PEER EFFECTS IN SIGNED NETWORKS: SEPARATING INFLUENCE THROUGH POSITIVE AND NEGATIVE TIES

Anonymous authors
Paper under double-blind review

ABSTRACT

Evaluating network interventions requires understanding how treatment affects people through their social relationships. Counting treated neighbors without distinguishing supportive and antagonistic ties can conceal opposing influences. We define effects through positive and negative ties, their interaction, and a sign-composition effect of reallocating treatment between the two types at a fixed total, and give their identification formulas. Under sign-blind assignment, we show how ignoring signs mixes the effects of the two tie types. We propose SiDE (**Si**gned-exposure **D**oubly robust **E**stimator), which combines sign-specific outcome models with exposure probabilities induced by individual treatment assignment. We establish double robustness of its score and assess approximate intervals that account for overlapping neighborhoods. Semi-synthetic experiments on six real signed networks demonstrate accurate effect estimation and examine the limits of interval coverage. An exploratory reanalysis of published school-experiment data yields a positive estimate of the peer effect through spend-time ties on wristband wearing, but the intervals for all four effects include zero after adjustment for multiple comparisons. This framework can inform network intervention design by showing when influences through the two tie types reinforce or offset one another.

1 INTRODUCTION

Social networks record trust and distrust on online platforms (Guha et al., 2004; Leskovec et al., 2010a) and friendship and antagonism in villages (Ghasemian and Christakis, 2024). Signed networks represent these relationships with positive ($+$) and negative ($-$) ties (Cartwright and Harary, 1956; Tang et al., 2016). A *peer effect* is a change in a person's outcome caused by changing peers' treatments while holding that person's own treatment fixed (Hudgens and Halloran, 2008). This dependence on others' treatments is network interference. Peer effects arise in contexts such as village health programs (Kim et al., 2015), online experiments (Eckles et al., 2017), and school interventions (Paluck et al., 2016).

Consider a village health program in which treatment is receipt of a preventive product through the program and the outcome is subsequent use. In Figure 1(a), households A and B each have four trusted and two avoided neighbors. A's treated neighbors are three trusted households and one avoided household, and B's are four trusted households. Figure 1(b) plots these signed exposures as $(3, 1)$ and $(4, 0)$. With a household's own treatment fixed, the *positive-tie effect* and *negative-tie effect* measure the response to treating an additional trusted or avoided neighbor, respectively, keeping the other treated-neighbor count fixed. Both effects can be positive or negative. Their *interaction* asks whether treating avoided neighbors changes the effect of treating trusted neighbors. The *sign-composition effect* compares a household's responses to A's and B's exposure patterns: treatment shifts from an avoided to a trusted neighbor while four neighbors remain treated.

An analysis that counts treated neighbors without distinguishing tie signs assigns A and B the same exposure. An estimated effect near zero can then reflect weak peer influence or opposing influences that cancel out. Existing frameworks for causal inference under interference accommodate exposure mappings and contrasts between exposure states (Aronow and Samii, 2017; Qu et al., 2026). Regression and graph neural estimators adjust for individual and neighborhood covariates (Forastiere

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.

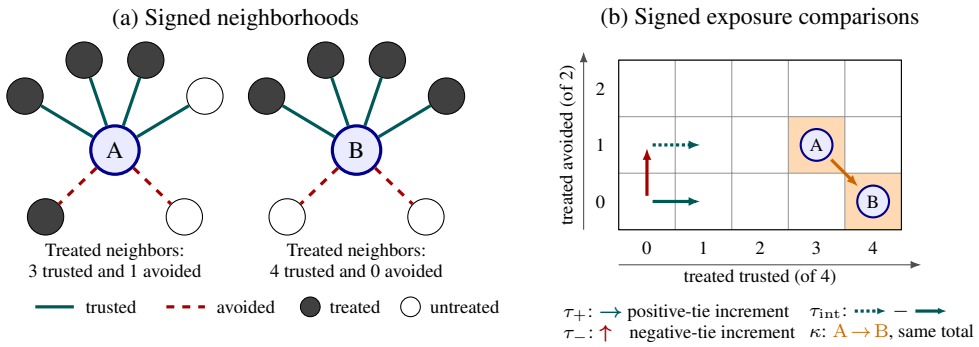

Figure 1: Exposure patterns with four treated neighbors. The grid counts treated trusted neighbors horizontally and treated avoided neighbors vertically. Mean outcome changes along horizontal and vertical arrows define the positive- and negative-tie effects. Their interaction subtracts the solid horizontal increment from the dotted one. The sign-composition effect compares outcomes under exposures A=$(3, 1)$ and B=$(4, 0)$.

et al., 2021; Ma and Tresp, 2021; Jiang and Sun, 2022; Zhao et al., 2024; Leung and Loupos, 2025). Building on these frameworks, we study effects through positive and negative ties, including the reallocation of treatment between peer types at a fixed total, and characterize the information lost when tie signs are discarded.

To construct a doubly robust score for these contrasts, we model both conditional outcomes and the joint probabilities of own treatment and the two signed exposure counts. Overlapping neighborhoods can also induce dependence between units, which must be considered when assessing uncertainty. We propose SiDE (**Si**gned-exposure **D**oubly robust **E**stimator), which combines separate encoders for positive and negative neighborhoods with exposure probabilities induced by individual treatment assignment. We also evaluate approximate intervals that account for dependence between units with overlapping neighborhoods.

Contributions. (1) We define and identify four signed peer effects: the positive-tie effect, the negative-tie effect, their interaction, and the sign-composition effect. Under sign-blind assignment, we derive a closed-form decomposition of the unsigned increment as a weighted average of positive- and negative-tie increments, with weights equal to the proportions of the two tie types. This decomposition shows when opposing increments cancel out. (2) We propose SiDE, establish the double robustness of its score, and construct approximate intervals that account for overlapping neighborhoods. (3) We evaluate SiDE against eleven baselines on semi-synthetic data built on six real signed networks and reanalyze data from an existing school field experiment to separate peer effects through students' spend-time and conflict ties.

2 Related Work

Causal inference under network interference. Potential outcomes describe an ego's outcome under specified treatments of the ego and its peers (Hudgens and Halloran, 2008). An exposure mapping summarizes peers' assignments, for example by the number of treated neighbors (Aronow and Samii, 2017). Observational analyses adjust for ego and neighbor covariates (Forastiere et al., 2021). Targeted learning updates outcome predictions using assignment probabilities (Ogburn et al., 2024), while graph neural models learn adjustments from neighborhood information (Ma and Tresp, 2021; Jiang and Sun, 2022; Zhao et al., 2024; Adhikari et al., 2026; Caljon et al., 2026). Doubly robust and AIPW estimators combine outcome and assignment models, with network-dependent variance estimates (Leung and Loupos, 2025) or type-specific treated-neighbor counts (Qu et al., 2026). Studies also distinguish tie strength (Bond et al., 2012; Rajkumar et al., 2022), edge type (Lin et al., 2023), and multiple networks (Egami, 2021). Our work builds on these foundations to estimate signed peer effects and to characterize how ignoring tie signs combines positive- and negative-tie increments under sign-blind assignment.

Signed networks. Positive and negative ties enter models of structural balance and social status (Cartwright and Harary, 1956; Leskovec et al., 2010b), trust propagation (Guha et al., 2004), and

network representation learning (Derr et al., 2018; Tang et al., 2016). Models of conflict and diffusion examine how allies and enemies affect behavior (König et al., 2017; Li et al., 2013). Empirical studies examine social influence in signed networks. Ghasemian and Christakis (2024) study information diffusion through antagonistic ties, while Bellotti et al. (2026) use cross-sectional data and a calibrated behavioral model to examine how negative ties may weaken reinforcement from positive contacts. Disliking a source need not reverse its influence (Takács et al., 2016). These findings lead us to define effects through each tie type, the interaction between the two types, and the reallocation of treatment between them at a fixed total.

3 Problem Setting

Let $V = \{1, \dots, n\}$ index the units in a directed network observed before treatment. The signed adjacency matrix $S = (S_{ij})_{i,j \in V}$ has entries $S_{ij} \in \{-1, 0, +1\}$ recording a negative tie, no tie, or a positive tie from unit i to unit j, with $S_{ii} = 0$. The two neighborhoods are $\mathcal{N}_i^+ = \{j : S_{ij} = +1\}$ and $\mathcal{N}_i^- = \{j : S_{ij} = -1\}$, with degrees $d_i^+ = |\mathcal{N}_i^+|$ and $d_i^- = |\mathcal{N}_i^-|$ and union $\mathcal{N}_i$. Each unit is an ego when its outcome is studied and a peer when it belongs to another unit's neighborhood. A peer effect runs from j's treatment to i's outcome.

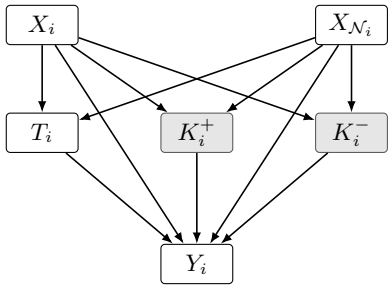

Figure 2: Causal graph for one ego. Shaded nodes are the numbers of treated neighbors within $\mathcal{N}_i^+$ and within $\mathcal{N}_i^-$.

Let $T_i \in \{0, 1\}$ be unit i's own treatment, Y_i the observed outcome, and X_i a vector of pre-treatment characteristics and network features. Write $\mathbf{T} = (T_1, \dots, T_n)$, $\mathbf{X} = (X_1, \dots, X_n)$, and $X_{\mathcal{N}_i} = (\{X_j\}_{j \in \mathcal{N}_i^+}, \{X_j\}_{j \in \mathcal{N}_i^-})$, keeping neighbor covariates separate by sign and invariant to their order. The signed exposures are $K_i^+(\mathbf{T}) = \sum_{j \in \mathcal{N}_i^+} T_j$ and $K_i^-(\mathbf{T}) = \sum_{j \in \mathcal{N}_i^-} T_j$. Together with T_i they form the joint state $Z_i = (T_i, K_i^+, K_i^-)$. Lowercase k^+ and k^- denote specified nonnegative integer values of these exposure counts. An unsigned analysis retains only $K_i^u = K_i^+ + K_i^-$ and cannot distinguish A's exposure $(3, 1)$ from B's $(4, 0)$ in Figure 1. Figure 2 shows the causal graph, and Appendix A, Table A1, collects the notation.

3.1 Assumptions

Let $Y_i(\mathbf{t})$ denote the potential outcome under assignment $\mathbf{t}$, and Y_i the outcome observed under $\mathbf{T}$. Expectations average over a uniformly drawn ego and over covariates, assignment, and potential outcomes given the fixed network. A nonempty set $\mathcal{U} \subseteq V$, defined by signed degrees, specifies the target population, and conditioning on $i \in \mathcal{U}$ restricts the average to it. We use four identification conditions (Hudgens and Halloran, 2008; Aronow and Samii, 2017; Forastiere et al., 2021).

Assumption 1 (Signed exposure mapping)**.** A neighbor's treatment affects ego i only through $K_i^+(\mathbf{t})$ and $K_i^-(\mathbf{t})$, the numbers of treated positive- and negative-tie neighbors, so that $Y_i(\mathbf{t}) = Y_i\big(t_i, K_i^+(\mathbf{t}), K_i^-(\mathbf{t})\big)$ for every assignment $\mathbf{t}$.

Under Assumption 1 we write $Y_i(t, k^+, k^-)$ for the potential outcome under own treatment t, k^+ treated positive-tie peers, and k^- treated negative-tie peers.

Assumption 2 (Consistency)**.** The observed outcome is the potential outcome at the realized exposure, $Y_i = Y_i\big(T_i, K_i^+, K_i^-\big)$.

Assumption 3 (Conditional exchangeability)**.** For every $t \in \{0, 1\}$, $k^+ \le d_i^+$, and $k^- \le d_i^-$, $Y_i(t, k^+, k^-) \perp\!\!\!\perp (T_i, K_i^+, K_i^-) \mid X_i, X_{\mathcal{N}_i}$. Conditional on ego and neighbor covariates, the joint state carries no further information about potential outcomes.

Assumption 4 (Positivity)**.** Every state used in a reported contrast has positive conditional probability, $\Pr(T_i = t, K_i^+ = k^+, K_i^- = k^- \mid X_i, X_{\mathcal{N}_i}) > 0$, at almost every covariate value in its target population.

3.2 THE FOUR EFFECTS

We now formalize the four effects introduced in Section 1. Treated-neighbor counts cannot exceed the corresponding degrees, giving the reachable set $\mathcal{V}(k^+, k^-) = \{i : d_i^+ \geq k^+,\ d_i^- \geq k^-\}$. Each effect compares potential outcomes over a common nonempty population that can reach every compared state. Appendix A.1 gives household examples.

Definition 1 (Positive-tie peer effect)**.** For own treatment t, negative-tie exposure k^-, and positive-tie levels $k_0^+ < k_1^+$,

$$\tau_+\big(k_0^+ \to k_1^+;\, t, k^-\big) = \mathbb{E}\big[\, Y_i(t, k_1^+, k^-) - Y_i(t, k_0^+, k^-) \;\big|\; i \in \mathcal{V}(k_1^+, k^-)\big]\,. \tag{1}$$

This measures the effect of increasing positive-tie exposure, holding own treatment and negative-tie exposure fixed. The negative-tie effect reverses the roles of the two exposures.

Definition 2 (Negative-tie peer effect)**.** For own treatment t, positive-tie exposure k^+, and negative-tie levels $k_0^- < k_1^-$,

$$\tau_-\big(k_0^- \to k_1^-;\, t, k^+\big) = \mathbb{E}\big[\, Y_i(t, k^+, k_1^-) - Y_i(t, k^+, k_0^-) \;\big|\; i \in \mathcal{V}(k^+, k_1^-)\big]\,. \tag{2}$$

The subscripts $+$ and $-$ in τ_+ and τ_- name the two tie types. Either effect can be positive, zero, or negative. Their interaction measures whether one exposure changes the effect of the other.

Definition 3 (Interaction between tie types)**.** For own treatment t, positive-tie levels $k_0^+ < k_1^+$, and negative-tie levels $k_0^- < k_1^-$,

$$\begin{aligned}\tau_{\text{int}}\big(k_0^+ \to k_1^+,\, k_0^- \to k_1^-;\, t\big) = \mathbb{E}\Big[\, &\big\{Y_i(t, k_1^+, k_1^-) - Y_i(t, k_0^+, k_1^-)\big\} \\ &- \big\{Y_i(t, k_1^+, k_0^-) - Y_i(t, k_0^+, k_0^-)\big\} \;\Big|\; i \in \mathcal{V}(k_1^+, k_1^-)\Big].\end{aligned} \tag{3}$$

A negative value means the positive-tie increment is smaller when negative-tie exposure is higher.

Definition 4 (Sign-composition effect)**.** For own treatment t and levels $k^+ \geq 0$ and $k^- \geq 1$,

$$\kappa\big(k^+, k^-;\, t\big) = \mathbb{E}\big[\, Y_i(t, k^+ + 1, k^- - 1) - Y_i(t, k^+, k^-) \;\big|\; i \in \mathcal{V}(k^+ + 1, k^-)\big]\,. \tag{4}$$

This contrast measures the effect of moving treatment from a negative-tie to a positive-tie neighbor at a fixed total, as from A to B in Figure 1. Both endpoints are reachable on $\mathcal{V}(k^+ + 1, k^-)$. There it equals the positive-tie increment minus the negative-tie increment from $(t, k^+, k^- - 1)$. The treated-neighbor total alone cannot express this comparison.

3.3 IDENTIFICATION

Let $Q_i = (X_i, X_{\mathcal{N}_i})$ collect ego i's adjustment covariates, and define the observed outcome regression by

$$m(t, k^+, k^-, q) = \mathbb{E}[Y_i \mid T_i = t, K_i^+ = k^+, K_i^- = k^-, Q_i = q]. \tag{5}$$

Proposition 1 (Identification)**.** *Under Assumptions 1–4, for any state (t, k^+, k^-) and any nonempty degree-defined set $\mathcal{U} \subseteq \mathcal{V}(k^+, k^-)$ on which this state satisfies Assumption 4,*

$$\mathbb{E}[Y_i(t, k^+, k^-) \mid i \in \mathcal{U}] = \mathbb{E}[m(t, k^+, k^-, Q_i) \mid i \in \mathcal{U}]. \tag{6}$$

Conditional exchangeability and consistency equate the conditional means of potential and observed outcomes, and averaging over Q_i gives Eq. (6). Appendix B proves Proposition 1 and gives the four identified effects in Corollary 1, replacing each potential outcome in Definitions 1–4 by m over the stated common population.

3.4 WHAT AN UNSIGNED ANALYSIS ESTIMATES

Fix a population with signed degrees (a, b), where $a, b \geq 1$ and $N = a + b$, and write $\mu(t, k^+, k^-) = \mathbb{E}[Y_i(t, k^+, k^-) \mid d_i^+ = a, d_i^- = b]$. An unsigned analysis adjusts for the same covariates but retains only the total K_i^u, targeting

$$\mu^u(t, k) = \mathbb{E}\big[\mathbb{E}[Y_i \mid T_i = t, K_i^u = k, Q_i] \mid d_i^+ = a, d_i^- = b\big]. \tag{7}$$

Proposition 2 (The unsigned peer effect combines the two tie types)**.** *Under Assumptions 1 to 4 and sign-blind assignment (uniform neighbor subsets conditional on own treatment, total exposure, and covariates), for $0 \le k < N$ such that $(T_i, K_i^u) = (t, k)$ and $(t, k+1)$ both have positive probability given Q_i,*

$$\mu^u(t, k+1) - \mu^u(t, k) = \frac{a}{N}\,\delta^+(t, k) + \frac{b}{N}\,\delta^-(t, k), \tag{8}$$

where $\delta^+(t,k) = \mathbb{E}[\mu(t, R_k^+ + 1, k - R_k^+) - \mu(t, R_k^+, k - R_k^+)]$ is the effect of one more treated positive-tie neighbor, averaged over the compositions an ego reaches at total k, with R_k^+ hypergeometric with population $N-1$, $a-1$ successes, and k draws, and $\delta^-(t,k) = \mathbb{E}[\mu(t, R_k^-, k - R_k^- + 1) - \mu(t, R_k^-, k - R_k^-)]$ is its counterpart for one more treated negative-tie neighbor, with R_k^- hypergeometric with population $N-1$, a successes, and k draws. No restriction is placed on μ.

Appendix B proves Proposition 2. The two δ terms average tie-type increments over their respective composition distributions within the degree stratum. Their weighted sum can vanish when $a\delta^+(t,k) = -b\delta^-(t,k)$ even if both tie types have large effects. Thus a zero unsigned effect need not imply absence of peer influence.

4 Estimation and Inference

SiDE predicts outcomes at the compared states and corrects them using observed prediction errors. Section 4.1 weights these errors by inverse state probabilities, giving rarer states greater weight. Section 4.2 fits the models and averages scores, and Section 4.3 estimates uncertainty from overlapping neighborhoods. Figure A1 and Algorithm 1 in Appendix C summarize the procedure.

4.1 Exposure propensities and doubly robust score

Each effect in Definitions 1–4 is a *contrast*: a weighted difference of mean potential outcomes. Write $z = (t, k^+, k^-)$, $Y_i(z) = Y_i(t, k^+, k^-)$, and use $m(z, q)$ from Eq. (5). Effect C has states $\mathcal{Z}_C$, coefficients c_z, and common reachable set $\mathcal{V}_C = \mathcal{V}(\max_z k^+, \max_z k^-)$ of size n_C:

$$\tau_C = \mathbb{E}\Big[\sum_{z \in \mathcal{Z}_C} c_z Y_i(z) \;\Big|\; i \in \mathcal{V}_C\Big]. \tag{9}$$

For the interaction in Definition 3, the four coefficients are $+1, -1, -1, +1$ in the order of Eq. (3). The score below applies the same coefficients to outcome predictions and residual corrections.

We model the binary treatments as mutually independent conditional on $(S, \mathbf{X})$, with $p_j = \Pr(T_j = 1 \mid S, \mathbf{X}) \in (0, 1)$. Since $\{i\}$, $\mathcal{N}_i^+$, and $\mathcal{N}_i^-$ are disjoint, this assignment model gives

$$e_i(z) = \Pr(Z_i = z \mid S, \mathbf{X}) = p_i^t (1 - p_i)^{1-t}\, \mathrm{PB}\big(k^+; \{p_j\}_{j \in \mathcal{N}_i^+}\big)\, \mathrm{PB}\big(k^-; \{p_j\}_{j \in \mathcal{N}_i^-}\big), \tag{10}$$

where $\mathrm{PB}(k; \{p_j\})$ is the Poisson–binomial probability that independent Bernoulli variables with success probabilities $\{p_j\}$ sum to k (Hong, 2013). Let q_i be the observed value of Q_i. Conditional on $(S, \mathbf{X})$, hold the candidate functions $(\bar{m}, \bar{e})$ constant as treatments and outcomes vary. The AIPW score (Robins et al., 1994; Cattaneo, 2010) is

$$\varphi_C(O_i; \bar{m}, \bar{e}) = \sum_{z \in \mathcal{Z}_C} c_z \left[\bar{m}(z, q_i) + \frac{\mathbf{1}\{Z_i = z\}}{\bar{e}_i(z)}\big(Y_i - \bar{m}(z, q_i)\big)\right], \tag{11}$$

where $O_i = (Y_i, Z_i, q_i)$. In Eq. (11), the prediction error $Y_i - \bar{m}(z, q_i)$ is weighted by $\mathbf{1}\{Z_i = z\}/\bar{e}_i(z)$ to correct the outcome prediction.

Proposition 3 (Double robustness)**.** *Let Assumptions 1 and 2 hold. Suppose $e_i(z) > 0$ and $\mathbb{E}[Y_i(z) \mid Z_i = z, S, \mathbf{X}] = \mathbb{E}[Y_i(z) \mid S, \mathbf{X}] = m(z, q_i)$ for the states of C. Given $(S, \mathbf{X})$, take $(\bar{m}, \bar{e})$ to be nonrandom, with $\bar{e}_i(z)$ bounded away from zero. If $\bar{m} = m$ or $\bar{e}_i = e_i$, then $\mathbb{E}[\varphi_C(O_i; \bar{m}, \bar{e}) \mid S, \mathbf{X}] = \sum_z c_z\, m(z, q_i)$ for every ego of $\mathcal{V}_C$. The average score over $\mathcal{V}_C$ then has mean $\tau_C^q = n_C^{-1} \sum_{i \in \mathcal{V}_C} \sum_z c_z\, m(z, q_i)$ given the covariates, and the mean of τ_C^q over the covariates is τ_C.*

Appendix B proves Proposition 3 by taking the conditional expectation of the weighted residual in Eq. (11).

4.2 SIGN-SPECIFIC NEIGHBORHOOD ENCODER AND ESTIMATOR

We implement SiDE's outcome and assignment models with separate encoders. Each uses pooling invariant to neighbor order (Zaheer et al., 2017) and distinguishes positive and negative neighborhoods, drawing on signed-network representation learning (Derr et al., 2018):

$$h_i = \rho\Big(X_i,\ \frac{1}{\max(d_i^+,1)}\sum_{j\in\mathcal{N}_i^+} g^+(X_j),\ \frac{1}{\max(d_i^-,1)}\sum_{j\in\mathcal{N}_i^-} g^-(X_j)\Big), \tag{12}$$

where g^+, g^-, and ρ are multilayer perceptrons. The outcome encoder gives h_i, and the assignment encoder has the same form with its own $g^\pm$ and ρ and gives $\tilde{h}_i$. The outcome head combines h_i with embeddings of the two exposure levels through main terms and a bilinear interaction. The assignment head uses a logistic function of $\tilde{h}_j$ to predict p_j, and Eq. (10) turns these predictions into exposure propensities.

SiDE partitions target egos into F folds and fits both models outside each fold before scoring its egos. With these held-out predictions, the unnormalized estimate is

$$\check{\tau}_C = n_C^{-1}\sum_{i\in\mathcal{V}_C}\varphi_C(O_i;\hat{m},\hat{e}). \tag{13}$$

Normalizing the residual weights gives the single-partition estimate $\tilde{\tau}_C$, and SiDE reports the median $\hat{\tau}_C$ of $\tilde{\tau}_C$ over P fold partitions, as in Algorithm 1.

4.3 VARIANCE ESTIMATION AND INFERENCE

Define the closed neighborhood $\bar{\mathcal{N}}_i = \{i\}\cup\mathcal{N}_i$ of every ego, and call i, j linked, written $i\sim j$, if $\bar{\mathcal{N}}_i\cap\bar{\mathcal{N}}_j\neq\varnothing$. Suppose that, given $(S,\mathbf{X})$, unit-level disturbances are mutually independent and independent of treatments, and each potential outcome depends only on disturbances in its closed neighborhood. Under the conditional mean condition of Proposition 3, the residuals using the true m, e,

$$r_i = \sum_{z\in\mathcal{Z}_C} c_z\frac{\mathbf{1}\{Z_i=z\}}{e_i(z)}\{Y_i - m(z,q_i)\}, \tag{14}$$

have mean zero and are independent for unlinked egos conditional on $(S,\mathbf{X},\mathbf{T})$. Replacing m, e by their held-out estimates gives $\hat{r}_i$ and the linked-pair variance estimate

$$\tilde{\Omega}_C = \frac{\sum_{i,j\in\mathcal{V}_C:\, i\sim j}\hat{r}_i\hat{r}_j}{n_C^2}, \tag{15}$$

where pairs include $i=j$ and $\hat{r}_i$ uses the same normalized inverse-probability weights as $\tilde{\tau}_C$, as given in Appendix C.

5 EXPERIMENTS

In this section, we first evaluate SiDE on semi-synthetic data built on six real signed networks and compare it with eleven baselines. Then, we examine the role of tie signs, the sensitivity to the hidden width of the neural networks, and the computational cost, followed by an ablation study and an evaluation of interval coverage. Finally, we reanalyze data from a published school field experiment. Unless stated otherwise, each semi-synthetic experiment is repeated 10 times, and we report the average and the standard deviation. Detailed settings are provided in Appendix D.

5.1 EXPERIMENTAL SETUP

Datasets and settings. We use six real signed networks: Bitcoin-Alpha, Bitcoin-OTC, Wiki-Elec, Slashdot, Epinions, and Reddit. The two Bitcoin networks record ratings between traders (Kumar et al., 2016). Wiki-Elec, Slashdot, and Epinions record votes, friend/foe relations, and trust/distrust, respectively (Leskovec et al., 2010a), and Reddit records signed hyperlinks between communities (Kumar et al., 2018a). Following prior evaluations of network causal estimators (Ma and Tresp,

Table 1: MAE on four networks for methods that use tie signs, with standard deviations of the absolute error below. Bold and underline mark the best and second-best values.

Method	Slashdot				Epinions				Reddit				Wiki-Elec			
	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ
Sat. reg.	0.059 ±0.033	0.123 ±0.048	0.068 ±0.060	0.180 ±0.042	0.056 ±0.021	0.123 ±0.048	0.044 ±0.049	0.178 ±0.055	0.048 ±0.034	0.139 ±0.120	0.148 ±0.078	0.156 ±0.108	0.064 ±0.041	0.157 ±0.089	0.143 ±0.075	0.208 ±0.128
GPS	0.100 ±0.065	1.532 ±0.074	0.270 ±0.066	1.579 ±0.077	0.109 ±0.032	2.191 ±0.065	0.405 ±0.090	2.331 ±0.064	0.880 ±0.193	2.490 ±0.243	0.384 ±0.153	2.540 ±0.297	0.154 ±0.079	0.717 ±0.167	0.235 ±0.146	0.760 ±0.154
TARNet	0.042 ±0.018	0.045 ±0.029	0.048 ±0.029	0.099 ±0.030	0.019 ±0.010	0.053 ±0.017	0.170 ±0.029	0.082 ±0.031	**0.032** ±0.026	0.532 ±0.083	0.298 ±0.041	0.473 ±0.081	0.046 ±0.033	0.065 ±0.038	0.161 ±0.023	0.191 ±0.063
CFR	0.032 ±0.017	**0.036** ±0.024	**0.039** ±0.023	0.108 ±0.038	0.017 ±0.011	0.017 ±0.010	0.106 ±0.028	0.075 ±0.024	0.036 ±0.027	0.319 ±0.072	0.240 ±0.037	0.324 ±0.066	0.043 ±0.032	0.098 ±0.064	0.176 ±0.033	0.185 ±0.088
SAGE-HSIC	0.134 ±0.053	0.097 ±0.042	0.162 ±0.096	0.259 ±0.057	0.073 ±0.026	0.226 ±0.047	0.096 ±0.081	0.298 ±0.069	0.440 ±0.074	0.312 ±0.279	0.245 ±0.182	0.808 ±0.312	0.131 ±0.029	0.148 ±0.090	**0.107** ±0.070	0.288 ±0.099
NetEst	0.084 ±0.043	0.444 ±0.038	0.081 ±0.048	0.709 ±0.079	0.065 ±0.019	0.383 ±0.073	0.300 ±0.044	0.614 ±0.090	0.042 ±0.029	0.788 ±0.113	0.348 ±0.035	0.835 ±0.112	0.067 ±0.035	0.329 ±0.107	0.112 ±0.077	0.485 ±0.112
HINITE	0.060 ±0.028	0.628 ±0.058	0.369 ±0.048	0.559 ±0.042	0.067 ±0.015	0.470 ±0.040	0.407 ±0.029	0.451 ±0.024	0.033 ±0.020	1.032 ±0.085	0.415 ±0.025	0.890 ±0.077	**0.024** ±0.020	0.730 ±0.142	0.390 ±0.042	0.468 ±0.116
DWR	0.056 ±0.031	0.582 ±0.062	0.230 ±0.092	0.528 ±0.088	0.028 ±0.020	0.371 ±0.029	0.423 ±0.039	0.199 ±0.059	0.049 ±0.029	1.011 ±0.118	0.389 ±0.059	0.757 ±0.150	0.039 ±0.027	0.462 ±0.128	0.125 ±0.094	0.301 ±0.126
GNN-DR	0.043 ±0.031	0.328 ±0.207	0.330 ±0.195	0.480 ±0.282	0.030 ±0.020	0.156 ±0.082	0.220 ±0.133	0.413 ±0.174	0.126 ±0.128	1.003 ±0.764	0.938 ±0.701	1.966 ±1.224	0.099 ±0.064	0.322 ±0.283	0.586 ±0.746	0.493 ±0.410
HINet	0.037 ±0.051	0.313 ±0.062	0.295 ±0.036	0.150 ±0.073	0.017 ±0.012	0.320 ±0.091	0.346 ±0.041	0.167 ±0.046	0.067 ±0.036	0.956 ±0.133	0.425 ±0.017	0.662 ±0.127	0.055 ±0.044	0.250 ±0.116	0.280 ±0.039	0.103 ±0.047
GAIPW	0.067 ±0.038	0.139 ±0.065	0.116 ±0.085	0.204 ±0.090	0.060 ±0.022	0.157 ±0.048	0.039 ±0.024	0.215 ±0.053	0.042 ±0.040	0.155 ±0.064	0.191 ±0.169	0.152 ±0.107	0.064 ±0.047	0.145 ±0.116	0.361 ±0.301	0.149 ±0.117
SiDE (ours)	**0.013** ±0.009	0.049 ±0.028	0.085 ±0.045	**0.064** ±0.034	**0.012** ±0.008	**0.014** ±0.009	**0.032** ±0.033	**0.027** ±0.018	0.042 ±0.031	**0.118** ±0.075	**0.134** ±0.104	**0.128** ±0.083	0.042 ±0.022	**0.060** ±0.052	0.195 ±0.135	**0.093** ±0.097

2021), we keep each observed network and generate treatments and outcomes on it, so that the true effects are known. For the first five networks, the six base covariates are standard normal draws independent of the network. For Reddit, they are the first six principal components of the released community embeddings. All models also receive signed degrees. Table A2 in Appendix D.1 reports network statistics, and Appendix D.2 describes how treatments and outcomes are generated. In the two Bitcoin networks, few egos reach the exposure states compared by the interaction, so the main comparison uses the other four networks.

Baselines. We compare SiDE with eleven baselines from four families. Saturated regression (Vazquez-Bare, 2023) and GPS (Forastiere et al., 2021) adjust for exposure counts through regression and generalized propensity scores. TARNet and CFR (Shalit et al., 2017) are neural treatment-effect estimators that take the exposures as inputs. SAGE-HSIC (Ma and Tresp, 2021), NetEst (Jiang and Sun, 2022), HINITE (Lin et al., 2023), DWR (Zhao et al., 2024), GNN-DR (Leung and Loupos, 2025), and HINet (Caljon et al., 2026) are graph neural estimators for network interference, and GAIPW (Qu et al., 2026) is a generalized AIPW estimator for type-specific exposures. Appendix D.3 describes each baseline.

Metrics. We evaluate the estimates by the bias $J^{-1}\sum_{s=1}^{J}(\hat{\tau}_C^{(s)} - \tau_C^q)$ and the MAE (mean absolute error) $J^{-1}\sum_{s=1}^{J}|\hat{\tau}_C^{(s)} - \tau_C^q|$ over $J = 10$ replications, where $\hat{\tau}_C^{(s)}$ is the estimate of contrast C in replication s and τ_C^q its true value. Lower MAE and smaller absolute bias indicate better estimation accuracy.

5.2 Results on semi-synthetic data

Effect estimation. In Table 1, SiDE has the lowest MAE averaged over the four networks and four effects. It has the lowest MAE for all four effects on Epinions, and individual baselines are more accurate for some effects on the other three networks, including the interaction on Wiki-Elec. Table A3 in Appendix D.4 reports the MAE on the two Bitcoin networks. The bias results in Table A4 also place SiDE first in average absolute bias across all six networks and four effects.

Why tie signs matter. As shown in Figure 3, removing the signs from an otherwise identical estimator attenuates the estimated negative-tie effect on Wiki-Elec, Slashdot, and Epinions and reverses its sign on average on Reddit. This doubly robust estimator combines SiDE's outcome network with a softmax exposure-state model, and its unsigned version, fitted on the same folds, pools positive and negative neighbors in both models. The unsigned version also makes the sign-composition effect identically zero, since the two composition states have the same treated-neighbor total. Figure A2 and Table A5 in Appendix D.4 show the two Bitcoin networks and the baselines fitted without tie signs.

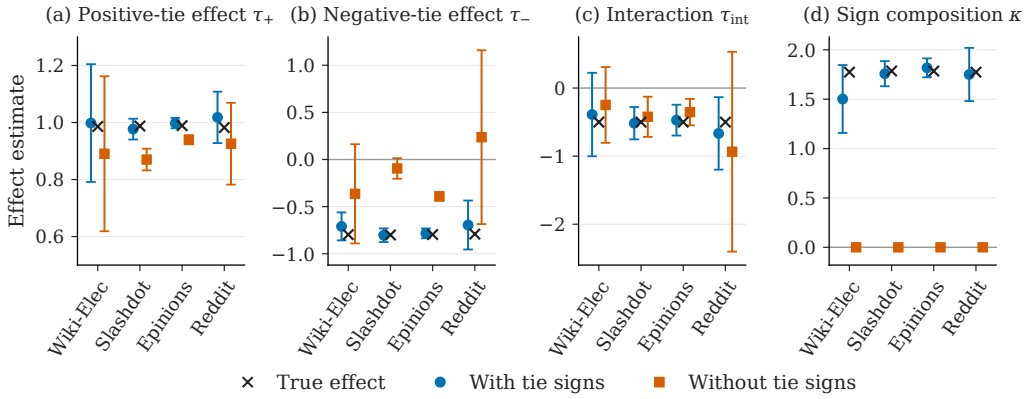

Figure 3: Effects with and without tie signs. Markers show mean estimates, bars show standard deviations, and crosses show true effects.

Table 2: Ablation results on four networks: (a) MAE with standard deviations of the absolute error below and (b) bias. The variant w/o weighting drops the residual correction, w/o outcome drops the outcome model, and w/o PB predicts joint-state probabilities with a softmax model. Bold and underline mark the smallest and second-smallest MAE in (a) and absolute bias in (b).

(a) MAE and standard deviation

	Slashdot				Epinions				Reddit				Wiki-Elec			
Variant	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ
SiDE	**0.013** ±0.009	0.049 ±0.028	0.085 ±0.045	0.064 ±0.034	**0.012** ±0.008	**0.014** ±0.009	**0.032** ±0.033	**0.027** ±0.018	0.042 ±0.031	0.118 ±0.075	0.134 ±0.104	**0.128** ±0.083	0.042 ±0.022	**0.060** ±0.052	0.195 ±0.135	**0.093** ±0.097
w/o weighting	0.015 ±0.004	**0.035** ±0.021	**0.053** ±0.046	**0.050** ±0.023	0.012 ±0.009	0.017 ±0.011	0.060 ±0.038	0.035 ±0.022	**0.040** ±0.027	**0.073** ±0.055	**0.112** ±0.089	0.143 ±0.060	**0.033** ±0.025	0.086 ±0.074	**0.121** ±0.065	0.167 ±0.117
w/o outcome	0.114 ±0.109	0.156 ±0.145	0.218 ±0.165	0.248 ±0.293	0.060 ±0.028	0.147 ±0.102	0.233 ±0.273	0.156 ±0.116	0.238 ±0.108	0.829 ±0.674	0.507 ±0.436	0.991 ±0.557	0.115 ±0.053	0.243 ±0.204	0.369 ±0.400	0.319 ±0.194
w/o PB	0.025 ±0.024	0.138 ±0.103	0.204 ±0.132	0.106 ±0.075	0.014 ±0.009	0.042 ±0.023	0.186 ±0.080	0.091 ±0.066	0.189 ±0.161	0.714 ±0.858	0.842 ±0.890	0.378 ±0.397	0.131 ±0.294	0.164 ±0.120	0.436 ±0.302	0.281 ±0.112

(b) Bias

	Slashdot				Epinions				Reddit				Wiki-Elec			
Variant	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ
SiDE	+0.003	**−0.001**	**−0.001**	**−0.002**	**+0.006**	+0.009	**+0.021**	**+0.003**	**−0.017**	**+0.001**	−0.031	**−0.051**	**+0.015**	**+0.055**	−0.063	−0.050
w/o weighting	+0.004	+0.026	+0.045	−0.013	+0.008	+0.013	+0.058	−0.015	−0.026	+0.021	+0.105	−0.143	−0.033	+0.073	+0.112	−0.167
w/o outcome	**+0.002**	−0.007	−0.026	+0.025	+0.034	−0.040	+0.172	+0.009	+0.159	+0.261	**−0.012**	+0.099	+0.070	−0.121	**−0.021**	**+0.026**
w/o PB	−0.014	−0.088	+0.083	−0.015	+0.013	**+0.006**	+0.077	+0.022	−0.044	+0.163	+0.106	−0.198	+0.078	+0.148	−0.256	−0.281

Sensitivity to hidden width. SiDE keeps the lowest average MAE among the methods in Table 1 when the hidden width of the outcome and assignment models, the number of units in each hidden layer, is 32, 64, or 128. Figure A3 in Appendix D.4 reports the width comparison. On the four main networks, the MAE of each effect changes by less than two Monte Carlo standard errors across these widths. The accuracy of SiDE is therefore stable across this range of model capacity.

Computational cost. With the depth of each multilayer perceptron fixed, each training epoch costs $O(|E|H + nH^2)$ for n units, the tie set $E = \{(i, j) : S_{ij} \neq 0\}$, and hidden width H. Table A6 reports runtimes of 11–94 seconds for one complete SiDE run after preprocessing on one NVIDIA RTX 6000 Ada GPU.

5.3 Ablation Study

In Table 2, dropping the outcome model (w/o outcome) or the Poisson-binomial exposure probabilities (w/o PB) raises the MAE for every network and effect. Dropping the residual correction (w/o weighting) gives a similar MAE but a larger bias, with average absolute bias 0.054 against 0.020 for SiDE. SiDE's bias stays within two Monte Carlo standard errors of zero for every network and effect except the negative-tie effect on Wiki-Elec. The price is a larger standard deviation of the estimates, 0.084 against 0.055. Table A7 reports all six networks.

Table 3: Coverage over 200 replications with 95% Clopper–Pearson (CP) confidence intervals. Underlining marks coverage below 0.919, two binomial standard errors below 0.95.

Network	τ_+		τ_-		τ_{int}		κ	
	Coverage	95% CP interval	Coverage	95% CP interval	Coverage	95% CP interval	Coverage	95% CP interval
Slashdot	0.930	[0.885, 0.961]	0.955	[0.916, 0.979]	0.965	[0.929, 0.986]	0.945	[0.904, 0.972]
Epinions	<u>0.905</u>	[0.856, 0.942]	0.930	[0.885, 0.961]	0.965	[0.929, 0.986]	0.965	[0.929, 0.986]
Reddit	<u>0.850</u>	[0.793, 0.896]	0.935	[0.891, 0.965]	0.945	[0.904, 0.972]	0.985	[0.957, 0.997]
Wiki-Elec	0.960	[0.923, 0.983]	0.920	[0.873, 0.954]	0.970	[0.936, 0.989]	0.940	[0.898, 0.969]

5.4 Coverage of approximate intervals

Table 3 reports coverage of at least 0.919, two binomial standard errors below 0.95, for every network and effect except the positive-tie effect on Epinions and Reddit. We generate 200 datasets on each of the four main networks, with independent treatment assignment and disturbances shared through overlapping neighborhoods, and coverage is the fraction of datasets whose nominal 95% interval $\hat{\tau}_C \pm 1.96\sqrt{\hat{\Omega}_C}$, with $\hat{\Omega}_C$ aggregated over partitions as in Algorithm 1, contains the known effect. On Reddit, the positive-tie coverage is 0.850, and on the two Bitcoin networks the positive-tie intervals also undercover, with coverage 0.880 and 0.840. Tables A8 and A9 in Appendix D.4 give details for the four main networks and the two Bitcoin networks.

5.5 Case study

In the 28 program schools of the anti-conflict experiment of Paluck et al. (2016; 2020), a student is treated if randomly selected for an invitation to the program, whether or not the student participated. Pre-treatment spend-time and conflict nominations among the 11,117 students of these schools define positive and negative ties. The outcome indicates whether a student wore an orange wristband during the week before the end-of-year survey, and program participants handed out these wristbands to recognize friendly or conflict-mitigating behavior. We estimate the four effects with SiDE at own invitation status zero, using known randomization probabilities. Each effect uses the students for whom every compared exposure state can occur, 6,375 for the spend-time effect and 941 for the interaction and sign-composition effects. Appendix E details the data, and Table A10 there reports all estimates with sensitivity analyses.

The estimated effect of one invited peer linked by a spend-time tie is 3.3 percentage points, about a fifth of the 15.4% mean, with a linked-pair 95% interval of [0.5, 6.1]. After a Bonferroni adjustment across the four contrasts, this interval becomes $[-0.3, 6.9]$, and we treat the analysis as exploratory, as Appendix E explains. The conflict-tie estimate is -0.9 points with interval $[-7.2, 5.5]$, too wide to determine its direction. For two secondary outcomes, in the program and in the control schools, no estimate has an interval that excludes zero, as Appendix E reports. A sign-blind comparison of untreated students with one invited peer and with none, weighted by the inverse of the known randomization probabilities, gives 3.2 points per invited peer. This estimate is close to the spend-time estimate, consistent with 85% of the invited peers in this comparison being linked by spend-time ties. The signed estimates associate the positive unsigned effect with spend-time ties, which makes these peers the first to target in a confirmatory trial.

6 Conclusion

We defined four contrasts for peer effects in signed networks and identified each from observational data as a contrast of outcome regressions over its reachable set. Under sign-blind assignment, we showed that an unsigned increment is a weighted average of positive- and negative-tie increments, so opposing influences can cancel. We proposed SiDE, which estimates the four effects with a doubly robust score that combines sign-specific outcome models with exposure probabilities induced by individual treatment assignment, and which constructs approximate intervals that account for overlapping neighborhoods. The sign-composition effect tells a program designer whether to treat a trusted or an avoided neighbor when the number of treated neighbors is fixed. We evaluated SiDE on semi-synthetic data built on six real signed networks, and an exploratory reanalysis of a school field experiment associates a positive peer effect on wristband wearing with spend-time ties. Extending the four effects to networks that change in response to treatment is a direction for future work.

AI USE STATEMENT

We used generative AI (artificial intelligence) tools to edit the manuscript for readability and to search the literature.

ETHICS STATEMENT

This work collects no new data from human subjects. The case study analyzes de-identified data from a published field experiment in schools. A public data archive distributes these data to registered researchers. We use them under the archive's terms of use and do not redistribute them with our code.

REPRODUCIBILITY STATEMENT

The identification conditions are stated in Section 3.1, and the assignment model and double-robustness conditions in Section 4.1. Complete proofs of the identification, double-robustness, and mixing results are given in Appendix B. The anonymous repository at `https://anonymous.4open.science/r/signed-peer-effects/README.md` contains the implementation of SiDE and of the data-generating process, together with five of the six signed networks. Implementation and experimental details are given in Appendices C, D, and E.

REFERENCES

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

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.

Elisa Bellotti, Federico Bianchi, Francesco Renzini, Carinthia Balabet Nengnong, Anna Maria van Eijk, Anne Kessler, Rajiv Sarkar, Jane M. Carlton, and Sandra Albert. Negative ties offset complex contagion effect on malaria preventive measure adoption in a hard-to-reach population. SocArXiv preprint, 2026.

Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

Daan Caljon, Jente Van Belle, and Wouter Verbeke. Estimating treatment effects in networks under unknown exposure mappings. arXiv:2510.21457v2, 2026.

Laura M. Callejas and Hana Shepherd. Conflict as a social status mobility mechanism in schools: A network approach. *Social Psychology Quarterly*, 83(4):319–341, 2020.

A. Colin Cameron and Douglas L. Miller. A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372, 2015.

Dorwin Cartwright and Frank Harary. Structural balance: A generalization of Heider's theory. *Psychological Review*, 63(5):277–293, 1956.

Matias D. Cattaneo. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, 155(2):138–154, 2010.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Tyler Derr, Yao Ma, and Jiliang Tang. Signed graph convolutional networks. In *Proceedings of the 18th IEEE International Conference on Data Mining*, pages 929–934, 2018.

Olive Jean Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64, 1961.

Dean Eckles, Brian Karrer, and Johan Ugander. Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, 5(1):20150021, 2017.

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

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.

Amir Ghasemian and Nicholas A. Christakis. The structure and function of antagonistic ties in village social networks. *Proceedings of the National Academy of Sciences*, 121(26):e2401257121, 2024.

R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th International Conference on World Wide Web*, pages 403–412, 2004.

Yili Hong. On computing the distribution function for the Poisson binomial distribution. *Computational Statistics & Data Analysis*, 59:41–51, 2013.

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

Song Jiang and Yizhou Sun. Estimating causal effects on networked observational data via representation learning. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, pages 852–861, 2022.

Joseph D. Y. Kang and Joseph L. Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4):523–539, 2007.

David A. Kim, Alison R. Hwong, Derek Stafford, D. Alex Hughes, A. James O'Malley, James H. Fowler, and Nicholas A. Christakis. Social network targeting to maximise population behaviour change: A cluster randomised controlled trial. *The Lancet*, 386(9989):145–153, 2015.

Michael D. König, Dominic Rohner, Mathias Thoenig, and Fabrizio Zilibotti. Networks in conflict: Theory and evidence from the Great War of Africa. *Econometrica*, 85(4):1093–1132, 2017.

Srijan Kumar, Francesca Spezzano, V. S. Subrahmanian, and Christos Faloutsos. Edge weight prediction in weighted signed networks. In *Proceedings of the 16th IEEE International Conference on Data Mining*, pages 221–230, 2016.

Srijan Kumar, William L. Hamilton, Jure Leskovec, and Dan Jurafsky. Community interaction and conflict on the web. In *Proceedings of the 2018 World Wide Web Conference*, pages 933–943, 2018a.

Srijan Kumar, Bryan Hooi, Disha Makhija, Mohit Kumar, Christos Faloutsos, and V. S. Subrahmanian. REV2: Fraudulent user prediction in rating platforms. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, pages 333–341, 2018b.

Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1361–1370, 2010a.

Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*, pages 641–650, 2010b.

Michael P. Leung and Pantelis Loupos. Graph neural networks for causal inference under network confounding. arXiv:2211.07823v5, 2025.

Yanhua Li, Wei Chen, Yajun Wang, and Zhi-Li Zhang. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, pages 657–666, 2013.

Xiaofeng Lin, Guoxi Zhang, Xiaotian Lu, Han Bao, Koh Takeuchi, and Hisashi Kashima. Estimating treatment effects under heterogeneous interference. In *Machine Learning and Knowledge Discovery in Databases: Research Track, ECML PKDD 2023, Proceedings, Part I*, volume 14169 of *Lecture Notes in Computer Science*, pages 576–592. Springer, 2023.

Yunpu Ma and Volker Tresp. Causal inference under networked interference and intervention policy enhancement. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, pages 3700–3708, 2021.

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.

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.

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, ICPSR 37070, version 2, 2020.

Zhaonan Qu, Ruoxuan Xiong, Jizhou Liu, and Guido Imbens. Semiparametric estimation of treatment effects in observational studies with heterogeneous partial interference. *Journal of Business & Economic Statistics*, pages 1–32, 2026.

Karthik Rajkumar, Guillaume Saint-Jacques, Iavor Bojinov, Erik Brynjolfsson, and Sinan Aral. A causal test of the strength of weak ties. *Science*, 377(6612):1304–1310, 2022.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994.

Uri Shalit, Fredrik D. Johansson, and David Sontag. Estimating individual treatment effect: Generalization bounds and algorithms. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3076–3085, 2017.

Károly Takács, Andreas Flache, and Michael Mäs. Discrepancy and disliking do not induce negative opinion shifts. *PLOS ONE*, 11(6):e0157948, 2016.

Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. A survey of signed network mining in social media. *ACM Computing Surveys*, 49(3):42, 2016.

Gonzalo Vazquez-Bare. Identification and estimation of spillover effects in randomized experiments. *Journal of Econometrics*, 237(1):105237, 2023.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems 30*, 2017.

Ziyu Zhao, Yuqi Bai, Ruoxuan Xiong, Qingyu Cao, Chao Ma, Ning Jiang, Fei Wu, and Kun Kuang. Learning individual treatment effects under heterogeneous interference in networks. *ACM Transactions on Knowledge Discovery from Data*, 18(8):199, 2024.

APPENDIX: PROOFS AND SUPPLEMENTARY RESULTS

CONTENTS

A Notation and the running example . . . 13
A.1 The four effects in the running example . . . 14
B Proofs of the main results . . . 14
C Implementation details . . . 16
D Experimental details . . . 17
D.1 Signed networks . . . 17
D.2 Treatment and outcome generation . . . 18
D.3 Baselines . . . 18
D.4 Additional results . . . 18
E Case study details . . . 21

A NOTATION AND THE RUNNING EXAMPLE

Table A1 lists the main notation. Appendix A.1 restates the four definitions and illustrates them using the households in Figure 1.

Table A1: Main notation.

Symbol	Meaning
$V = \{1, \dots, n\}$	Set of units (egos and their peers).
$S = (S_{ij}),\ S_{ij} \in \{-1, 0, +1\}$	Signed adjacency matrix and pre-treatment sign of the tie from ego i to peer j.
$\mathcal{N}_i^+, \mathcal{N}_i^-, \mathcal{N}_i$	Positive-tie and negative-tie neighborhoods of unit i, and their union.
d_i^+, d_i^-	Numbers of positive-tie and negative-tie neighbors of unit i.
$T_i, \mathbf{T}$	Treatment of unit i and the full treatment vector.
Y_i	Observed post-treatment outcome of unit i.
$X_i, X_{\mathcal{N}_i}$	Ego pre-treatment characteristics and network features (including signed degrees), and neighbor covariates.
K_i^+, K_i^-	Numbers of treated positive-tie and negative-tie peers.
k^+, k^-	Specified nonnegative integer values of the exposure counts.
K_i^u	Unsigned treated-neighbor total $K_i^+ + K_i^-$.
$Y_i(t, k^+, k^-)$	Potential outcome at state (t, k^+, k^-).
$\mathcal{V}(k^+, k^-)$	Reachable set $\{i : d_i^+ \geq k^+,\ d_i^- \geq k^-\}$.
$\mathcal{U}$	A target population defined by signed degrees.
$\tau_+, \tau_-, \tau_{\text{int}}, \kappa$	Positive-tie peer effect, negative-tie peer effect, interaction between tie types, and sign-composition effect.
$m(t, k^+, k^-, q)$	Outcome regression $\mathbb{E}[Y_i \mid Z_i = (t, k^+, k^-), Q_i = q]$.
$z,\ Z_i$	A joint-state value and the observed state (T_i, K_i^+, K_i^-).
$\mathbf{X}$	Covariates $(X_1, \dots, X_n)$ of all units.
$p_j,\ e_i(z)$	Unit-level assignment probability $\Pr(T_j = 1 \mid S, \mathbf{X})$, and exposure propensity $\Pr(Z_i = z \mid S, \mathbf{X})$ of ego i.
$C,\ \tau_C,\ \varphi_C$	Effect indexed as a contrast, its target value, and its doubly robust score.
$\mathcal{V}_C,\ n_C$	Common reachable set of contrast C and its size.
$Q_i,\ q_i$	Adjustment covariates $(X_i, X_{\mathcal{N}_i})$ of ego i, and their observed value.
$\check{\tau}_C$	Unnormalized estimate from one fold partition.
$\tilde{\tau}_C,\ \tilde{\Omega}_C$	Normalized estimate and linked-pair variance from one fold partition.
$\hat{\tau}_C,\ \hat{\Omega}_C$	Estimate and variance aggregated over P partitions by the median.
$i \sim j$	Linked egos: $\bar{\mathcal{N}}_i \cap \bar{\mathcal{N}}_j \neq \varnothing$, where $\bar{\mathcal{N}}_i = \{i\} \cup \mathcal{N}_i$ is the closed neighborhood of i.

A.1 THE FOUR EFFECTS IN THE RUNNING EXAMPLE

In the village health program of Figure 1, households are egos, trusted and avoided neighbors form $\mathcal{N}_i^+$ and $\mathcal{N}_i^-$, receiving the product through the program is the treatment, and subsequent use is the outcome. A and B each have four trusted and two avoided neighbors, and their depicted exposures are $(3,1)$ and $(4,0)$. We restate Definitions 1–4 below. The examples set own treatment to $t=0$. Each contrast averages alternative outcomes for the same households that can reach every compared exposure.

Definition 1 (Positive-tie peer effect)**.** For own treatment t, negative-tie exposure k^-, and positive-tie levels $k_0^+ < k_1^+$,

$$\tau_+\big(k_0^+ \to k_1^+;\, t, k^-\big) = \mathbb{E}\big[\, Y_i(t, k_1^+, k^-) - Y_i(t, k_0^+, k^-) \;\big|\; i \in \mathcal{V}(k_1^+, k^-)\big]\,.$$

Example. The solid horizontal arrow depicts $\tau_+(0\to 1;\, 0,0)$: treating one trusted neighbor compared with treating none, with all avoided neighbors untreated. It averages over $\mathcal{V}(1,0)$, households with at least one trusted neighbor, including A and B. The same one-neighbor step taken at A's exposure, $\tau_+(3\to 4;\, 0,1)$, moves from $(3,1)$ to $(4,1)$ and averages over $\mathcal{V}(4,1)$.

Definition 2 (Negative-tie peer effect)**.** For own treatment t, positive-tie exposure k^+, and negative-tie levels $k_0^- < k_1^-$,

$$\tau_-\big(k_0^- \to k_1^-;\, t, k^+\big) = \mathbb{E}\big[\, Y_i(t, k^+, k_1^-) - Y_i(t, k^+, k_0^-) \;\big|\; i \in \mathcal{V}(k^+, k_1^-)\big]\,.$$

Example. The vertical arrow depicts $\tau_-(0\to 1;\, 0,0)$: treating one avoided neighbor compared with treating none, with all trusted neighbors untreated. It averages over $\mathcal{V}(0,1)$, households with at least one avoided neighbor. The negative tie does not fix the sign of this effect, since treating an avoided neighbor can raise, leave unchanged, or lower product use.

Definition 3 (Interaction between tie types)**.** For own treatment t, positive-tie levels $k_0^+ < k_1^+$, and negative-tie levels $k_0^- < k_1^-$,

$$\begin{aligned}\tau_{\text{int}}\big(k_0^+ \to k_1^+,\, k_0^- \to k_1^-;\, t\big) = \mathbb{E}\Big[\, &\big\{Y_i(t, k_1^+, k_1^-) - Y_i(t, k_0^+, k_1^-)\big\} \\ &- \big\{Y_i(t, k_1^+, k_0^-) - Y_i(t, k_0^+, k_0^-)\big\} \;\Big|\; i \in \mathcal{V}(k_1^+, k_1^-)\Big].\end{aligned}$$

Example. The interaction $\tau_{\text{int}}(0\to 1, 0\to 1;\, 0)$ subtracts the lower-row increment $Y_i(0,1,0) - Y_i(0,0,0)$ from the upper-row increment $Y_i(0,1,1) - Y_i(0,0,1)$. All four potential outcomes are averaged over $\mathcal{V}(1,1)$, households with at least one neighbor of each type. A negative value means that treating an avoided neighbor reduces the increase in product use from treating a trusted neighbor.

Definition 4 (Sign-composition effect)**.** For own treatment t and levels $k^+ \geq 0$ and $k^- \geq 1$,

$$\kappa\big(k^+, k^-;\, t\big) = \mathbb{E}\big[\, Y_i(t, k^+ + 1, k^- - 1) - Y_i(t, k^+, k^-) \;\big|\; i \in \mathcal{V}(k^+ + 1, k^-)\big]\,.$$

Example. The contrast $\kappa(3,1;0)$ compares each household's product use at B's exposure $(4,0)$ with its product use at A's exposure $(3,1)$, averaging over $\mathcal{V}(4,1)$. The two assignments differ by reallocating treatment from an avoided to a trusted neighbor while keeping the total at four. An unsigned analysis assigns the same exposure to both endpoints. The experiments report $\kappa(0,1;0)$, which compares $(1,0)$ with $(0,1)$ over $\mathcal{V}(1,1)$.

B PROOFS OF THE MAIN RESULTS

We prove Propositions 1–3 and Corollary 1, which expresses the four effects in terms of the outcome regression. We use $Q_i = (X_i, X_{\mathcal{N}_i})$ and the expectation convention of Section 3.1.

Proof of Proposition 1. Write q for a value of Q_i and $G_{\mathcal{U}}$ for the distribution of Q_i among the egos of $\mathcal{U}$. Membership in $\mathcal{U}$ is a function of the signed degrees, which are part of X_i, so Q_i determines membership in $\mathcal{U}$, and conditioning on Q_i and on membership is the same as conditioning on Q_i. By the tower property,

$$\mathbb{E}\big[Y_i(t, k^+, k^-) \mid i \in \mathcal{U}\big] = \int \mathbb{E}\big[Y_i(t, k^+, k^-) \mid Q_i = q\big]\, \mathrm{d}G_{\mathcal{U}}(q). \tag{16}$$

For $G_{\mathcal{U}}$-almost every q the state (t, k^+, k^-) has positive probability by Assumption 4, so the conditional expectation in the integrand stays well defined when the state is added to the conditioning set. Assumption 3 allows exactly that addition, and Assumption 2 then replaces the potential outcome by the observed one:

$$\begin{aligned}\mathbb{E}\big[Y_i(t,k^+,k^-) \mid Q_i = q\big] &= \mathbb{E}\big[Y_i(t,k^+,k^-) \mid T_i = t,\, K_i^+ = k^+,\, K_i^- = k^-,\, Q_i = q\big] \\ &= \mathbb{E}\big[Y_i \mid T_i = t,\, K_i^+ = k^+,\, K_i^- = k^-,\, Q_i = q\big]\,.\end{aligned} \tag{17}$$

Substituting Eq. (17) into Eq. (16) gives Eq. (6). □

Corollary 1 (Identified form of the four effects)**.** *Under Assumptions 1 to 4, for the exposure levels and nonempty reachable sets of Definitions 1–4,*

$$\tau_+\big(k_0^+ \to k_1^+;\, t, k^-\big) = \mathbb{E}\big[\, m(t,k_1^+,k^-,Q_i) - m(t,k_0^+,k^-,Q_i) \;\big|\; i \in \mathcal{V}(k_1^+,k^-)\big], \tag{18}$$

$$\tau_-\big(k_0^- \to k_1^-;\, t, k^+\big) = \mathbb{E}\big[\, m(t,k^+,k_1^-,Q_i) - m(t,k^+,k_0^-,Q_i) \;\big|\; i \in \mathcal{V}(k^+,k_1^-)\big], \tag{19}$$

$$\begin{aligned}\tau_{\mathrm{int}}\big(k_0^+ \to k_1^+,\, k_0^- \to k_1^-;\, t\big) = \mathbb{E}\big[\, & m(t,k_1^+,k_1^-,Q_i) - m(t,k_0^+,k_1^-,Q_i) \\ & - m(t,k_1^+,k_0^-,Q_i) + m(t,k_0^+,k_0^-,Q_i) \;\big|\; i \in \mathcal{V}(k_1^+,k_1^-)\big],\end{aligned} \tag{20}$$

$$\begin{aligned}\kappa\big(k^+,k^-;\, t\big) = \mathbb{E}\big[\, & m(t,k^+{+}1,k^-{-}1,Q_i) - m(t,k^+,k^-,Q_i) \\ & \;\big|\; i \in \mathcal{V}(k^+{+}1,k^-)\big].\end{aligned} \tag{21}$$

Proof of Corollary 1. The set $\mathcal{V}(k_1^+,k^-)$ is defined by the signed degrees and is contained in $\mathcal{V}(k_0^+,k^-)$, because an ego that can reach k_1^+ treated positive-tie neighbors can reach $k_0^+ < k_1^+$ of them. Proposition 1 therefore applies with $\mathcal{U} = \mathcal{V}(k_1^+,k^-)$ to both potential outcomes in Eq. (1), and linearity of expectation gives Eq. (18). The same argument with $\mathcal{U} = \mathcal{V}(k^+,k_1^-)$ gives Eq. (19). The set $\mathcal{V}(k_1^+,k_1^-)$ is contained in the reachable set of each of the four levels in Eq. (3), so taking $\mathcal{U} = \mathcal{V}(k_1^+,k_1^-)$ gives Eq. (20). Finally, $\mathcal{V}(k^+{+}1,k^-)$ is defined by the signed degrees and is contained in $\mathcal{V}(k^+{+}1,k^-{-}1)$ and in $\mathcal{V}(k^+,k^-)$, so Proposition 1 applies with $\mathcal{U} = \mathcal{V}(k^+{+}1,k^-)$ to both potential outcomes in Eq. (4), and linearity of expectation gives Eq. (21). □

Proof of Proposition 2. The events $\{K_i^+ = k^+,\, K_i^- = k - k^+\}$ for $k^+ = 0,\dots,k$ partition $\{K_i^u = k\}$, so by the law of total expectation

$$\begin{aligned}\mathbb{E}\big[Y_i \mid T_i = t,\, K_i^u = k,\, Q_i\big] = \sum_{k^+} & \pi(k^+;t,k,Q_i) \\ & \times \mathbb{E}\big[Y_i \mid T_i = t,\, K_i^+ = k^+,\, K_i^- = k - k^+,\, Q_i\big],\end{aligned} \tag{22}$$

with $\pi(k^+;t,k,Q_i) = \Pr(K_i^+ = k^+ \mid T_i = t, K_i^u = k, Q_i)$. Compositions the ego cannot reach have $\pi = 0$ and drop out. For every remaining k^+, consistency and exchangeability turn the conditional expectation on the right into $\mathbb{E}[Y_i(t,k^+,k-k^+) \mid Q_i]$, as in the proof of Proposition 1. Sign-blind assignment makes $\pi(k^+;t,k,Q_i)$ the hypergeometric probability $\binom{a}{k^+}\binom{b}{k-k^+}/\binom{N}{k}$ of drawing k^+ positive-tie neighbors among k treated ones, which depends on neither Q_i nor the ego. Averaging Eq. (22) over Q_i among the egos with degrees (a,b) therefore gives $\mu^u(t,k) = \mathbb{E}[\mu(t,R,k-R)]$, where R, the number of positive-tie neighbors in a uniformly drawn set $\mathcal{D}$ of k treated neighbors, is hypergeometric with population N, a successes, and k draws.

A uniform set of size $k+1$ can be drawn by taking a uniform set $\mathcal{D}$ of size k and adding one peer uniformly from the $N-k$ that remain, so the two treated sets differ by a single peer. Given $\mathcal{D}$, the added peer is a positive-tie neighbor with probability $(a-R)/(N-k)$ and a negative-tie neighbor with probability $(b-k+R)/(N-k)$. With $\Delta^+(k^+) = \mu(t,k^+{+}1,k-k^+) - \mu(t,k^+,k-k^+)$ and $\Delta^-(k^+) = \mu(t,k^+,k-k^+{+}1) - \mu(t,k^+,k-k^+)$,

$$\mu^u(t,k+1) - \mu^u(t,k) = \mathbb{E}\left[\frac{a-R}{N-k}\,\Delta^+(R)\right] + \mathbb{E}\left[\frac{b-k+R}{N-k}\,\Delta^-(R)\right]. \tag{23}$$

The probability that the added peer is a positive-tie neighbor falls as more positive-tie neighbors are already treated, and the identities $(a-k^+)\binom{a}{k^+} = a\binom{a-1}{k^+}$ and $(N-k)\binom{N}{k} = N\binom{N-1}{k}$ absorb that dependence exactly. They turn the first term into $(a/N)\,\mathbb{E}[\Delta^+(R_k^+)]$, and the same computation with b in place of a turns the second into $(b/N)\,\mathbb{E}[\Delta^-(R_k^-)]$, which is Eq. (8). □

Proof of Proposition 3. Fix an ego $i \in \mathcal{V}_C$ and a state $z \in \mathcal{Z}_C$, and condition on $(S, \mathbf{X})$. On the event $Z_i = z$ the observed outcome is $Y_i(z)$ by Assumptions 1 and 2. The conditional mean condition of Proposition 3 gives $\mathbb{E}[Y_i(z) \mid Z_i = z, S, \mathbf{X}] = m(z, q_i)$, so

$$\mathbb{E}\left[\frac{\mathbf{1}\{Z_i = z\}}{\bar{e}_i(z)}\big(Y_i - \bar{m}(z, q_i)\big) \;\middle|\; S, \mathbf{X}\right] = \frac{e_i(z)}{\bar{e}_i(z)}\big\{m(z, q_i) - \bar{m}(z, q_i)\big\}.$$

The mean of the bracket in Eq. (11) is therefore $\bar{m}(z, q_i) + \{e_i(z)/\bar{e}_i(z)\}\{m(z, q_i) - \bar{m}(z, q_i)\}$, which equals $m(z, q_i)$ when $\bar{m} = m$ and also when $\bar{e}_i = e_i$. Summing over the states with coefficients c_z and averaging over the egos of $\mathcal{V}_C$ gives τ_C^q. The condition $\mathbb{E}[Y_i(z) \mid S, \mathbf{X}] = m(z, q_i)$ of Proposition 3 gives $\sum_z c_z\, m(z, q_i) = \mathbb{E}[\sum_z c_z\, Y_i(z) \mid S, \mathbf{X}]$ by linearity, so the mean of τ_C^q over the covariates is the average of $\mathbb{E}[\sum_z c_z\, Y_i(z)]$ over the egos of $\mathcal{V}_C$, which is τ_C under the convention of Section 3.1. □

C Implementation Details

Figure A1 summarizes SiDE, and Algorithm 1 gives the estimation procedure.

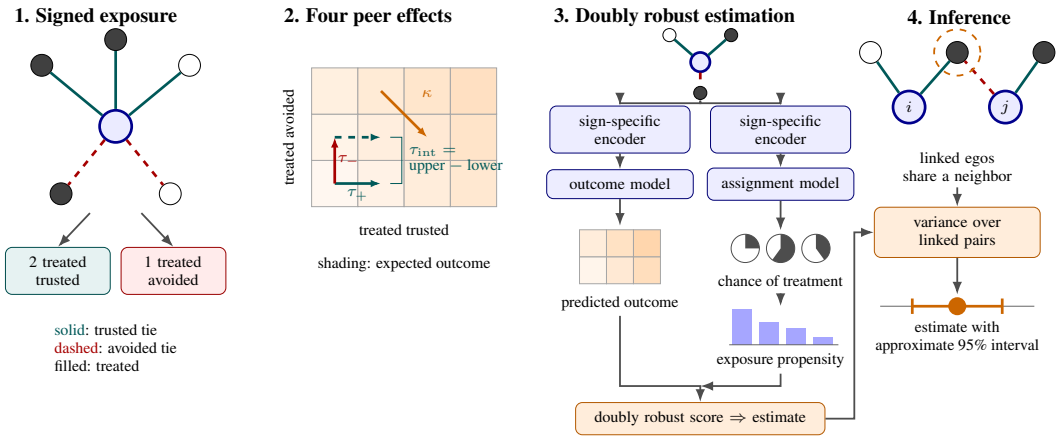

Figure A1: SiDE estimation. (1) Construct the signed exposure state. (2) Define the four outcome contrasts. (3) Fit separate outcome and assignment encoders, then combine outcome predictions with corrections based on exposure probabilities. (4) Estimate uncertainty using residual products over overlapping neighborhoods.

Outcome head. The outcome head embeds each exposure level as a free vector, $\psi^+(k^+)$ and $\psi^-(k^-)$, with one vector shared by all levels above five, the cap of the experiments, and sets

$$m_\theta(z, q_i) = \ell\Big(u_t(h_i) + v_t^+(h_i)^\top \psi^+(k^+) + v_t^-(h_i)^\top \psi^-(k^-) + \psi^+(k^+)^\top \Gamma_t(h_i)\, \psi^-(k^-)\Big), \quad (24)$$

where ℓ is the identity for a continuous outcome and the logistic function for a binary one, u_t, v_t^+, v_t^-, and Γ_t are heads on h_i with one set of parameters for each treatment value $t \in \{0, 1\}$, and θ collects all parameters. The fitted m_θ is the $\hat{m}$ of Section 4.2. The last term lets the effect of one exposure vary with the level of the other.

Assignment head and exposure propensity. The assignment head is a logistic model $\hat{p}_j$ of T_j on the representation $\tilde{h}_j$ of the assignment encoder. The exposure propensity of ego i at state $z = (t, k^+, k^-)$ is Eq. (10) evaluated at $\hat{p}$. Each Poisson-binomial factor starts from $\Pi(k) = \mathbf{1}\{k = 0\}$ and adds the neighbors j of one side one at a time, replacing $\Pi(k)$ by $(1 - \hat{p}_j)\Pi(k) + \hat{p}_j \Pi(k - 1)$ for every k, with $\Pi(-1) = 0$. The two factors cost $O\{(d_i^+)^2 + (d_i^-)^2\}$ operations for ego i.

Training settings. SiDE's outcome and assignment models use hidden width 64, Adam with learning rate 0.01, weight decay 10^{-4}, and at most 300 epochs.

Variance implementation. For the normalized estimator, the fitted residual in Eq. (15) is $\hat{r}_i = \sum_{z \in \mathcal{Z}_C} c_z \mathbf{1}\{Z_i = z\}\{Y_i - \hat{m}(z, q_i)\}/\{\hat{e}_i(z)\bar{w}_z\}$, where $w_{iz} = \mathbf{1}\{Z_i = z\}/\hat{e}_i(z)$ and $\bar{w}_z = n_C^{-1} \sum_{i \in \mathcal{V}_C} w_{iz}$. If the resulting variance is negative, the indicator of a linked pair is replaced

by the positive semidefinite overlap weight of Leung and Loupos (2025), computed on one-hop neighborhoods, $|\bar{\mathcal{N}}_i \cap \bar{\mathcal{N}}_j|/(|\bar{\mathcal{N}}_i|\,|\bar{\mathcal{N}}_j|)^{1/2}$. The intervals are approximate and their coverage is evaluated empirically.

Algorithm 1 SiDE: doubly robust estimation of a set of contrasts on a signed network

Input: Signed network S, data $\{(Y_i, T_i, X_i)\}_{i\in V}$, a set of contrasts $\mathcal{C} = \{C_1, \ldots, C_M\}$ with states $\mathcal{Z}_C$ and coefficients c_z, number of folds F, number of partitions P

Output: Estimates $\hat{\tau}_C$ and variance estimates $\hat{\Omega}_C$ for every $C \in \mathcal{C}$

Compute K_i^+ and K_i^- for every ego, the reachable sets $\mathcal{V}_C$, their union $\mathcal{V}_\cup = \bigcup_{C\in\mathcal{C}} \mathcal{V}_C$, and the links $i \sim j$
for $\nu = 1, \ldots, P$ **do**
  Randomly partition $\mathcal{V}_\cup$ into folds $I_1, \ldots, I_F$
  for $f = 1, \ldots, F$ **do**
    Train $\hat{m}$ and $\hat{p}$ on the units of $V \setminus I_f$
    Clip fitted treatment probabilities to $[0.01, 0.99]$ and compute $\hat{e}_i$ for $i \in I_f$ by Eq. (10)
  end for
  for each $C \in \mathcal{C}$ **do**
    Mark the contrast undefined if any compared state is unobserved
    Otherwise, normalize each state's inverse-probability weights to have mean one over $\mathcal{V}_C$
    Compute Eq. (11) and residuals with these weights and each ego's held-out models
    $\tilde{\tau}_C^{(\nu)} \leftarrow$ the average of the normalized scores over $\mathcal{V}_C$
    $\tilde{\Omega}_C^{(\nu)} \leftarrow n_C^{-2} \sum_{i,j\in\mathcal{V}_C:\, i\sim j} \hat{r}_i \hat{r}_j$, using the normalized residuals and variance fallback defined above
  end for
end for
for each $C \in \mathcal{C}$ **do**
  $\hat{\tau}_C \leftarrow \mathrm{median}_\nu\, \tilde{\tau}_C^{(\nu)}$ (Chernozhukov et al., 2018)
  $\hat{\Omega}_C \leftarrow \mathrm{median}_\nu \left\{\tilde{\Omega}_C^{(\nu)} + \left(\tilde{\tau}_C^{(\nu)} - \hat{\tau}_C\right)^2\right\}$
end for
return $\hat{\tau}_C$ and $\hat{\Omega}_C$ for every $C \in \mathcal{C}$

D Experimental details

D.1 Signed networks

The six networks are released by the Stanford Network Analysis Project. **Bitcoin-Alpha** and **Bitcoin-OTC** record trader ratings from -10 to $+10$, with the rating sign defining the tie sign (Kumar et al., 2016; 2018b). **Wiki-Elec** records votes for or against Wikipedia administrator candidates, with neutral votes dropped (Leskovec et al., 2010a;b). **Slashdot** records friend/foe marks in February 2009 (Leskovec et al., 2010a). **Epinions** records trust/distrust nominations on a consumer review site (Leskovec et al., 2010a). **Reddit** records hyperlinks between communities, labeled negative or neutral-to-positive (Kumar et al., 2018a). We pool title and body hyperlinks and retain communities with released embeddings. Their first six principal components, together with signed degrees, form the Reddit covariates.

Table A2: Network statistics. Neg. is the share of negative ties, and Scarcest is the smallest expected number of egos among the four states of the interaction contrast. Reachable-set sizes and scarcest-state counts use binary exposure contrasts with at most five ties per sign.

Network	Tie	Units	Ties	Neg.	$\lvert\mathcal{V}(1,0)\rvert$	$\lvert\mathcal{V}(0,1)\rvert$	$\lvert\mathcal{V}(1,1)\rvert$	Scarcest
Bitcoin-Alpha	trust / distrust	3,783	23,690	5.4%	2,444	93	81	6
Bitcoin-OTC	trust / distrust	5,881	34,876	9.2%	3,628	235	202	15
Wiki-Elec	support / oppose	7,115	103,091	21.0%	3,523	1,437	823	68
Slashdot	friend / foe	82,140	545,304	22.4%	28,094	5,986	4,291	336
Epinions	trust / distrust	131,580	835,393	14.5%	71,281	11,897	4,858	303
Reddit	negative / other	30,744	271,667	7.0%	19,242	1,257	1,028	109

D.2 Treatment and outcome generation

Target population. The experiments report the first-step contrasts $\tau_+(0\to1;0,0)$, $\tau_-(0\to1;0,0)$, $\tau_{\text{int}}(0\to1,0\to1;0)$, and $\kappa(0,1;0)$ for untreated egos, written without arguments in tables and figures. Each effect averages over the egos of its reachable set with at most five ties of each sign, a degree-defined set covered by Proposition 1 that keeps the compared states away from zero probability. True values and all methods use the same egos, and the network and covariates stay fixed across replications.

Covariates and treatment. Each X_i contains six base covariates and the standardized logarithms of the two signed degrees. The first four coordinates are transformed to $\exp(X_{i1}/2)$, $X_{i2}/\{1+\exp(X_{i1})\}$, $(X_{i1}X_{i3}/25+0.6)^3$, and $(X_{i2}+X_{i4}+20)^2$, adapted from Kang and Schafer (2007), and the six base coordinates are then standardized. Clipping all eight coordinates to $[-5,5]$ gives the vector W_i that generates treatments and outcomes. The estimators receive X_i only.

Let $\bar W_i^+$ and $\bar W_i^-$ be their averages over positive and negative neighbors. The index $\zeta_i = W_i^\top\gamma + \bar W_i^{+\top}\gamma^+ + \bar W_i^{-\top}\gamma^-$ is scaled to standard deviation 0.75. Treatments are independent Bernoulli draws with $\operatorname{logit} p_i = \gamma_0 + \zeta_i$, where the intercept γ_0 is chosen so that the probabilities average 0.25, and each p_i is then clipped to $[0.05, 0.95]$.

Outcome. With $\omega(k) = \log_2(1+k)$, we generate

$$Y_i = \alpha_0(i) + \beta_0(W_i)T_i + \beta_+(W_i)\omega(K_i^+) - \beta_-(W_i)\omega(K_i^-) + \lambda_\times\,\omega(K_i^+)\omega(K_i^-) + \varepsilon_i. \quad (25)$$

Here $\varepsilon_i = 0.7\{\xi_i + 0.7\max(d_i,1)^{-1/2}\sum_{j\in\mathcal{N}_i}\xi_j\}$ with $d_i = d_i^+ + d_i^-$ and the ξ_j independent and uniform on $[-\sqrt{3},\sqrt{3}]$, so egos whose closed neighborhoods overlap share random disturbances. The baseline is $\alpha_0(i) = W_i^\top\eta + \bar W_i^{+\top}\eta^+ + \bar W_i^{-\top}\eta^- + 0.5\sin(W_{i1}) + 0.8\zeta_i$. The effects are $\beta_0 = 1+0.3\tanh(W_{i2})$, $\beta_+ = 1+0.3\tanh(W_{i3})$, and $\beta_- = \lambda_-\{1+0.3\tanh(W_{i4})\}$. Coefficient vectors are drawn once from centered normal distributions with standard deviations 0.5 for γ, 0.4 for $\gamma^\pm$, 0.6 for η, and 0.4 for $\eta^\pm$, and fixed. We set $\lambda_- = 0.8$ and $\lambda_\times = -0.5$, allowing the two exposures to interact.

Estimation settings. SiDE is run with $F = 5$ folds and $P = 5$ fold partitions, with the training settings of Appendix C.

D.3 Baselines

Saturated regression (Vazquez-Bare, 2023) uses exposure-state indicators and their interactions with own treatment, with covariate adjustment. GPS (Forastiere et al., 2021) adjusts for individual and neighborhood generalized propensity scores. TARNet (Shalit et al., 2017), a treatment-agnostic representation network, learns a shared representation with treatment-specific outcome heads, and CFR (Shalit et al., 2017), counterfactual regression, adds a distribution-balancing penalty to TARNet. SAGE-HSIC (Ma and Tresp, 2021) averages treatment-masked neighbor representations and penalizes the dependence of the representation on own treatment with the Hilbert–Schmidt independence criterion (HSIC). NetEst (Jiang and Sun, 2022) uses adversarial representation learning to adjust for treatment and exposure. HINITE (Lin et al., 2023) uses attention over multiple edge types. DWR (Zhao et al., 2024) combines learned exposures with sample reweighting. GNN-DR (Leung and Loupos, 2025) combines graph neural outcome and propensity models in a doubly robust estimator. HINet (Caljon et al., 2026) uses treatment-aware message passing and adversarial balancing. GAIPW (Qu et al., 2026) uses tie signs as neighbor types in a generalized AIPW estimator with parametric propensity and outcome models. SAGE-HSIC, NetEst, HINITE, DWR, and HINet receive tie signs as two edge types, and GNN-DR receives them as an edge attribute. Saturated regression, GPS, TARNet, and CFR take separate positive- and negative-tie exposures.

D.4 Additional results

MAE is accompanied by the sample standard deviation of the absolute error across replications. Bold and underline are assigned from unrounded errors. Daggers mark effects with fewer than 30 expected egos in their scarcest exposure state.

MAE and bias on all networks. Table A3 compares MAE on the two Bitcoin networks. SiDE has the lowest MAE for the positive- and negative-tie effects on Bitcoin-OTC and the negative-tie effect

on Bitcoin-Alpha, while baselines are more accurate for the sparse interaction on both networks. Across the six networks and four effects in Table A4, SiDE has the smallest average absolute bias.

Table A3: MAE on the two Bitcoin networks for the methods of Table 1, with standard deviations of the absolute error below. Bold and underline mark the best and second-best values. Daggers mark sparse exposure states.

	Bitcoin-Alpha				Bitcoin-OTC			
Method	τ_+	$\tau_-{}^\dagger$	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$	τ_+	τ_-	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$
Sat. reg.	0.071 ±0.044	0.316 ±0.325	0.545 ±0.333	0.365 ±0.368	0.087 ±0.051	0.200 ±0.156	0.302 ±0.187	0.192 ±0.140
GPS	0.086 ±0.066	4.103 ±1.598	1.235 ±2.378	4.159 ±1.573	0.120 ±0.095	4.021 ±0.764	0.652 ±0.440	4.130 ±0.752
TARNet	0.075 ±0.047	0.489 ±0.165	0.270 ±0.042	0.622 ±0.174	0.085 ±0.044	0.421 ±0.063	**0.227** ±0.032	0.460 ±0.101
CFR	0.060 ±0.039	0.309 ±0.155	0.291 ±0.033	0.483 ±0.226	0.064 ±0.039	0.246 ±0.072	0.301 ±0.038	0.337 ±0.119
SAGE-HSIC	0.151 ±0.078	0.542 ±0.445	**0.166** ±0.118	0.933 ±0.429	0.200 ±0.076	0.412 ±0.211	0.229 ±0.158	0.741 ±0.250
NetEst	0.112 ±0.056	0.788 ±0.180	0.380 ±0.068	0.946 ±0.252	0.114 ±0.061	0.683 ±0.131	0.376 ±0.074	0.844 ±0.115
HINITE	**0.044** ±0.027	0.288 ±0.242	0.613 ±0.028	0.387 ±0.186	0.054 ±0.041	0.415 ±0.137	0.560 ±0.043	**0.096** ±0.090
DWR	0.071 ±0.055	0.679 ±0.300	0.429 ±0.056	0.479 ±0.288	0.051 ±0.046	0.667 ±0.250	0.387 ±0.119	0.572 ±0.251
GNN-DR	0.145 ±0.066	9.485 ±11.224	10.833 ±10.632	8.739 ±9.701	0.427 ±0.487	5.178 ±5.385	4.675 ±3.778	4.989 ±4.815
HINet	0.072 ±0.043	0.604 ±0.176	0.392 ±0.080	**0.202** ±0.138	0.054 ±0.033	0.661 ±0.132	0.425 ±0.043	0.304 ±0.115
GAIPW	0.075 ±0.048	0.276 ±0.333	1.143 ±0.912	0.718 ±0.474	0.093 ±0.061	0.761 ±0.731	1.396 ±1.477	0.892 ±0.637
SiDE (ours)	0.061 ±0.041	**0.272** ±0.268	0.497 ±0.320	0.362 ±0.261	**0.047** ±0.038	**0.174** ±0.110	0.567 ±0.606	0.188 ±0.221

Table A4: Bias on six networks. The eleven baselines use tie signs. Bold and underline mark the smallest and second-smallest absolute bias.

(a) Bias

	Slashdot				Epinions				Reddit			
Method	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ
Sat. reg.	+0.055	−0.123	+0.007	+0.180	+0.056	−0.123	−0.017	+0.178	+0.027	−0.129	+0.030	+0.156
GPS	+0.060	+1.532	+0.270	−1.579	+0.109	+2.191	+0.405	−2.331	+0.880	+2.490	+0.384	−2.540
TARNet	+0.042	−0.045	+0.047	+0.099	+0.012	−0.053	+0.170	+0.082	+0.017	−0.532	+0.298	+0.473
CFR	+0.032	−0.027	+0.015	+0.108	+0.012	−0.011	+0.106	+0.075	**+0.008**	−0.319	+0.240	+0.324
SAGE-HSIC	+0.134	−0.095	−0.162	+0.259	+0.073	−0.226	−0.090	+0.298	+0.440	−0.260	−0.087	+0.808
NetEst	+0.084	−0.444	+0.044	+0.709	+0.065	−0.383	+0.300	+0.614	−0.027	−0.788	+0.348	+0.835
HINITE	+0.058	−0.628	+0.369	+0.559	+0.067	−0.470	+0.407	+0.451	−0.018	−1.032	+0.415	+0.890
DWR	+0.030	−0.582	+0.230	+0.528	+0.018	−0.371	+0.423	+0.199	−0.033	−1.011	+0.389	+0.757
GNN-DR	+0.034	−0.328	+0.170	+0.480	+0.028	−0.156	**−0.013**	+0.413	+0.108	−0.902	**−0.015**	+1.966
HINet	−0.016	−0.313	+0.295	+0.145	**+0.002**	−0.320	+0.346	+0.167	−0.063	−0.956	+0.425	+0.662
GAIPW	+0.066	−0.139	−0.040	+0.204	+0.060	−0.157	+0.036	+0.215	+0.031	−0.070	−0.044	+0.054
SiDE	**+0.003**	**−0.001**	**−0.001**	**−0.002**	+0.006	**+0.009**	+0.021	**+0.003**	−0.017	**+0.001**	−0.031	**−0.051**

(b) Bias

	Wiki-Elec				Bitcoin-Alpha				Bitcoin-OTC			
Method	τ_+	τ_-	τ_{int}	κ	τ_+	$\tau_-{}^\dagger$	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$	τ_+	τ_-	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$
Sat. reg.	+0.055	−0.144	**+0.016**	+0.207	+0.032	−0.217	+0.084	+0.247	+0.030	−0.141	**+0.100**	+0.163
GPS	+0.154	+0.717	+0.217	−0.760	+0.045	+3.270	−1.104	−3.367	+0.088	+4.021	−0.652	−4.130
TARNet	+0.044	−0.054	+0.161	+0.191	+0.046	−0.489	+0.270	+0.622	+0.061	−0.421	+0.227	+0.460
CFR	+0.029	−0.096	+0.176	+0.185	+0.006	−0.309	+0.291	+0.483	+0.019	−0.246	+0.301	+0.337
SAGE-HSIC	+0.131	−0.138	−0.102	+0.288	+0.139	−0.529	**+0.013**	+0.933	+0.200	−0.412	+0.220	+0.741
NetEst	+0.065	−0.329	+0.054	+0.485	+0.078	−0.788	+0.380	+0.946	+0.100	−0.683	+0.376	+0.844
HINITE	**+0.004**	−0.730	+0.390	+0.468	**−0.002**	−0.152	+0.613	−0.291	+0.023	−0.415	+0.560	**+0.068**
DWR	+0.010	−0.462	+0.056	+0.301	−0.057	−0.679	+0.429	+0.452	+0.013	−0.667	+0.387	+0.572
GNN-DR	+0.096	**−0.018**	+0.050	**+0.004**	+0.071	+0.603	−1.195	−0.733	−0.221	−3.564	+2.974	+2.864
HINet	+0.024	−0.250	+0.280	+0.085	+0.044	−0.604	+0.392	+0.192	+0.016	−0.661	+0.425	+0.304
GAIPW	+0.056	−0.135	+0.029	+0.148	+0.040	+0.240	+0.041	−0.300	+0.028	−0.519	+1.155	+0.343
SiDE	+0.015	+0.055	−0.063	−0.050	−0.020	**−0.039**	+0.157	**+0.041**	**−0.010**	**−0.044**	+0.441	−0.143

Unsigned comparison. The Bitcoin comparison in Figure A2 indicates that removing tie signs attenuates the negative-tie effect on Bitcoin-Alpha but not on Bitcoin-OTC. Both versions estimate the sparse interaction imprecisely. Figures 3 and A2 compare two versions of one estimator trained on the same folds with one partition. Table A5 reports mean estimates on all six networks for the

eleven baselines fitted without tie signs and for SiDE, which uses the signs, and the sign-composition effect is zero by construction without signs.

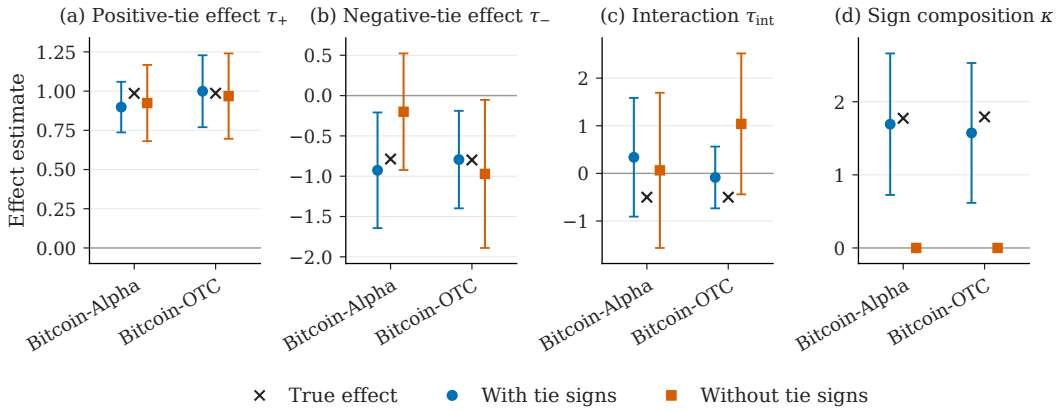

Figure A2: Effects with and without tie signs on the two Bitcoin networks, in the format of Figure 3.

Table A5: Mean effect estimates on six networks. The eleven baselines are fitted without tie signs. Bold and underline mark the smallest and second-smallest absolute bias. The sign-composition effect is zero by construction without signs and is not marked.

(a)

Method	Slashdot				Epinions				Reddit			
	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ
True value	+0.99	−0.80	−0.50	+1.79	+0.99	−0.80	−0.50	+1.79	+0.98	−0.79	−0.50	+1.77
Sat. reg.	+0.80	+0.80	−0.43	0.00	+0.81	+0.81	−0.37	0.00	+0.94	+0.94	−0.37	0.00
GPS	+0.98	+0.86	−0.48	0.00	+0.94	+0.89	−0.61	0.00	+1.32	+0.72	+0.20	0.00
TARNet	+0.89	+0.03	−0.33	0.00	+0.94	−0.37	−0.17	0.00	+0.96	+0.40	−0.27	0.00
CFR	+0.87	+0.08	−0.38	0.00	+0.94	−0.33	−0.33	0.00	+0.94	+0.42	−0.31	0.00
SAGE-HSIC	+1.01	−0.11	−0.55	0.00	+1.00	−0.60	**−0.49**	0.00	+1.38	+0.04	−0.28	0.00
NetEst	+1.25	+1.09	−1.46	0.00	+1.15	+0.87	−1.25	0.00	+1.07	+0.77	−0.24	0.00
HINITE	+0.87	−0.32	−0.11	0.00	+0.96	−0.53	−0.02	0.00	+0.91	+0.05	−0.03	0.00
DWR	+0.83	+0.31	−0.21	0.00	+0.89	+0.41	−0.09	0.00	+0.86	+0.43	−0.04	0.00
GNN-DR	+0.86	−0.10	−0.30	0.00	+0.95	−0.43	−0.40	0.00	+0.90	−0.05	−0.17	0.00
HINet	+0.85	+0.34	−0.15	0.00	+0.95	−0.42	0.00	0.00	+0.89	+0.21	−0.02	0.00
GAIPW	+0.92	−0.15	−0.36	0.00	**+0.99**	−0.51	−0.37	0.00	**+0.98**	+0.16	−0.28	0.00
SiDE (ours)	**+0.99**	**−0.80**	**−0.50**	+1.78	+0.99	**−0.79**	−0.48	+1.79	+0.97	**−0.79**	**−0.53**	+1.72

(b)

Method	Wiki-Elec				Bitcoin-Alpha				Bitcoin-OTC			
	τ_+	τ_-	τ_{int}	κ	τ_+	$\tau_-{}^\dagger$	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$	τ_+	τ_-	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$
True value	+0.99	−0.80	−0.50	+1.77	+0.99	−0.79	−0.50	+1.77	+0.99	−0.80	−0.50	+1.79
Sat. reg.	+0.59	+0.59	−0.39	0.00	**+0.99**	+0.99	−0.44	0.00	+0.94	+0.94	−0.41	0.00
GPS	+0.76	+0.65	−0.45	0.00	+1.04	+0.97	−0.32	0.00	+1.03	+0.89	−0.33	0.00
TARNet	+0.73	+0.06	−0.34	0.00	+1.02	+0.91	−0.61	0.00	+1.01	+0.73	**−0.49**	0.00
CFR	+0.72	+0.12	−0.38	0.00	+0.97	+1.02	−0.73	0.00	+0.95	+0.64	−0.42	0.00
SAGE-HSIC	+0.70	+0.32	**−0.53**	0.00	+1.11	+1.02	**−0.53**	0.00	+1.14	+0.88	−0.49	0.00
NetEst	+0.80	+0.47	−0.65	0.00	+1.27	+1.08	−0.84	0.00	+1.29	+1.25	−1.20	0.00
HINITE	+0.39	+0.33	−0.04	0.00	+0.97	+0.65	−0.07	0.00	+0.96	+0.62	−0.06	0.00
DWR	+0.51	+0.37	−0.01	0.00	+0.81	+0.41	0.00	0.00	+0.82	+0.35	−0.01	0.00
GNN-DR	+0.92	−0.17	−0.67	0.00	+0.98	+0.19	−1.23	0.00	+1.01	−0.30	+0.77	0.00
HINet	+0.54	+0.44	−0.05	0.00	+1.02	+0.67	−0.17	0.00	+0.97	+0.68	−0.12	0.00
GAIPW	+0.84	−0.31	−0.38	0.00	+1.02	+0.35	−0.59	0.00	+0.96	−0.40	+0.67	0.00
SiDE (ours)	**+1.00**	**−0.74**	−0.56	+1.72	+0.97	**−0.83**	−0.34	+1.81	**+0.98**	**−0.84**	−0.06	+1.65

Hidden width. Figure A3 compares SiDE's MAE across widths 32, 64, and 128. On the four main networks, the MAE of each effect moves by less than two Monte Carlo standard errors, and on the two Bitcoin networks only the sparse sign-composition effect on Bitcoin-OTC moves by slightly more. The outcome and assignment models use the same width, with every other setting fixed. All three widths are rerun in one batch with the data seeds and fold partitions of Table 1, and comparisons across widths use this batch only.

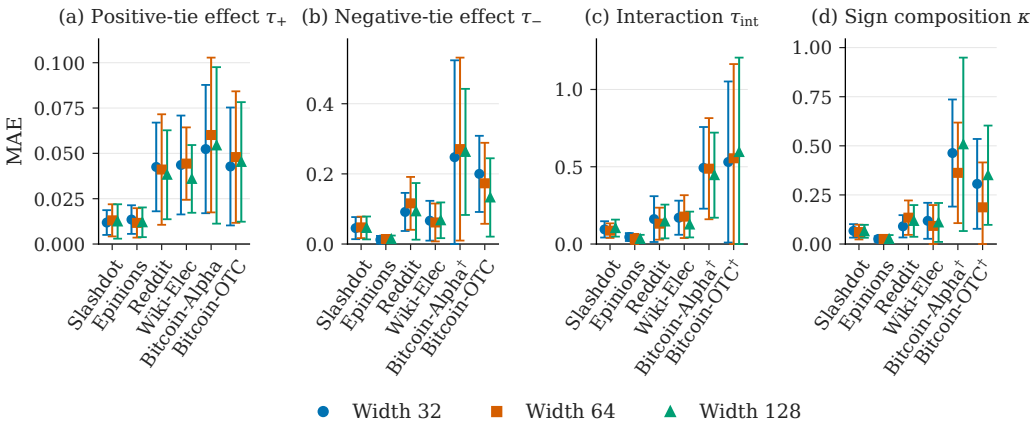

Figure A3: Sensitivity of SiDE to hidden width on six networks. Each panel shows one effect. Markers show MAE and bars show one standard deviation of the absolute error. Daggers mark sparse exposure states.

Runtime. The runtimes in Table A6 cover model fitting and effect estimation after preprocessing. One complete SiDE run takes 11–94 seconds on one NVIDIA RTX 6000 Ada GPU.

Table A6: SiDE running times. Egos is the number of target egos in the union of reachable sets, and Replication (s) is the time of one complete SiDE run, a mean with standard deviation over 5 runs. Epoch time is averaged over all fitted models.

Network	Units	Ties	Egos	Setup (s)	Replication (s)	Epochs per fit	Epoch (ms)
Bitcoin-Alpha	3,783	23,690	2,456	0.1	$11_{\pm 1}$	67	3.2
Bitcoin-OTC	5,881	34,876	3,661	0.2	$12_{\pm 1}$	74	3.3
Wiki-Elec	7,115	103,091	4,137	0.7	$12_{\pm 1}$	75	3.3
Reddit	30,744	271,667	19,471	0.1	$30_{\pm 3}$	165	3.6
Slashdot	82,140	545,304	29,789	0.1	$49_{\pm 1}$	167	5.9
Epinions	131,580	835,393	78,320	0.2	$94_{\pm 4}$	201	9.4

Ablations. The pattern of Table 2 holds on all six networks of Table A7. Averaged over networks and effects, the MAE is 0.136 for SiDE, 0.153 for w/o weighting, 0.359 for w/o PB, and 0.501 for w/o outcome, and the average absolute bias of w/o weighting is 0.129 against 0.051 for SiDE. All four methods aggregate the same five fold partitions of each replication. The variants w/o weighting and w/o PB share SiDE's outcome fit, while w/o outcome shares its assignment fit. The variant w/o PB fits a separate joint-state probability model.

Coverage details. On the four main networks, the average standard error of the intervals lies between 0.76 and 1.14 times the standard deviation of the estimates, as shown in Table A8, and no replication needs the positive semidefinite fallback of the variance estimator.

Bitcoin networks. Table A9 reports undercoverage for the positive-tie effect: 0.880 on Bitcoin-Alpha and 0.840 on Bitcoin-OTC. For the sparse effects, the standard errors exceed the standard deviation of the estimates, with coverage at least 0.955.

E Case study details

We use the ICPSR (Inter-university Consortium for Political and Social Research) archive, study 37070 (Paluck et al., 2020) and retain students in the 28 program schools with valid identifiers and recorded invitation status. The sample contains 11,117 students, including 728 invited students. Wave-1 spend-time and conflict nominations, the measures analyzed by Callejas and Shepherd (2020), produce 85,085 spend-time ties and 16,195 conflict ties. Giving conflict priority for the 3,428 overlapping nominations leaves 81,657 positive and 16,195 negative directed ties. We retain the archived orange-wristband indicator `WRISTOW2`. Among 1,589 students without a valid answer to the underlying wristband question, this derived indicator records 1,586 zeros and 3 ones. The

Table A7: Ablation results on six networks: (a, b) MAE with standard deviations of the absolute error below and (c, d) bias. Variants are named as in Table 2. Bold and underline mark the smallest and second-smallest MAE in (a, b) and absolute bias in (c, d). Daggers mark sparse exposure states.

(a) MAE and standard deviation

	Slashdot				Epinions				Reddit			
Variant	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ
SiDE	**0.013** ±0.009	0.049 ±0.028	0.085 ±0.045	0.064 ±0.034	**0.012** ±0.008	**0.014** ±0.009	**0.032** ±0.033	**0.027** ±0.018	0.042 ±0.031	0.118 ±0.075	0.134 ±0.104	**0.128** ±0.083
w/o weighting	0.015 ±0.004	**0.035** ±0.021	**0.053** ±0.046	**0.050** ±0.023	0.012 ±0.009	0.017 ±0.011	0.060 ±0.038	0.035 ±0.022	**0.040** ±0.027	**0.073** ±0.055	**0.112** ±0.089	0.143 ±0.060
w/o outcome	0.114 ±0.109	0.156 ±0.145	0.218 ±0.165	0.248 ±0.293	0.060 ±0.028	0.147 ±0.102	0.233 ±0.273	0.156 ±0.116	0.238 ±0.108	0.829 ±0.674	0.507 ±0.436	0.991 ±0.557
w/o PB	0.025 ±0.024	0.138 ±0.103	0.204 ±0.132	0.106 ±0.075	0.014 ±0.009	0.042 ±0.023	0.186 ±0.080	0.091 ±0.066	0.189 ±0.161	0.714 ±0.858	0.842 ±0.890	0.378 ±0.397

(b) MAE and standard deviation

	Wiki-Elec				Bitcoin-Alpha				Bitcoin-OTC			
Variant	τ_+	τ_-	τ_{int}	κ	τ_+	$\tau_-{}^\dagger$	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$	τ_+	τ_-	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$
SiDE	0.042 ±0.022	**0.060** ±0.052	0.195 ±0.135	**0.093** ±0.097	**0.061** ±0.041	**0.272** ±0.268	0.497 ±0.320	**0.362** ±0.261	**0.047** ±0.038	**0.174** ±0.110	0.567 ±0.606	**0.188** ±0.221
w/o weighting	**0.033** ±0.025	0.086 ±0.074	**0.121** ±0.065	0.167 ±0.117	0.076 ±0.068	0.463 ±0.159	**0.273** ±0.174	0.709 ±0.150	0.053 ±0.045	0.334 ±0.234	**0.228** ±0.167	0.488 ±0.272
w/o outcome	0.115 ±0.053	0.243 ±0.204	0.369 ±0.400	0.319 ±0.194	0.163 ±0.101	0.983 ±0.627	1.122 ±0.829	1.358 ±0.826	0.107 ±0.116	0.811 ±0.832	1.651 ±1.045	0.894 ±0.750
w/o PB	0.131 ±0.294	0.164 ±0.120	0.436 ±0.302	0.281 ±0.112	0.102 ±0.093	0.548 ±0.474	0.962 ±0.877	0.634 ±0.604	0.230 ±0.201	0.624 ±0.347	0.850 ±0.491	0.726 ±0.687

(c) Bias

	Slashdot				Epinions				Reddit			
Variant	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ	τ_+	τ_-	τ_{int}	κ
SiDE	+0.003	**−0.001**	**−0.001**	**−0.002**	**+0.006**	+0.009	**+0.021**	**+0.003**	**−0.017**	**+0.001**	−0.031	**−0.051**
w/o weighting	+0.004	+0.026	+0.045	−0.013	+0.008	+0.013	+0.058	−0.015	−0.026	+0.021	+0.105	−0.143
w/o outcome	**+0.002**	−0.007	−0.026	+0.025	+0.034	−0.040	+0.172	+0.009	+0.159	+0.261	**−0.012**	+0.099
w/o PB	−0.014	−0.088	+0.083	−0.015	+0.013	**+0.006**	+0.077	+0.022	−0.044	+0.163	+0.106	−0.198

(d) Bias

	Wiki-Elec				Bitcoin-Alpha				Bitcoin-OTC			
Variant	τ_+	τ_-	τ_{int}	κ	τ_+	$\tau_-{}^\dagger$	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$	τ_+	τ_-	$\tau_{\text{int}}{}^\dagger$	$\kappa^\dagger$
SiDE	**+0.015**	**+0.055**	−0.063	−0.050	−0.020	**−0.039**	+0.157	**+0.041**	**−0.010**	**−0.044**	+0.441	−0.143
w/o weighting	−0.033	+0.073	+0.112	−0.167	−0.074	+0.463	**+0.015**	−0.709	−0.029	+0.332	**+0.116**	−0.488
w/o outcome	+0.070	−0.121	**−0.021**	**+0.026**	+0.059	−0.087	−0.191	+0.460	+0.092	−0.156	+0.714	**+0.009**
w/o PB	+0.078	+0.148	−0.256	−0.281	**+0.004**	+0.095	+0.292	−0.277	+0.030	−0.207	+0.605	+0.032

indicator has no missing-value code. We use it as recorded for all students, including these students. Dropping them gives a positive-tie estimate of 3.7 percentage points with linked-pair interval [0.6, 6.8].

Half of the invitation-eligible students were selected within each school randomization block. We compute joint state probabilities from multivariate hypergeometric distributions convolved over blocks. Each contrast includes students for whom every compared state has positive design probability, giving 6,375 students for the positive-tie effect, 1,659 for the negative-tie effect, and 941 for the interaction and sign-composition effects. The binary outcome model uses width 32 and at most 300 epochs, with covariates comprising gender, grade, missing-value indicators, signed degrees, invitation eligibility, and eligible-neighbor counts. The sign-blind comparison of Section 5.5 uses the 7,093 students with at least one invitation-eligible named schoolmate and has no standard error. Cross-fitting holds out entire schools. We use the normalized score and aggregate repeated school partitions by the median rule. This aggregation rule was selected after inspecting initial estimates, so the analysis is exploratory. We report approximate linked-pair intervals and apply a Bonferroni adjustment across the four contrasts (Dunn, 1961). For the spend-time effect, the school-clustered interval is [−0.1, 6.6], and the Bonferroni-adjusted linked-pair interval is [−0.3, 6.9]. Both include zero.

Table A10 reports all four effects with linked-pair intervals and school-clustered intervals, for the main analysis and for two sensitivity analyses, which drop non-respondents or count the overlapping nominations as spend-time ties. The school-clustered standard errors allow any dependence within a school, including the dependence created by drawing invitations at a fixed share within randomization

Table A8: Coverage details on the four main networks. Coverage is the fraction of the 200 datasets whose interval, aggregated over five fold partitions, contains the true effect. The 95% CP interval is the Clopper–Pearson (CP) confidence interval for coverage. Length is the average interval length, and SE/SD and Bias/SD divide the average standard error and the bias by the standard deviation of the estimates. Underlined values are below 0.919.

Network	Effect	Coverage	95% CP interval	Length	SE/SD	Bias/SD
Slashdot	τ_+	0.930	[0.885, 0.961]	0.077	0.92	0.05
	τ_-	0.955	[0.916, 0.979]	0.201	0.97	0.11
	τ_{int}	0.965	[0.929, 0.986]	0.403	1.08	0.11
	κ	0.945	[0.904, 0.972]	0.270	1.08	−0.17
Epinions	τ_+	<u>0.905</u>	[0.856, 0.942]	0.060	0.92	0.20
	τ_-	0.930	[0.885, 0.961]	0.129	0.94	−0.07
	τ_{int}	0.965	[0.929, 0.986]	0.387	1.02	−0.05
	κ	0.965	[0.929, 0.986]	0.242	1.03	0.00
Reddit	τ_+	<u>0.850</u>	[0.793, 0.896]	0.104	0.76	−0.28
	τ_-	0.935	[0.891, 0.965]	0.627	1.01	0.02
	τ_{int}	0.945	[0.904, 0.972]	1.032	1.07	0.09
	κ	0.985	[0.957, 0.997]	0.716	1.14	−0.05
Wiki-Elec	τ_+	0.960	[0.923, 0.983]	0.222	1.03	−0.14
	τ_-	0.920	[0.873, 0.954]	0.407	0.96	0.37
	τ_{int}	0.970	[0.936, 0.989]	0.985	1.07	0.01
	κ	0.940	[0.898, 0.969]	0.645	1.05	−0.36

Table A9: Coverage details on the two Bitcoin networks, with the columns of Table A8. The 95% CP interval is the Clopper–Pearson (CP) confidence interval for coverage. Daggers mark sparse exposure states.

Network	Effect	Coverage	95% CP interval	Length	SE/SD	Bias/SD
Bitcoin-Alpha	τ_+	<u>0.880</u>	[0.827, 0.922]	0.251	0.88	−0.48
	$\tau_-{}^{\dagger}$	0.955	[0.916, 0.979]	2.362	1.13	0.15
	$\tau_{\mathrm{int}}{}^{\dagger}$	0.980	[0.950, 0.995]	4.379	1.17	−0.06
	$\kappa^{\dagger}$	0.965	[0.929, 0.986]	2.796	1.20	−0.16
Bitcoin-OTC	τ_+	<u>0.840</u>	[0.782, 0.888]	0.205	0.78	−0.34
	τ_-	0.965	[0.929, 0.986]	1.342	1.21	0.24
	$\tau_{\mathrm{int}}{}^{\dagger}$	0.995	[0.972, 1.000]	2.640	1.23	0.01
	$\kappa^{\dagger}$	0.960	[0.923, 0.983]	1.684	1.22	−0.26

blocks. They multiply the cluster-robust variance by 28/27, the usual small-sample correction for 28 school clusters, and use the t distribution with 27 degrees of freedom (Cameron and Miller, 2015), so the interval is the estimate plus or minus 2.09 times the uncorrected cluster-robust standard error. We also estimate the four effects for two secondary outcomes, reported conflict with other students and whether friends talked with the student about changing how students act, each coded as any positive answer against none, in the program schools and in the control schools. None of the resulting 16 estimates has an interval that excludes zero under either standard error. The wristband outcome is almost never positive in control schools (mean 0.001), so it provides no placebo test.

Table A10: Case-study estimates in percentage points, with approximate 95% intervals. Students is the size of the contrast population.

Analysis	Effect	Students	Estimate	Linked-pair interval	School-clustered interval
Main analysis	τ_+	6,375	3.3	[0.5, 6.1]	[−0.1, 6.6]
	τ_-	1,659	−0.9	[−7.2, 5.5]	[−6.6, 4.8]
	τ_{int}	941	10.3	[−2.4, 22.9]	[−0.7, 21.3]
	κ	941	−0.3	[−7.8, 7.2]	[−9.8, 9.2]
Non-respondents dropped	τ_+	5,521	3.7	[0.6, 6.8]	[−0.2, 7.7]
	τ_-	1,435	−0.4	[−7.9, 7.0]	[−7.5, 6.7]
	τ_{int}	819	12.1	[−2.2, 26.5]	[−0.7, 24.9]
	κ	819	−1.8	[−10.5, 7.0]	[−13.1, 9.5]
Overlaps as spend-time ties	τ_+	6,588	3.3	[0.6, 6.0]	[0.0, 6.6]
	τ_-	1,302	−2.6	[−10.1, 4.8]	[−10.1, 4.8]
	τ_{int}	797	14.4	[−0.3, 29.1]	[0.7, 28.2]
	κ	797	0.9	[−7.3, 9.0]	[−10.1, 11.8]