace each empirical average by its expectation under $\mathbb{E}_D$. The exposure-t contribution becomes

$$\begin{aligned}
&\left[\frac{1}{n}\sum_{i=1}^{n}\widehat{U}_i\left\{1 - \frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)}\right\}\right]^\top \\
&\qquad \times \left[\frac{1}{n}\sum_{j=1}^{n}\pi_j(t)\widehat{U}_j\widehat{U}_j^\top\right]^{-1}\left[\frac{1}{n}\sum_{j=1}^{n}\pi_j(t)\widehat{U}_j\{\mu_j(t) - \widehat{m}_j(t)\}\right] \\
&= \left[\frac{1}{n}\sum_{i=1}^{n}\widehat{U}_i\left\{1 - \frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)}\right\}\right]^\top (-v) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\widehat{U}_i^\top v.
\end{aligned} \tag{B.18}$$

The factor $-v$ follows from (B.17) because $\mu_j(t) - \widehat{m}_j(t) = -\{\widehat{m}_j(t) - \mu_j(t)\}$. The last line of (B.18) has the opposite sign to the sum multiplying $\widehat{U}_i^\top v$ in (B.16). Consequently,

$$\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\{\widehat{m}_i(t) - \mu_i(t)\} \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\widehat{U}_i^\top v \\
&= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1\right]\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}.
\end{aligned} \tag{B.19}$$

After replacing the empirical averages separately by their expectations under $\mathbb{E}_D$, the exposure-t contribution of the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t, t')$ cancels the sum contain-

ing $\widehat{U}_i^\top v$ in (B.16), where v satisfies (B.17). The final sum in (B.19) therefore multiplies $\{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)\}/\{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)\} - 1$ by $\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v$, rather than by $\widehat{m}_i(t) - \mu_i(t)$.

The vector v in (B.19) satisfies (B.17). Since $\sum_i \pi_i(t)\widehat{U}_i\widehat{U}_i^\top$ is positive definite, v minimizes $n^{-1}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2$ over $\mathbb{R}^r$. Under Assumption B.1, work on the event $\min_i \widehat{\pi}_i^{\text{plug-in}}(t) \geq \epsilon/2$. The bounds $\pi_i(t) \geq \epsilon$ and $\widehat{\pi}_i^{\text{plug-in}}(t) \leq 1$ imply $n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t) \geq \epsilon$. Equation (B.14) and $\widehat{\pi}_i^{\text{plug-in}}(t) \geq \epsilon/2$ therefore give

$$\max_i \left| \frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1 \right| \leq \frac{4}{\epsilon^2} \max_i |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)|. \tag{B.20}$$

Applying the weighted Cauchy–Schwarz inequality to the final sum in (B.19) yields

$$\begin{aligned}
&\left| -\frac{1}{n}\sum_{i=1}^n \left[\frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1 \right] \{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\} \right| \\
&\leq \left[\frac{1}{n}\sum_{i=1}^n \frac{1}{\pi_i(t)} \left\{ \frac{\pi_i(t)/\widehat{\pi}_i^{\text{plug-in}}(t)}{n^{-1}\sum_j \pi_j(t)/\widehat{\pi}_j^{\text{plug-in}}(t)} - 1 \right\}^2 \right]^{1/2} \left[\frac{1}{n}\sum_{i=1}^n \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2 \right]^{1/2} \\
&\leq C \max_i |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)| \left[\frac{1}{n}\sum_{i=1}^n \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2 \right]^{1/2} \\
&= C \max_i |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)| \inf_{v\in\mathbb{R}^r} \left[\frac{1}{n}\sum_{i=1}^n \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2 \right]^{1/2}.
\end{aligned} \tag{B.21}$$

Equation (B.15) bounds the remaining propensity–outcome product error after the first-order estimator $\widehat{\mu}^{(1)}(t)$ by

$$C \max_i |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)| \left[\frac{1}{n}\sum_{i=1}^n \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t)\}^2 \right]^{1/2},$$

whereas (B.21) bounds the remaining propensity–outcome product error after adding the exposure-t contribution of the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$ by

$$C \max_i |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)| \inf_{v\in\mathbb{R}^r} \left[\frac{1}{n}\sum_{i=1}^n \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2 \right]^{1/2}.$$

The two bounds have the same factor $C \max_i |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)|$. Therefore, comparing (B.15) and (B.21) reduces to comparing

$$\inf_{v\in\mathbb{R}^r} \left[\frac{1}{n}\sum_{i=1}^n \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2 \right]^{1/2}$$

with

$$\left[\frac{1}{n}\sum_{i=1}^n \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t)\}^2 \right]^{1/2}.$$

Proposition B.1 gives this comparison.

B.4 Theoretical Result

Proposition B.1. *For every $t \in \mathcal{T}$ with $\min_i \pi_i(t) > 0$, the bounds on the remaining propensity–outcome product error before and after adding the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$*

satisfy

$$\max_i |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)| \inf_{v\in\mathbb{R}^r} \left[\frac{1}{n}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2\right]^{1/2}$$
$$\leq \max_i |\widehat{\pi}_i^{\text{plug-in}}(t) - \pi_i(t)| \left[\frac{1}{n}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t)\}^2\right]^{1/2}. \tag{B.22}$$

The left-hand side of (B.22) is the bound on the remaining propensity–outcome product error after adding $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$ in (B.21). The right-hand side is the bound on the remaining propensity–outcome product error before adding $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$ in (B.15). The common constant C is omitted. For $r>0$, the inequality is strict precisely when $\max_i |\widehat{\pi}_i^{\text{plug-in}}(t)-\pi_i(t)|>0$ and $\sum_i \pi_i(t)\widehat{U}_i\{\widehat{m}_i(t)-\mu_i(t)\}\neq 0$. Applying (B.22) at t and t' and adding compares the bounds on the remaining propensity–outcome product error for the estimand $\tau(t,t')$.

The comparison is between upper bounds on the remaining propensity–outcome product error. The comparison does not order the absolute values of the signed sums in (B.3) and (B.19), or the realized errors $\widehat{\tau}^{(1)}(t,t') - \tau(t,t')$ and $\widehat{\tau}_{\text{PIN}}(t,t') - \tau(t,t')$.

B.5 Proof

Proof of Proposition B.1. For any $v\in\mathbb{R}^r$ minimizing $n^{-1}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2$, the normal equation (B.17) implies $\sum_i \pi_i(t)\widehat{U}_i\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\} = 0$. Expanding the square of the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$ therefore gives

$$\frac{1}{n}\sum_i \pi_i(t)\{\widehat{m}_i(t)-\mu_i(t)\}^2 = \frac{1}{n}\sum_i \pi_i(t)\{\widehat{m}_i(t)-\mu_i(t)-\widehat{U}_i^\top v\}^2 + \frac{1}{n}\sum_i \pi_i(t)(\widehat{U}_i^\top v)^2. \tag{B.23}$$

Since $n^{-1}\sum_i \pi_i(t)(\widehat{U}_i^\top v)^2 \geq 0$,

$$\inf_{v\in\mathbb{R}^r}\left[\frac{1}{n}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t) - \widehat{U}_i^\top v\}^2\right]^{1/2} \leq \left[\frac{1}{n}\sum_i \pi_i(t)\{\widehat{m}_i(t) - \mu_i(t)\}^2\right]^{1/2}. \tag{B.24}$$

Multiplying by the nonnegative $\max_i |\widehat{\pi}_i^{\text{plug-in}}(t)-\pi_i(t)|$ proves (B.22). Positive $\pi_i(t)$ and the normal equation imply that $n^{-1}\sum_i \pi_i(t)(\widehat{U}_i^\top v)^2 > 0$ exactly when $\sum_i \pi_i(t)\widehat{U}_i\{\widehat{m}_i(t) - \mu_i(t)\}\neq 0$. The inequality for the bounds on the remaining propensity–outcome product error is therefore strict exactly under the two stated conditions. For $r=0$, both sides of (B.22) are equal. □

C Propagation-Integrated Standard Error: Derivation, Theory and Proof

Appendix C derives and analyzes the propagation-integrated standard error $\widehat{\text{SE}}_{\text{PIN}}(t,t')$ in Step 3. The proofs use the final PIN point estimator $\widehat{\tau}_{\text{PIN}}(t,t')$ in (B.7) and the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$ in (B.6). The network A, the pretreatment features, and the potential-outcome functions $Y_i(\cdot)$ remain fixed. Assignment expectations $\mathbb{E}_D$ do not condition on the fitted arrays.

C.1 Derivation

C.1.1 The centered node-level score

Fix the contrast t,t'. The node-level contribution to the outcome-side moment $\widehat{p}_i$ is indexed by the node supplying the residual $Y_i - \widehat{m}_i(T_i)$:

$$\widehat{p}_i = \widehat{B}_i(T_i)c_{t,t'}(T_i)\{Y_i - \widehat{m}_i(T_i)\}, \qquad \widehat{\Delta}^{(2)}_{\text{prop}}(t,t') = \frac{1}{n(n-1)}\sum_{i\neq j}\widehat{q}_i^\top\widehat{\Omega}^\dagger\widehat{p}_j. \tag{C.1}$$

The second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$ in (C.1) equals (B.6). Replacing the node-level contribution to the outcome-side moment $\widehat{p}_j$ by $\widehat{p}_i$ throughout the ordered-pair sum would define a different estimator.

Holding the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$, the fitted outcome predictions $\widehat{m}_i(t)$, and the propagation directions $\widehat{U}_i$ fixed, differentiation of $n^{-1}\sum_k \widehat{m}_k(t)$ with respect to node i's weight gives $n^{-1}\{\widehat{m}_i(t) - n^{-1}\sum_k \widehat{m}_k(t)\}$. The quotient rule for the normalized inverse-exposure average

of the residuals $Y_i - \widehat{m}_i(t)$ subtracts $\widehat{\mu}^{(1)}(t) - n^{-1}\sum_k \widehat{m}_k(t)$ from $Y_i - \widehat{m}_i(t)$ inside the normalized inverse-exposure contribution. In the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$, node i appears in both positions of an ordered pair. Summing both node-level contributions over i counts each ordered pair twice, so centering subtracts $2\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')/n$ from each node's contribution. Combining the contributions gives the centered node-level score $\widehat{s}_i^{\text{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}})$:

$$\begin{aligned}
&\widehat{s}_i^{\text{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}}) \\
&= \frac{1}{n}\left[\widehat{m}_i(t) - \widehat{m}_i(t') - \frac{1}{n}\sum_{k=1}^{n}\{\widehat{m}_k(t) - \widehat{m}_k(t')\}\right] \\
&\quad + \frac{c_{t,t'}(T_i)}{n\widehat{\pi}_i^{\text{plug-in}}(T_i)\widehat{h}(T_i)}\left\{Y_i - \widehat{m}_i(T_i) - \widehat{\mu}^{(1)}(T_i) + \frac{1}{n}\sum_{k=1}^{n}\widehat{m}_k(T_i)\right\} \\
&\quad + \frac{1}{n(n-1)}\sum_{j\neq i}\left\{\widehat{q}_i^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_j + \widehat{q}_j^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_i\right\} - \frac{2}{n}\widehat{\Delta}^{(2)}_{\text{prop}}(t,t').
\end{aligned} \tag{C.2}$$

Each of the three parts of the centered node-level score $\widehat{s}_i^{\text{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}})$ in (C.2) sums to zero, so $\sum_i \widehat{s}_i^{\text{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}}) = 0$. The calculation for the second-order propagation correction holds the evaluated bilinear products $\widehat{q}_i^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_j$ fixed rather than refitting the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$ and fitted outcome predictions $\widehat{m}_i(t)$ or recomputing the Hájek normalization $\widehat{h}(t)$ and the second-moment matrix $\widehat{\Omega}$. The proof below establishes the stochastic approximation of the centered node-level score in (C.2) directly.

C.1.2 The remaining pairwise component

For the remaining pairwise component, the matrix $\widehat{R}(t,t')$ averages both orderings of $\widehat{q}_i^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_j$, subtracts the two node-specific pair averages, and restores the overall pair average. For $i \neq j$,

$$\begin{aligned}
\widehat{R}_{ij}(t,t') &= \frac{1}{2}\left\{\widehat{q}_i^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_j + \widehat{q}_j^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_i\right\} \\
&\quad - \frac{1}{2(n-1)}\sum_{k\neq i}\left\{\widehat{q}_i^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_k + \widehat{q}_k^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_i\right\} \\
&\quad - \frac{1}{2(n-1)}\sum_{k\neq j}\left\{\widehat{q}_j^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_k + \widehat{q}_k^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_j\right\} + \widehat{\Delta}^{(2)}_{\text{prop}}(t,t'), \qquad \widehat{R}_{ii}(t,t') = 0.
\end{aligned} \tag{C.3}$$

Interchanging i and j leaves (C.3) unchanged, so $\widehat{R}(t,t')$ is symmetric.

C.1.3 The propagation-integrated standard error

The network variance term $\widehat{\text{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\text{PIN}}(t,t')\}$ is

$$\widehat{\text{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\text{PIN}}(t,t')\} = \sum_{i,j} K^{+}_{ij}\widehat{s}_i^{\text{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}})\widehat{s}_j^{\text{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}}). \tag{C.4}$$

Symmetry of $\widehat{R}(t,t')$ and positive semidefiniteness of K^{+} give

$$\text{tr}\{\widehat{R}(t,t')K^{+}\widehat{R}(t,t')K^{+}\} = \|(K^{+})^{1/2}\widehat{R}(t,t')(K^{+})^{1/2}\|_F^2 \geq 0, \tag{C.5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Therefore, the propagation-integrated standard error $\widehat{\text{SE}}_{\text{PIN}}(t,t')$ in (7) satisfies

$$\widehat{\text{SE}}_{\text{PIN}}(t,t') \geq \left[\widehat{\text{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\text{PIN}}(t,t')\}\right]^{1/2}. \tag{C.6}$$

The coverage proof uses (C.6) and a derived lower bound for the network variance term $\widehat{\text{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\text{PIN}}(t,t')\}$ in (C.4), not separate consistency of the two terms in the propagation-integrated standard error $\widehat{\text{SE}}_{\text{PIN}}(t,t')$ in (7).

C.2 Conditions

Reuse Assumptions B.1–B.4 from Appendix B, including the exponents a, b, the neighborhood bound with $\eta < 1/4$, and the common normalization of the propagation directions $\widehat{U}_i$. Two additional assumptions are needed.

Assumption C.1. With probability tending to one, the arrays $\widehat{\pi}_i^{\text{plug-in}}(u), \widehat{m}_i(u), \widehat{U}_{i_{i,},u\in\mathcal{T}}$ belong to deterministic families of uniformly bounded numerical arrays. Within the deterministic families, the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(u)$ are fixed numerical values in $[\epsilon/2, 1]$. For every $0 < \delta < 1$, each deterministic family has at most $C(n/\delta)^C$ approximating arrays. Every array in the deterministic family is within δ of an approximating array in maximum absolute entry distance. The constant C does not depend on n or δ.

Assumption C.1 permits simultaneous control of the arrays of plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$, fitted outcome predictions $\widehat{m}_i(t)$, and propagation directions $\widehat{U}_i$.

Define the assignment contribution $V_i(t,t')$ and the variance $\sigma_n^2(t,t')$ of $n^{-1}\sum_i V_i(t,t')$ by

$$V_i(t,t') = \frac{c_{t,t'}(T_i)}{\pi_i(T_i)}\{Y_i(D) - \mu_i(T_i)\}, \qquad \sigma_n^2(t,t') = \text{Var}_D\left\{\frac{1}{n}\sum_i V_i(t,t')\right\}. \tag{C.7}$$

Equation (B.13) gives $\mathbb{E}_D V_i(t,t') = 0$. Write $\sigma_n(t,t') = \{\sigma_n^2(t,t')\}^{1/2}$.

Assumption C.2. For the fixed contrast t, t',

$$\liminf_{n\to\infty} n\sigma_n^2(t,t') > 0. \tag{C.8}$$

C.3 Theoretical Results

Theorem C.1. *Under Assumptions B.1–B.4, C.1, and C.2, suppose*

$$\min(a,b) > \eta/2, \qquad a+b > 1/2. \tag{C.9}$$

Here, η controls the growth of local network neighborhoods, while a and b are the convergence-rate exponents for the plug-in exposure probabilities and the fitted outcome predictions, respectively. The rate conditions require both errors to converge faster than $n^{-\eta/2}$ and their product to be $o(n^{-1/2})$. Then

$$\frac{\widehat{\tau}_{\text{PIN}}(t,t') - \tau(t,t')}{\sigma_n(t,t')} \xrightarrow{d} N(0,1). \tag{C.10}$$

Theorem C.2. *Under Assumptions B.1–B.4 and C.1, suppose*

$$\min(a,b) > \eta, \qquad a+b > 1/2. \tag{C.11}$$

The network variance term $\widehat{\text{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\text{PIN}}(t,t')\}$ satisfies

$$\begin{aligned}
&\widehat{\text{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\text{PIN}}(t,t')\} - \sigma_n^2(t,t') \\
&= \frac{1}{n^2}\sum_{i,j}(K_{ij}^+ - K_{ij})\,\text{Cov}_D\{V_i(t,t'), V_j(t,t')\} \\
&\quad + \frac{1}{n^2}\sum_{i,j} K_{ij}^+\{\mu_i(t) - \mu_i(t') - \tau(t,t')\}\{\mu_j(t) - \mu_j(t') - \tau(t,t')\} + o_p(n^{-1}).
\end{aligned} \tag{C.12}$$

Both double sums in (C.12) are nonnegative. If Assumption C.2 also holds, then, for every $\varepsilon > 0$,

$$\Pr\left\{\widehat{\text{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\text{PIN}}(t,t')\} \geq (1-\varepsilon)\sigma_n^2(t,t')\right\} \longrightarrow 1. \tag{C.13}$$

For every $\alpha \in (0,1)$,

$$\liminf_{n\to\infty} \Pr\left\{|\widehat{\tau}_{\text{PIN}}(t,t') - \tau(t,t')| \leq \Phi^{-1}(1-\alpha/2)\widehat{\text{SE}}_{\text{PIN}}(t,t')\right\} \geq 1-\alpha, \tag{C.14}$$

where Φ is the standard normal distribution function.

Proposition C.3. *Under Assumptions B.1–B.4 and C.1,*

$$0 \leq \left[\frac{2\,\text{tr}\{\widehat{R}(t,t')K^+\widehat{R}(t,t')K^+\}}{\{n(n-1)\}^2}\right]^{1/2} = O_p(n^{\eta-1}) = o_p(n^{-1/2}). \tag{C.15}$$

For finite observed quantities and $\xi \sim N(0, K^+)$ drawn conditional on the data,

$$\text{Var}_\xi\left[\frac{\xi^\top\widehat{R}(t,t')\xi - \text{tr}\{\widehat{R}(t,t')K^+\}}{n(n-1)}\right] = \frac{2\,\text{tr}\{\widehat{R}(t,t')K^+\widehat{R}(t,t')K^+\}}{\{n(n-1)\}^2}. \tag{C.16}$$

C.4 Proof

C.4.1 Bounds used by the proofs

Connect nodes i, j whenever $\ell_A(i,j) \leq 2R$. Assumptions B.1 and B.2 make collections without a connecting edge independent and bound each closed neighborhood by Cn^η. The dependency graph

therefore has a coloring with at most Cn^{η} colors and independent variables within each color. For deterministic candidate functions $f_i(D)$ depending on treatments within distance R and bounded by Cn^{-u}, the bounded-variable exponential-moment inequality within colors and Hölder's inequality across colors give

$$\Pr\left\{\left|\frac{1}{n}\sum_i\{f_i(D)-\mathbb{E}_D f_i(D)\}\right|>x\right\}\le 2\exp(-cn^{1-\eta+2u}x^2). \tag{C.17}$$

Here $u\ge 0$ is fixed. Apply a union bound to a polynomial-size approximating set with maximum function-value distance n^{-u-2}. Approximation changes the centered average by at most $2n^{-u-2}$, yielding the uniform bound

$$\left|\frac{1}{n}\sum_i\{f_i(D)-\mathbb{E}_D f_i(D)\}\right| = O_p\left(n^{-u}n^{-(1-\eta)/2}\sqrt{\log n}\right). \tag{C.18}$$

For a single deterministic array, counting at most $Cn^{1+\eta}$ nonzero covariances gives variance at most $Cn^{\eta-1-2u}$ and removes $\sqrt{\log n}$ from the bound. Reciprocation on $[\epsilon/2,1]$ and multiplication of bounded entries preserve the approximation counts in Assumption C.1. The uniform bound therefore permits substitution of the fitted arrays without conditioning on the fitted arrays.

Applying (C.18) to the Hájek normalization $\widehat{h}(u)$ and using (B.11) gives, uniformly over $u\in\mathcal{T}$,

$$\begin{aligned}\widehat{h}(u)-\frac{1}{n}\sum_i\frac{\pi_i(u)}{\widehat{\pi}_i^{\text{plug-in}}(u)} &= O_p\left(n^{-(1-\eta)/2}\sqrt{\log n}\right),\\ \widehat{h}(u)-1 &= O_p\left(n^{-a}+n^{-(1-\eta)/2}\sqrt{\log n}\right).\end{aligned} \tag{C.19}$$

The second line follows because the absolute difference between $n^{-1}\sum_i\pi_i(u)/\widehat{\pi}_i^{\text{plug-in}}(u)$ and one is at most $(2/\epsilon)\max_i|\widehat{\pi}_i^{\text{plug-in}}(u)-\pi_i(u)|$. Consequently, every Hájek normalization $\widehat{h}(u)$ exceeds $1/2$ with probability tending to one.

C.4.2 The second-moment matrix, the propensity-side moment, and the outcome-side moment

In the common normalization of the propagation directions $\widehat{U}_i$ in (B.10), entrywise application of (C.18) and bounded r give

$$\left\|\frac{1}{n}\sum_i\{\mathbf{1}\{T_i=u\}-\pi_i(u)\}\widehat{U}_i\widehat{U}_i^{\top}\right\|_{\text{op}} = O_p\left(n^{-(1-\eta)/2}\sqrt{\log n}\right), \tag{C.20}$$

where $\|\cdot\|_{\text{op}}$ is the Euclidean operator norm. The lower bound on the exposure probabilities $\pi_i(u)$ in Assumption B.1 gives

$$\epsilon I_r \preceq \frac{1}{n}\sum_i\pi_i(u)\widehat{U}_i\widehat{U}_i^{\top}\preceq I_r. \tag{C.21}$$

Equations (C.20) and (C.21) make every exposure block of the second-moment matrix $\widehat{\Omega}$ nonsingular with probability tending to one and give $\|\widehat{\Omega}^{\dagger}\|_{\text{op}}=O_p(1)$ in the common normalization of the propagation directions $\widehat{U}_i$. On the nonsingular event, $\widehat{\Omega}^{\dagger}$ is the ordinary inverse. Under a common nonsingular linear transformation of the propagation directions $\widehat{U}_i$, the block transformations of $\widehat{q}_i$ and $\widehat{p}_j$ cancel with the inverse transformations of $\widehat{\Omega}^{\dagger}$. The bilinear products $\widehat{q}_i^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_j$, fitted outcome predictions $\widehat{m}_i(t)$, the centered node-level score $\widehat{s}_i^{\text{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}})$, matrix $\widehat{R}(t,t')$, and propagation-integrated standard error $\widehat{\text{SE}}_{\text{PIN}}(t,t')$ are therefore preserved on the nonsingular event. No transformation rule for a singular Moore–Penrose inverse is needed.

Replacing empirical averages in the propensity-side moment $n^{-1}\sum_i\widehat{q}_i$ by their assignment expectations $\mathbb{E}_D$ gives (B.5), which is $O_p(n^{-a})$ by (B.20). Replacing empirical averages in the outcome-side moment $n^{-1}\sum_i\widehat{p}_i$ by their assignment expectations $\mathbb{E}_D$ gives the exposure-u block $c_{t,t'}(u)n^{-1}\sum_i\pi_i(u)\widehat{U}_i\{\mu_i(u)-\widehat{m}_i(u)\}$, which is $O_p(n^{-b})$. The replacements hold the numerical

values of $\widehat{\pi}_i^{\text{plug-in}}(u)$, $\widehat{m}_i(u)$, and $\widehat{U}_i$ fixed. Applying (C.18), (C.19), and (B.13) therefore gives

$$\begin{aligned}\left\|\frac{1}{n}\sum_i \widehat{q}_i\right\| &= O_p\Big(n^{-a} + n^{-(1-\eta)/2}\sqrt{\log n}\Big),\\ \left\|\frac{1}{n}\sum_i \widehat{p}_i\right\| &= O_p\Big(n^{-b} + n^{-(1-\eta)/2}\sqrt{\log n}\Big),\\ \max_{u\in\mathcal{T}}\left|\widehat{\mu}^{(1)}(u) - \frac{1}{n}\sum_i \widehat{m}_i(u)\right| &= O_p\Big(n^{-b} + n^{-(1-\eta)/2}\sqrt{\log n}\Big).\end{aligned} \tag{C.22}$$

The largest norms of the node-level contribution to the propensity-side moment $\widehat{q}_i$ and the node-level contribution to the outcome-side moment $\widehat{p}_i$ are bounded in probability in the common normalization of the propagation directions $\widehat{U}_i$. The exact factorization (B.8) then gives

$$|\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')| = O_p\Big[\{n^{-a} + n^{-(1-\eta)/2}\sqrt{\log n}\}\{n^{-b} + n^{-(1-\eta)/2}\sqrt{\log n}\} + n^{-1}\Big], \tag{C.23}$$

because subtracting the diagonal self-products removes n bounded bilinear products $\widehat{q}_i^{\top}\widehat{\Omega}^{\dagger}\widehat{p}_i$ divided by $n(n-1)$. If $r = 0$, the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$ is zero and the calculations involving $\widehat{U}_i$ are omitted.

C.4.3 Expansion and asymptotic normality

Proof of Theorem C.1. For fixed numerical values of the plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$ and the fitted outcome predictions $\widehat{m}_i(t)$, (B.13) centers the expression containing $Y_i - \mu_i(t)$ below. The uniform bound (C.18) gives

$$\begin{aligned}&\frac{1}{n}\sum_i \mathbf{1}\{T_i = t\}\left\{\frac{1}{\widehat{\pi}_i^{\text{plug-in}}(t)} - \frac{1}{\pi_i(t)}\right\}\{Y_i - \mu_i(t)\} = O_p\Big(n^{-a}n^{-(1-\eta)/2}\sqrt{\log n}\Big),\\ &\frac{1}{n}\sum_i \{\mathbf{1}\{T_i = t\} - \pi_i(t)\}\frac{\widehat{m}_i(t) - \mu_i(t)}{\widehat{\pi}_i^{\text{plug-in}}(t)} = O_p\Big(n^{-b}n^{-(1-\eta)/2}\sqrt{\log n}\Big).\end{aligned} \tag{C.24}$$

To apply the probability-order bounds in (B.11), first restrict the numerical arrays to $\max_{i,u}|\widehat{\pi}_i^{\text{plug-in}}(u) - \pi_i(u)| \le C_1 n^{-a}$ and $\max_{i,u}|\widehat{m}_i(u) - \mu_i(u)| \le C_1 n^{-b}$, then increase the fixed C_1. The second restriction bounds the outcome-side error $\widehat{m}_i(u) - \mu_i(u)$. Choosing representatives inside nonempty covering balls preserves polynomial approximation counts.

Insert (C.24) into (B.12). In the averages of the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$ in (B.12), replacing the Hájek normalization $\widehat{h}(t)$ using the first line of (C.19) contributes $O_p(n^{-b}n^{-(1-\eta)/2}\sqrt{\log n})$. Replacing the empirical averages separately by their assignment expectations $\mathbb{E}_D$ leaves the remaining propensity–outcome product error in (B.3), bounded by $O_p(n^{-a-b})$ using (B.15). The one-array bound gives $n^{-1}\sum_i \mathbf{1}\{T_i = t\}\{Y_i - \mu_i(t)\}/\pi_i(t) = O_p(n^{-(1-\eta)/2})$. Replacing $\widehat{h}(t)$ by one in the term containing $Y_i - \mu_i(t)$ therefore costs $O_p(n^{-a}n^{-(1-\eta)/2}\sqrt{\log n} + n^{\eta-1}\log n)$. Subtracting the expressions at t and t' and adding the second-order propagation correction $\widehat{\Delta}^{(2)}_{\text{prop}}(t,t')$, bounded in (C.23), proves

$$\begin{aligned}&\widehat{\tau}_{\text{PIN}}(t,t') - \tau(t,t') - \frac{1}{n}\sum_i V_i(t,t')\\ &\quad= O_p\Big(n^{-a-b} + (n^{-a} + n^{-b})n^{-(1-\eta)/2}\sqrt{\log n} + n^{\eta-1}\log n\Big).\end{aligned} \tag{C.25}$$

The expansion uses the existing bound on the remaining propensity–outcome product error and does not require an additional rate improvement after subtracting the best $\pi_i(t)$-weighted fit $\widehat{U}_i^{\top}v$ from the outcome-side error $\widehat{m}_i(t) - \mu_i(t)$.

Each assignment contribution $V_i(t,t')$ is centered, uniformly bounded, and has the dependency graph used for (C.18). Counting nonzero covariances and using Assumption C.2 gives $c/n \le \sigma_n^2(t,t') \le Cn^{\eta-1}$. For each fixed integer $k \ge 3$, joint cumulants are bounded and vanish for indices separated into independent collections. Nonzero cumulants require connected collections, with at most $C_k n^{1+\eta(k-1)}$ ordered k-tuples: choose one index in n ways and attach each remaining distinct index along a spanning tree in at most Cn^{η} ways. Repeated indices do not increase the bound. Hence,

$$\left|\text{cum}_{D,k}\left\{\frac{\sum_i V_i(t,t')}{n\sigma_n(t,t')}\right\}\right| \le C_k n^{1+\eta(k-1)-k/2} \longrightarrow 0. \tag{C.26}$$

The exponent is negative for every $k \geq 3$ because $\eta < 1/4$. The first two normalized cumulants are zero and one. The moment–cumulant identity and the method of moments give $n^{-1}\sum_i V_i(t,t')/\sigma_n(t,t') \xrightarrow{d} N(0,1)$. Under (C.9), every term on the right of (C.25) is $o_p(n^{-1/2})$. Since $\sigma_n(t,t') \geq \sqrt{c/n}$, Slutsky's theorem proves (C.10). □

C.4.4 Conservativeness and coverage

Proof of Theorem C.2. Approximation of the centered node-level score. After multiplication by n, the part of the centered node-level score $\widehat{s}_i^{\mathrm{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}})$ containing the fitted outcome predictions $\widehat{m}_i(t)$ and $\widehat{m}_i(t')$ differs from $\mu_i(t) - \mu_i(t') - \tau(t,t')$ by at most $4\max_{k,u}|\widehat{m}_k(u) - \mu_k(u)|$. Equations (C.19) and (C.22), the lower bounds on $\pi_i(u)$ and $\widehat{\pi}_i^{\text{plug-in}}(u)$, and bounded outcomes make the part containing the residual $Y_i - \widehat{m}_i(T_i)$ differ from the assignment contribution $V_i(t,t')$ by $O_p(n^{-a} + n^{-b} + n^{-(1-\eta)/2}\sqrt{\log n})$, uniformly over i. For the two node-level contributions of the second-order propagation correction $\widehat{\Delta}^{(2)}_{\mathrm{prop}}(t,t')$, excluding node i from either $n^{-1}\sum_j \widehat{p}_j$ or $n^{-1}\sum_j \widehat{q}_j$ changes the average by $O_p(n^{-1})$. The first two lines of (C.22) therefore bound the node-level contribution with i supplying $\widehat{q}_i$ by $O_p(n^{-b} + n^{-(1-\eta)/2}\sqrt{\log n} + n^{-1})$ and the node-level contribution with i supplying $\widehat{p}_i$ by $O_p(n^{-a} + n^{-(1-\eta)/2}\sqrt{\log n} + n^{-1})$. Equation (C.23) controls the centering subtraction $2\widehat{\Delta}^{(2)}_{\mathrm{prop}}(t,t')$. Consequently,

$$\begin{aligned}\max_i \Big| n\widehat{s}_i^{\mathrm{PIN}}(t,t';\widehat{\pi}^{\text{plug-in}}) - \{\mu_i(t) - \mu_i(t') - \tau(t,t') + V_i(t,t')\}\Big| \\ = O_p\Big(n^{-a} + n^{-b} + n^{-(1-\eta)/2}\sqrt{\log n}\Big).\end{aligned} \tag{C.27}$$

Bounds for the kernel K^+. The row-sum bound for K and the replacement of negative eigenvalues give

$$\begin{aligned}\|K^+\|_{\mathrm{op}} \leq Cn^{\eta}, \qquad \|K^+\|_F^2 \leq Cn^{1+\eta}, \\ \max_{i,j}|K^+_{ij}| \leq Cn^{\eta/2}, \quad K^+ - K \succeq 0.\end{aligned} \tag{C.28}$$

For the entrywise bound, $(K^+)^2 \preceq K^2$ gives $(K^+_{ii})^2 \leq \{(K^+)^2\}_{ii} \leq \sum_j K^2_{ij} \leq Cn^{\eta}$, and positive semidefiniteness gives $|K^+_{ij}|^2 \leq K^+_{ii}K^+_{jj}$. The matrix inequality $|x^\top Wx - y^\top Wy| \leq \|W\|_{\mathrm{op}}\|x-y\|(2\|y\| + \|x-y\|)$ and (C.27) now yield

$$\begin{aligned}&\widehat{\mathrm{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\mathrm{PIN}}(t,t')\} \\ &= \frac{1}{n^2}\sum_{i,j} K^+_{ij}\{\mu_i(t) - \mu_i(t') - \tau(t,t') + V_i(t,t')\} \\ &\quad \times \{\mu_j(t) - \mu_j(t') - \tau(t,t') + V_j(t,t')\} + o_p(n^{-1}).\end{aligned} \tag{C.29}$$

The replacement error is $O_p[n^{\eta-1}\{n^{-a} + n^{-b} + n^{-(1-\eta)/2}\sqrt{\log n}\}] = o_p(n^{-1})$ under (C.11) and $\eta < 1/4$.

Concentration after replacing the centered node-level score. Let $\Sigma_n(t,t')$ denote the covariance matrix of the assignment contributions $(V_i(t,t'))_{i=1}^n$. Bounded entries and at most Cn^{η} nonzero covariances per row give $\|\Sigma_n(t,t')\|_{\mathrm{op}} \leq Cn^{\eta}$. The fourth-moment expansion gives a term $2\,\mathrm{tr}\{K^+\Sigma_n(t,t')K^+\Sigma_n(t,t')\}$ and a sum of fourth joint cumulants weighted by $K^+_{ij}K^+_{kl}$. The trace is at most $Cn^{1+3\eta}$ by (C.28). Only connected quadruples contribute to the cumulant sum. The connected-index count used in (C.26) gives at most $Cn^{1+3\eta}$ quadruples, including repetitions, and requires no variance lower bound. Bounded cumulants and $|K^+_{ij}K^+_{kl}| \leq Cn^{\eta}$ consequently give

$$\mathrm{Var}_D\left\{\sum_{i,j} K^+_{ij}V_i(t,t')V_j(t,t')\right\} \leq Cn^{1+4\eta}. \tag{C.30}$$

The variance of the linear contribution $2\sum_{i,j}K^+_{ij}\{\mu_i(t) - \mu_i(t') - \tau(t,t')\}V_j(t,t')$ is at most

$$4\|\Sigma_n(t,t')\|_{\mathrm{op}}\|K^+\|^2_{\mathrm{op}}\sum_i\{\mu_i(t) - \mu_i(t') - \tau(t,t')\}^2 \leq Cn^{1+3\eta}.$$

The contribution containing only the fixed exposure-specific means $\mu_i(t)$ and $\mu_i(t')$ is deterministic. Applying $\mathrm{Var}(Z+Z') \leq 2\,\mathrm{Var}(Z) + 2\,\mathrm{Var}(Z')$ and dividing by n^4 bounds the variance of the double sum in (C.29) by $Cn^{-3+4\eta}$. Chebyshev's inequality makes the difference between the double sum and its assignment expectation $\mathbb{E}_D$ equal to $O_p(n^{-3/2+2\eta}) = o_p(n^{-1})$.

The assignment expectation and the variance lower bound. Since $\mathbb{E}_D V_i(t,t') = 0$, the assignment expectation in (C.29) consists of the covariance sum weighted by K^+ and the quadratic form of $\mu_i(t) - \mu_i(t') - \tau(t,t')$. Covariances vanish when $\ell_A(i,j) > 2R$, and $b_n \geq 2R$ gives

$$\frac{1}{n^2}\sum_{i,j} K_{ij}\,\mathrm{Cov}_D\{V_i(t,t'), V_j(t,t')\} = \sigma_n^2(t,t'). \tag{C.31}$$

Subtracting (C.31) proves (C.12). The first double sum in (C.12) is nonnegative because $K^+ - K$ and $\Sigma_n(t,t')$ are positive semidefinite. The second double sum is a quadratic form of $K^+ \succeq 0$. Hence the network variance term $\widehat{\mathrm{Var}}_{\text{plug-in}}\{\widehat{\tau}_{\text{PIN}}(t,t')\}$ is at least $\sigma_n^2(t,t') + o_p(n^{-1})$. Under Assumption C.2, division by $\sigma_n^2(t,t') \geq c/n$ proves (C.13).

Coverage. The derived bounds on the Hájek normalizations $\widehat{h}(t)$, the matrix $\widehat{\Omega}^\dagger$, and the bilinear products $\widehat{q}_i^\top \widehat{\Omega}^\dagger \widehat{p}_j$ make the matrix $\widehat{R}(t,t')$ in (C.3) finite with probability tending to one. For fixed $\varepsilon \in (0,1)$, (C.6) and (C.13) give $\widehat{\mathrm{SE}}_{\text{PIN}}(t,t') \geq \sqrt{1-\varepsilon}\,\sigma_n(t,t')$ with probability tending to one. Conditions (C.11) imply (C.9), so Theorem C.1 yields

$$\begin{aligned}\liminf_{n\to\infty} \Pr\Big\{|\widehat{\tau}_{\text{PIN}}(t,t') - \tau(t,t')| \leq \Phi^{-1}(1-\alpha/2)\widehat{\mathrm{SE}}_{\text{PIN}}(t,t')\Big\} \\ \geq 2\Phi\big\{\Phi^{-1}(1-\alpha/2)\sqrt{1-\varepsilon}\big\} - 1.\end{aligned} \tag{C.32}$$

Letting ε decrease to zero proves (C.14) without requiring independence between the two terms in the propagation-integrated standard error $\widehat{\mathrm{SE}}_{\text{PIN}}(t,t')$. □

C.4.5 The remaining pairwise component

Proof of Proposition C.3. The bounds for $\widehat{q}_i$, $\widehat{p}_j$, and $\widehat{\Omega}^\dagger$ give $\max_{i\neq j} |\widehat{q}_i^\top \widehat{\Omega}^\dagger \widehat{p}_j| = O_p(1)$. Equation (C.3) implies $\|\widehat{R}(t,t')\|_F = O_p(n)$. Equations (C.5) and (C.28) then give

$$0 \leq \mathrm{tr}\{\widehat{R}(t,t')K^+\widehat{R}(t,t')K^+\} \leq \|K^+\|_{\text{op}}^2 \|\widehat{R}(t,t')\|_F^2 = O_p(n^{2+2\eta}),$$

proving (C.15) after division by $\{n(n-1)\}^2$ and taking square roots. For the Gaussian auxiliary-variance identity in (C.16), write $\xi = (K^+)^{1/2} g$ with $g \sim N(0, I_n)$. If λ_j are the eigenvalues of $(K^+)^{1/2}\widehat{R}(t,t')(K^+)^{1/2}$, orthogonal invariance gives

$$\xi^\top \widehat{R}(t,t')\xi - \mathrm{tr}\{\widehat{R}(t,t')K^+\} \overset{d}{=} \sum_j \lambda_j (g_j^2 - 1).$$

The independent $g_j^2 - 1$ have variance two, so the variance is $2\sum_j \lambda_j^2 = 2\,\mathrm{tr}\{\widehat{R}(t,t')K^+\widehat{R}(t,t')K^+\}$. Division by $\{n(n-1)\}^2$ proves (C.16). □

The guarantees require dependence on treatments within distance R, the stated restrictions on the arrays of plug-in exposure probabilities $\widehat{\pi}_i^{\text{plug-in}}(t)$, fitted outcome predictions $\widehat{m}_i(t)$, and propagation directions $\widehat{U}_i$, and the stated rates. Persistent misspecification of both $\widehat{\pi}_i^{\text{plug-in}}(t)$ and $\widehat{m}_i(t)$ and interference extending beyond every fixed radius are not covered. In particular, a linear-in-means outcome with nonzero feedback at all network distances requires an additional approximation argument before Assumption B.2 applies.

D Simulation Design and Implementation

The notation follows Sections 2 and 4. To distinguish the full synthetic network from the population used to estimate the exposure contrast, let $\mathcal{N}_N = 1, \ldots, N$, with $N = 1456$, denote the full set of simulated nodes. A fixed subset of the full network is defined below as the analysis population and is relabeled as $\mathcal{N}_n = 1, \ldots, n$, consistent with the notation in the main text.

For node i, A denotes the network, H_i denotes the pretreatment features used to generate and estimate the treatment propensity $e_i = \Pr(D_i = 1 \mid H_i)$, D_i denotes the realized treatment, and $T_i = T(i, D, A)$ denotes the realized exposure. For any network-wide assignment $\mathbf{d}$, the outcome equations below define the potential outcome $Y_i(\mathbf{d})$. The observed outcome in a Monte Carlo replication is $Y_i = Y_i(D)$. The network maps the treatment-propensity vector $e = (e_1, \ldots, e_N)^\top$ into the exposure probabilities $\pi_i(t; e, A)$.

Each node i is assigned a two-dimensional latent network position

$$\rho_i = \begin{pmatrix}\rho_{i1}\\ \rho_{i2}\end{pmatrix}, \qquad \rho_i \overset{\text{iid}}{\sim} \mathrm{Unif}([0,1]^2),$$

and a scalar pretreatment covariate

$$X_i = x_i, \qquad x_i \overset{\text{iid}}{\sim} N(0,1).$$

The latent network positions ρ_i and pretreatment covariates X_i are generated once and held fixed across all Monte Carlo replications.

Let κ denote the average observed out-degree in the two largest treated schools from Paluck et al. (2016), and define

$$r_N = \left(\frac{\kappa}{\pi N}\right)^{1/2}.$$

The synthetic network is constructed from the latent network positions as

$$A_{ij} = \mathbf{1}\left\{i \neq j,\ \|\rho_i - \rho_j\|_2 \leq r_N\right\}.$$

The resulting network is undirected and is held fixed across Monte Carlo replications. Let

$$\deg_i = \sum_{j=1}^{N} A_{ij}$$

denote the degree of node i. The treatment-eligibility indicator $E_i \in 0, 1$ is retained from the Paluck data and is also held fixed across Monte Carlo replications.

The pretreatment features are

$$H_i = \begin{pmatrix} x_i \\ \rho_{i1} \\ \rho_{i2} \\ \deg_i \end{pmatrix}.$$

For an eligible node, the baseline treatment propensity is

$$e_i^{\text{base}} = E_i \left[\text{logit}^{-1}\left\{-0.25 + 0.8x_i + 0.6\rho_{i1} - 0.5\rho_{i2} + 0.35\deg_i\right\}\right],$$

Thus, an ineligible node with $E_i = 0$ has treatment propensity zero. Conditional on the fixed pretreatment quantities, treatment assignments are independent:

$$D_i \mid H_i, E_i \sim \text{Bernoulli}(e_i), \qquad i = 1, \ldots, N.$$

At the baseline difficulty level, $e_i = e_i^{\text{base}}$. The three treatment-propensity shifts defined in Appendix D.3 modify e_i across difficulty levels while retaining the same treatment-assignment mechanism.

All three simulations use the binary spillover exposure

$$T_i = \mathbf{1}\left\{\sum_{j=1}^{N} A_{ij} D_j > 0\right\}, \qquad \mathcal{T} = \{0, 1\}.$$

The event $T_i = 1$ means that node i has at least one treated neighbor, while $T_i = 0$ means that node i has no treated neighbor. Under the conditionally independent Bernoulli assignment model, the exposure probabilities are

$$\pi_i(0; e, A) = \prod_{j: A_{ij}=1} (1 - e_j), \qquad \pi_i(1; e, A) = 1 - \pi_i(0; e, A).$$

Hence, the exposure probability of node i depends on the treatment propensities of the neighbors of node i.

The outcome disturbance is generated once as

$$\varepsilon_i = \nu_i + (\rho_{i1} - 0.5), \qquad \nu_i \overset{\text{iid}}{\sim} N(0,1),$$

and is held fixed across Monte Carlo replications. The two outcome models are defined in Appendix D.1.

Let

$$\pi_i^{\text{base}}(t) = \pi_i(t; e^{\text{base}}, A).$$

The fixed analysis population is

$$\mathcal{N}_n = \left\{ i \in \mathcal{N}_N : \sum_{j=1}^{N} A_{ij} E_j > 0,\ \min_{t \in \{0,1\}} \pi_i^{\text{base}}(t) \geq 10^{-3} \right\}.$$

The analysis population is defined once using the baseline treatment propensities and remains unchanged across difficulty levels. For the retained population, the target estimand is

$$\tau(1, 0) = \mu(1) - \mu(0), \qquad \mu(t) = \frac{1}{n} \sum_{i \in \mathcal{N}_n} \mu_i(t).$$

Thus, $\tau(1,0)$ compares exposure to at least one treated neighbor with exposure to no treated neighbors.

The latent network positions ρ_i, network A, treatment-eligibility indicators E_i, pretreatment covariates X_i, pretreatment features H_i, and outcome disturbances ε_i are generated once and held fixed across Monte Carlo replications. For each difficulty level, the treatment-propensity vector $e = (e_1, \ldots, e_N)^\top$ is fixed before the Monte Carlo replications begin. In each replication, a new treatment vector D is drawn, the exposure vector T is recomputed, and the outcomes are evaluated under the realized treatment vector. Simulations 1–3 therefore differ only in how the treatment-propensity vector e changes across difficulty levels.

D.1 Outcome Models

Let $\widetilde{A}_{ij} = A_{ij}/\deg_i$ for $\deg_i > 0$ and $\widetilde{A}_{ij} = 0$ otherwise, so that $\sum_j \widetilde{A}_{ij} d_j$ and $\sum_j \widetilde{A}_{ij} Y_j(\mathbf{d})$ denote the average treatment and average outcome among node i's neighbors, respectively. For any network-wide treatment assignment $\mathbf{d} = (d_1, \ldots, d_N)^\top$, we consider two potential-outcome models.

The linear-in-means model:

$$Y_i(\mathbf{d}) = 0.8 \sum_{j=1}^{N} \widetilde{A}_{ij} Y_j(\mathbf{d}) + \sum_{j=1}^{N} \widetilde{A}_{ij} d_j + d_i + 3X_i + \varepsilon_i. \tag{D.1}$$

Thus, the continuous outcome depends on the node's own treatment, average treatment among its neighbors, average outcome among its neighbors, the pretreatment covariate X_i, and the disturbance ε_i. The system is solved jointly for $Y(\mathbf{d})$.

The complex-contagion model, let $s = 0, 1, 2, \ldots$ index the iterative updates of the binary outcomes. We initialize the outcomes at (s=0), and then repeatedly update each node using the neighboring outcomes from the preceding iteration until the outcome vector no longer changes:

$$Y_i^{(0)}(\mathbf{d}) = \mathbf{1}\left\{ -1 + \sum_{j=1}^{N} \widetilde{A}_{ij} d_j + d_i + 3X_i + \varepsilon_i \ \geq \ 0 \right\},$$

and for $s \geq 1$ update

$$Y_i^{(s)}(\mathbf{d}) = \mathbf{1}\left\{ -1 + 1.5 \sum_{j=1}^{N} \widetilde{A}_{ij} Y_j^{(s-1)}(\mathbf{d}) + \sum_{j=1}^{N} \widetilde{A}_{ij} d_j + d_i + 3X_i + \varepsilon_i \ \geq \ 0 \right\}.$$

The iterations continue until $Y^{(s)}(\mathbf{d}) = Y^{(s-1)}(\mathbf{d})$. The resulting stable binary vector defines $Y(\mathbf{d})$. In each Monte Carlo replication, the observed outcome is $Y_i = Y_i(D)$ under the realized treatment vector D.

D.2 Regression Specifications

Following Gao and Ding (2025), we consider three regression specifications, denoted by Unadj, Add, and Sat. The Unadj, Add, and Sat specifications determine how the pretreatment covariates X_i enter the outcome regression and do not change the potential-outcome models in Appendix D.1. Define the centered pretreatment covariates as

$$\widetilde{X}_i = X_i - \frac{1}{n} \sum_{j \in \mathcal{N}_n} X_j, \qquad i \in \mathcal{N}_n.$$

The covariates X_i are centered over the fixed analysis population $\mathcal{N}_n$, rather than separately within the realized exposure groups.

Unadj. The Unadj specification includes an intercept α_t for each exposure t but does not include the pretreatment covariates $\widetilde{X}_i$:

$$Y_i \approx \sum_{t \in \mathcal{T}} \mathbf{1}\{T_i = t\} \alpha_t.$$

Unadj omits X_i from the outcome regression, not H_i from treatment-propensity estimation.

Add. The Add specification includes the centered pretreatment covariates $\widetilde{X}_i$ with a coefficient vector β shared across all exposures:

$$Y_i \approx \sum_{t \in \mathcal{T}} \mathbf{1}\{T_i = t\} \alpha_t + \beta^\top \widetilde{X}_i.$$

The intercept α_t can differ across exposures, while the coefficient vector β is the same at every exposure.

Sat. The Sat specification allows both the intercept α_t and the coefficient vector β_t to differ across exposures:

$$Y_i \approx \sum_{t \in \mathcal{T}} \mathbf{1}\{T_i = t\} \left\{ \alpha_t + \beta_t^\top \widetilde{X}_i \right\}.$$

Unlike Add, Sat allows the relationship between $\widetilde{X}_i$ and Y_i to vary with the exposure T_i.

D.3 Treatment-Propensity Shifts

Simulations 1–3 use the same network A, pretreatment features H_i, outcome models, and treatment-assignment mechanism, and differ only in how the treatment-propensity vector $e = (e_1, \ldots, e_N)^\top$ changes across difficulty levels. Starting from the same baseline treatment propensities e_i^{base}, Simulation 1 uses a Treatment-propensity logit-scale shift, Simulation 2 uses a Fixed-overlap nonlinear rank shift, and Simulation 3 uses a Fixed-overlap network-aligned rank shift. The three shifts are constructed to change which nodes are more or less likely to receive treatment while holding the remaining simulation components fixed.

D.3.1 Simulation 1: Treatment-Propensity Logit-Scale Shift

Simulation 1 changes the spread of the node-level treatment propensities while preserving the average treatment propensity among eligible nodes. For a logit multiplier $\lambda \geq 1$, define

$$e_i^{(\lambda)} = E_i \, \text{logit}^{-1}\{\alpha_\lambda + \lambda \, (0.8 x_i + 0.6\rho_{i1} - 0.5\rho_{i2} + 0.35 \deg_i)\}.$$

Here, λ controls the spread of the treatment propensities, while α_λ is the intercept chosen separately for each value of λ to preserve the average treatment propensity among eligible nodes.

Let

$$\bar{e}^{\text{base}} = \frac{\sum_{i=1}^N E_i e_i^{\text{base}}}{\sum_{i=1}^N E_i}.$$

For each value of λ, α_λ is selected to satisfy

$$\frac{\sum_{i=1}^N E_i e_i^{(\lambda)}}{\sum_{i=1}^N E_i} = \bar{e}^{\text{base}}.$$

At $\lambda = 1$, setting $\alpha_1 = -0.25$ recovers the baseline treatment propensities:

$$e_i^{(1)} = e_i^{\text{base}}.$$

Increasing λ increases the differences in treatment propensity across eligible nodes: nodes with larger baseline treatment-assignment values receive larger treatment propensities, while nodes with smaller baseline treatment-assignment values receive smaller treatment propensities. The adjustment through α_λ prevents this increase in dispersion from changing the average treatment propensity among eligible nodes. The reported difficulty levels use

$$\lambda \in \{1.0, 1.1, 1.2, 1.3, 1.4, 1.5\}.$$

D.3.2 Simulation 2: Fixed-Overlap Nonlinear Rank Shift

Simulation 2 keeps the collection of baseline treatment-propensity values fixed but changes which eligible nodes receive the larger and smaller treatment propensities. For nodes with $\deg_i > 0$, define the neighbor-averaged pretreatment covariate as

$$\bar{X}_i = \frac{\sum_{j=1}^N A_{ij} X_j}{\deg_i},$$

with $\bar{X}_i = 0$ when $\deg_i = 0$.

Using $\bar{X}_i$, define

$$g_i = \bar{X}_i + 0.5\bar{X}_i^3 + 0.5\bar{X}_i \left(\rho_{i1}^2 + \rho_{i2}^2\right).$$

Among eligible nodes, let

$$(a, b_1, b_2, b_3) = \underset{a, b_1, b_2, b_3}{\arg\min} \sum_{i=1}^N E_i \left(g_i - a - b_1\rho_{i1} - b_2\rho_{i2} - b_3 \deg_i\right)^2.$$

The nonlinear ranking component is the corresponding residual

$$r_i^{\text{NL}} = g_i - a - b_1\rho_{i1} - b_2\rho_{i2} - b_3 \deg_i.$$

Thus, r_i^{NL} is the part of the nonlinear combination g_i that is not linearly represented by ρ_{i1}, ρ_{i2}, and $\deg_i$.

Recall that the nonintercept part of the baseline treatment-propensity logit is

$$0.8x_i + 0.6\rho_{i1} - 0.5\rho_{i2} + 0.35\deg_i .$$

For each $\nu > 0$, the baseline treatment propensities among eligible nodes are sorted from smallest to largest, and eligible nodes are separately sorted according to

$$0.8x_i + 0.6\rho_{i1} - 0.5\rho_{i2} + 0.35\deg_i + \nu r_i^{\mathrm{NL}}.$$

The smallest baseline treatment propensity is assigned to the first node in this ordering, the second-smallest baseline treatment propensity is assigned to the second node, and the procedure continues until every eligible node receives one baseline treatment-propensity value. Therefore,

$$\left\{\left\{e_i^{(\nu)} : E_i = 1\right\}\right\} = \left\{\left\{e_i^{\mathrm{base}} : E_i = 1\right\}\right\} .$$

Changing ν therefore changes which eligible nodes receive the larger and smaller treatment propensities without changing the collection or marginal distribution of treatment-propensity values among eligible nodes. The reported difficulty levels use

$$\nu \in \{0, 0.25, 0.40\}.$$

D.3.3 Simulation 3: Fixed-Overlap Network-Aligned Rank Shift

Simulation 3 also keeps the collection of baseline treatment-propensity values fixed but changes which eligible nodes receive the larger and smaller treatment propensities according to neighboring pretreatment information. For nodes with $\deg_i > 0$, define the neighbor-averaged pretreatment covariate as

$$\bar{X}_i = \frac{\sum_{j=1}^{N} A_{ij} X_j}{\deg_i},$$

with $\bar{X}_i = 0$ when $\deg_i = 0$.

Recall that the nonintercept part of the baseline treatment-propensity logit is

$$q_i = 0.8x_i + 0.6\rho_{i1} - 0.5\rho_{i2} + 0.35\deg_i .$$

Among eligible nodes, regress the neighbor-averaged pretreatment covariate $\bar{X}_i$ on an intercept and the baseline treatment-assignment expression q_i. Let

$$(a, b) = \arg\min_{a,b} \sum_{i=1}^{N} E_i \left(\bar{X}_i - a - bq_i\right)^2 .$$

The network-aligned ranking component is the corresponding residual

$$r_i^{\mathrm{NET}} = \bar{X}_i - a - bq_i.$$

Thus, r_i^{NET} is the part of the neighbor-averaged pretreatment covariate that is not linearly represented by the baseline treatment-assignment expression.

For rotation strength ω, eligible nodes are ordered according to

$$\sqrt{1-\omega^2}\, q_i + \omega r_i^{\mathrm{NET}}.$$

Thus, ω controls the rotation from the baseline treatment-propensity ordering toward the neighboring pretreatment information represented by r_i^{NET}. At $\omega = 0$, the ordering is determined entirely by q_i. As ω increases, r_i^{NET} receives greater weight, while the weight on q_i decreases.

For each $\omega > 0$, the baseline treatment propensities among eligible nodes are sorted from smallest to largest, and eligible nodes are separately sorted according to the ordering above. The smallest baseline treatment propensity is assigned to the first node in this ordering, the second-smallest baseline treatment propensity is assigned to the second node, and the procedure continues until every eligible node receives one baseline treatment-propensity value. Therefore,

$$\left\{\left\{e_i^{(\omega)} : E_i = 1\right\}\right\} = \left\{\left\{e_i^{\mathrm{base}} : E_i = 1\right\}\right\} .$$

Changing ω therefore changes where the larger and smaller treatment propensities are located in the network without changing the collection or marginal distribution of treatment-propensity values among eligible nodes. The reported difficulty levels use

$$\omega \in \{0, 0.15, 0.30\}.$$

E REAL-NETWORK SEMI-SYNTHETIC STUDY

Following Gao and Ding (2025), let $\mathcal{N}_N = \{1, \ldots, N\}$, with $N = 3306$, denote the full set of students in the five largest treated schools from Paluck et al. (2016). A fixed subset of the full network is defined below as the analysis population and is denoted by $\mathcal{N}_n \subseteq \mathcal{N}_N$.

For student i, A denotes the observed school network, X_i denotes the pretreatment variables used for outcome-regression adjustment, H_i denotes the pretreatment features used to generate and estimate the treatment propensity $e_i = \Pr(D_i = 1 \mid H_i)$, and $E_i \in \{0, 1\}$ denotes treatment eligibility. The realized treatment is denoted by D_i. For any network-wide treatment assignment $\mathbf{d}$, the outcome equation below defines the potential outcome $Y_i(\mathbf{d})$, and the observed outcome in a Monte Carlo replication is $Y_i = Y_i(D)$.

The observed network A, pretreatment variables X_i and H_i, and treatment-eligibility indicators E_i are retained from the Paluck data and held fixed across Monte Carlo replications. The original Paluck treatment assignments and observed outcomes are not used as D_i or Y_i. For each difficulty level, new treatment assignments are generated from the treatment-assignment mechanism defined below, the corresponding exposures are recomputed, and the synthetic outcomes are evaluated under the realized treatment vector. The observed school network is represented by the directed adjacency matrix $A = (A_{ij})$, where $A_{ij} = 1$ indicates that student i named student j as a student with whom student i had recently spent time. Students could nominate up to ten students within the same school, and links across different schools are set to zero. The directed network is held fixed across all difficulty levels and Monte Carlo replications.

Let

$$\deg_i = \sum_{j=1}^{N} A_{ij}$$

denote the observed out-degree of student i. The outcome-adjustment variables are

$$X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \end{pmatrix},$$

where X_{i1} and X_{i2} denote gender and grade, respectively. The pretreatment features used for treatment-propensity generation and estimation are

$$H_i = \left(\text{Gender}_i,\ \text{Grade}_i,\ \text{SchoolDummies}_i^\top,\ \text{BlockDummies}_i^\top,\ \deg_i\right)^\top.$$

School membership and randomization-block membership are one-hot encoded, with one level omitted from each set to avoid duplicating the intercept.

The distinction between X_i and H_i reflects their different roles: X_i contains the pretreatment variables used for outcome-regression adjustment, while H_i contains the pretreatment features used to generate and estimate treatment propensities. No realized treatment, exposure, or outcome is included in either X_i or H_i.

Starting from these pretreatment features, define

$$r_i = 0.45\,\text{Gender}_i - 0.35\,\text{Grade}_i + 0.85\,\deg_i + \sum_{a \in \mathcal{A}_{\text{school}}} 0.25\sin(a)\,\text{School}_{ia} + \sum_{b \in \mathcal{A}_{\text{block}}} 0.35\cos(b)\,\text{Block}_{ib}.$$

To construct the treatment-propensity shift, we use r_i to form the nonlinear outcome-prognostic component

$$G_i = \sin(r_i) + 0.35\left[r_i^2 - \frac{1}{N}\sum_{k=1}^{N} r_k^2\right].$$

The component G_i enters the synthetic potential outcome defined below and therefore creates outcome-relevant heterogeneity.

We then construct S_i from the part of G_i that can be represented by H_i after removing the linear component represented by X_i. Among eligible students, let

$$(a_H, b_H) = \arg\min_{a,b} \sum_{i=1}^{N} E_i \left(G_i - a - b^\top H_i\right)^2,$$

and

$$(a_X, b_X) = \arg\min_{a,b} \sum_{i=1}^{N} E_i \left(a_H + b_H^\top H_i - a - b^\top X_i\right)^2.$$

Define

$$S_i = a_H + b_H^\top H_i - a_X - b_X^\top X_i, \qquad E_i = 1.$$

The score S_i is determined entirely by pretreatment information and is related to the generated outcome through G_i, while excluding the linear component represented by the outcome-adjustment variables X_i.

Let $\zeta \geq 0$ denote the treatment-propensity shift. The treatment propensity of student i at shift level ζ is

$$e_i^{(\zeta)} = E_i \left[\text{logit}^{-1}\left\{\zeta S_i\right\}\right].$$

For each value of ζ. At $\zeta = 0$, eligible students have the same treatment propensity, while increasing ζ assigns larger treatment propensities to students with larger S_i and smaller treatment propensities to students with smaller S_i. The reported shift levels are

$$\zeta \in \{0, 1.0, 1.5\}.$$

Following Gao and Ding (2025), we consider one-dimensional and two-dimensional exposure mappings $T_i = T(i, D, A)$. The one-dimensional analyses use either the direct exposure $T_i = D_i$, indicating whether student i receives treatment, or the binary spillover exposure $T_i = \mathbf{1}\{\sum_{j=1}^N A_{ij} D_j > 0\}$, indicating whether student i has at least one treated neighbor. Each one-dimensional exposure mapping has exposure set $\mathcal{T} = \{0, 1\}$. The two-dimensional exposure mapping jointly records own treatment and binary spillover exposure as $T_i = (D_i, \mathbf{1}\{\sum_{j=1}^N A_{ij} D_j > 0\})$, with exposure set $\mathcal{T} = \{(0,0), (0,1), (1,0), (1,1)\}$. The first component of T_i indicates whether student i receives treatment, and the second component indicates whether student i has at least one treated neighbor.

F Additional Numerical Results

The design follows Simulation 3 in Appendix D.3.3. The same network, pretreatment features, outcome models, treatment-assignment mechanism, treatment-propensity values, and difficulty levels are used. The additional comparison changes only the set of methods included in the reported results. All other simulation components remain unchanged. The treatment-propensity shifts are 0, 1.0, and 1.5, and the same population, Monte Carlo draws, propensity-estimation procedure, and confidence level are used across the three difficulty levels. For PIN, WLS+, WLS, EHW, Leung, and Kojevnikov, the reported quantities are averaged over the Unadj, Add, and Sat specifications. Forastiere, NDR, and netAIPW are reported using their respective implementations.

Table F.1: Linear-in-Means outcome performance under observational covariate shift.

Difficulty	Method	Average coverage rate	Average |Bias|	Average CI length
Baseline (shift 0)	**Proposed**	**1.00**	**0.00**	**2.23**
	WLS+	1.00	0.01	1.92
	WLS	0.98	0.01	1.35
	EHW	0.99	0.01	1.37
	Leung	0.99	0.01	1.37
	Kojevnikov	0.97	0.01	1.22
	Forastiere	0.99	0.00	335.73
	NDR	0.95	0.03	1.11
	netAIPW	0.94	0.02	1.02
Moderate (shift 1.0)	**Proposed**	**0.99**	**0.31**	**2.40**
	WLS+	0.94	0.48	2.01
	WLS	0.75	0.48	1.41
	EHW	0.75	0.48	1.40
	Leung	0.74	0.48	1.39
	Kojevnikov	0.68	0.48	1.30
	Forastiere	1.00	1.09	1126.85
	NDR	0.86	0.25	1.26
	netAIPW	0.84	0.24	1.09
High (shift 1.5)	**Proposed**	**0.99**	**0.34**	**2.49**
	WLS+	0.82	0.74	2.11
	WLS	0.51	0.74	1.47
	EHW	0.51	0.74	1.46
	Leung	0.46	0.74	1.43
	Kojevnikov	0.43	0.74	1.35
	Forastiere	1.00	1.22	1215.20
	NDR	0.75	0.42	1.34
	netAIPW	0.72	0.39	1.20

Notes: Results use 200 paired draws, and 95% confidence intervals.