Appendices to *Causal Inference for Social Network Data*

Elizabeth L. Ogburn

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

E-mail: eogburn@jhsph.edu

Oleg Sofrygin

Department of Biostatistics, University of California, Berkeley, CA, USA

Iván Díaz

Division of Biostatistics and Epidemiology, Weill Cornell Medicine, New York, NY, USA

Mark J. van der Laan

Department of Biostatistics, University of California, Berkeley, CA, USA

Appendix I : Estimation procedure

Below we propose a targeted maximum loss-based estimator (TMLE) of ψ, however all of the results that follow are equally applicable to a standard estimating equation approach. TMLEs are substitution estimators and are not as sensitive to the near violations of the positivity assumption that can occur in finite samples and result in extreme values of $\bar{h}_{x^*}(v_i)/\bar{h}(v_i)$. Targeted maximum likelihood estimation is a general template for estimation of smooth parameters in semi- and nonparametric models. The estimation algorithm is constructed to solve the efficient influence function estimating equation, thereby yielding, under regularity conditions, asymptotically linear estimators with the same semiparametric efficiency property as the estimating equation approach described above. In our setting, a TMLE is constructed using three elements: (i) a valid loss function L for the outcome regression model m, (ii) initial working estimators $\hat{m}$ of m and and $\hat{g}$ of g, and (iii) a parametric submodel m_ϵ of $\mathcal{M}$, the score of which corresponds to a particular component of the score based on the efficient influence function $D(\mathbf{o})$ and such that $m_0 = m(\cdot)$. The TMLE is then defined by an iterative procedure that, at each step, estimates ϵ by minimizing the empirical risk of the loss function L at m_ϵ. An updated estimate is then computed as $\hat{m}_{\hat{\epsilon}}$, and the process is repeated until convergence. The TMLE is the estimator obtained in the final step of the iteration. The result of the previous iterative procedure is that, at the final step, the efficient influence function estimating equation is solved. For more details about targeted maximum likelihood estimation, see Van der Laan and Rose (2011). In the present setting, the TMLE for ψ based on $D'(\mathbf{o})$ requires only one iteration for convergence (Van der Laan and Rose, 2011). We use influence function $D'(\mathbf{o})$ to derive the TMLE, instead of $D(\mathbf{o})$, because it is computationally more tractable and because the choice of influence function does not matter for the conditional parameter that we are interested in when latent variable dependence is present.

Initial estimators $\hat{m}$ and $\hat{g}$ of m and g may be found through maximum likelihood or loss-based estimation methods like standard regression models; under the conditions required for Theorem 1 to hold, a similar argument shows that m-estimator for either of the nuisance models will be CAN for its expectation. Under a conditional independence structure analogous to that implied by assumptions (A1), (A4), and (A5), Benkeser et al. (2018) showed that super learning (van der Laan Mark et al., 2007) can be used to estimate the nuisance models. The empirical distribution $\hat{p}_C$ is used to estimate p_C. An estimate $\hat{\bar{h}}$ of $\bar{h}(v)$ optimizes the log likelihood function $\sum_{i=1}^{n} \log \bar{h}(V_i|W_i)$, as if the pooled sample (V_i, W_i) were i.i.d. It can be shown that this results in a valid loss function for $\bar{h}$, even for dependent observations (V_i, W_i), for $i = 1, \ldots, n$ (van der Laan, 2014; Sofrygin and van der Laan, 2015). Similarly, one can construct a direct estimator $\hat{\bar{h}}_{x^*}$ of $\bar{h}_{x^*}$, by first creating a sample (V_i^*, W_i) and then directly optimizing the log likelihood function $\sum_{i=1}^{n} \log \bar{h}_{x^*}(V_i^*|W_i)$, as if the pooled sample (V_i^*, W_i) were i.i.d. We perform estimation of the conditional mixture density $\bar{h}$ using a conditional histogram approach, previously described for i.i.d. data in Munoz and van der Laan (2011). The approach relies on fitting the conditional hazards of individual bins from the support of V_i (given W_i) using separate parametric logistic regression models.

In our highly-dependent network settings, the operational characteristics of the direct estimator of $\bar{h}$ are unclear. Similarly, it is unclear how to appropriately conduct cross-validation with our proposed direct estimation approach for $\bar{h}$. However, lacking any other reasonable estimation alternatives, we believe that the enormous computational advantages offered by this direct estimation route, along with the encouraging results obtained from our extensive simulations, merit the description of this estimator. We also realize that more theoretical work is needed to justify and improve upon this approach. For additional simulation results that demonstrate the performance of the direct estimation approach for mixture density $\bar{h}$, we refer to Sofrygin et al. (2017, 2018).

Now the TMLE of ψ is computed as follows:

(a) Define the auxiliary weights H_i as the ratio of estimated densities of V^* and V evaluated at the observed value V_i. Compute the auxiliary weights as

$$H_i = \frac{\hat{\bar{h}}_{x^*}(V_i)}{\hat{\bar{h}}(V_i)}.$$

(b) Compute initial predicted outcome values $\hat{Y}_i \equiv \hat{m}(V_i)$ and predicted potential outcome values $\hat{Y}_i^* \equiv \hat{m}(V_i^*)$ evaluated at the counterfactual value $V_i^* = s_{Y,i}(\mathbf{C}, \mathbf{x}^*)$.

(c) Construct a TMLE model update $\hat{m}_{\hat{\epsilon}}$ of $\hat{m}$ by running a weighted intercept-only logistic regression model with weights H_i defined in step (1), Y_i as the outcome and including $\hat{Y}_i$ as an offset. That is, define $\hat{\epsilon}$ as the estimate of the intercept parameter ϵ from the following *weighted* logistic regression model

$$\text{logit}\hat{m}_\epsilon(v) = \text{logit}\hat{m}(v) + \epsilon,$$

where $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$.

(d) Compute updated predicted potential outcomes $\tilde{Y}_i^*$ as the fitted values of the regression from step (c), evaluated at v^* rather than v (that is, at $\hat{Y}_i^*$ instead of $\hat{Y}_i$):

$$\tilde{Y}_i^* = \text{expit}\{\text{logit}\hat{Y}_i^* + \hat{\epsilon}\},$$

where $\text{expit}(x) = \frac{1}{1+e^{-x}}$, i.e., the inverse of the logit function.

(e) Compute the TMLE $\hat{\psi}$ as

$$\hat{\psi} = \frac{1}{n}\sum_{i=1}^{n} \tilde{Y}_i^*.$$

The TMLE is doubly robust: it will be consistent for ψ if either the working model $\hat{g}$ for g or the working model $\hat{m}$ for m is correctly specified. This resulting estimator remains CAN for ψ under assumptions (A2) and (A3) instead of (A4) and (A5), and the same procedure can be used to estimate the parameter conditional on $\mathbf{C}$.

Appendix II : Proof of Theorem 1

A. Regularity conditions

For a real-valued function $\mathbf{c} \mapsto f(\mathbf{c})$, let the $L^2(P)$-norm of $f(\mathbf{c})$ be denoted by $\|f\| = E[f(\mathbf{C})^2]^{1/2}$. Define $\mathcal{M}_m$ and $\mathcal{M}_{\tilde{h}}$ as the classes of possible functions that can be used for estimating the two nuisance parameters m and $\tilde{h} \equiv \bar{h}_{x^*}/\bar{h}$, respectively. Note that a model for g plus the empirical distribution of covariates $\mathbf{C}$ determines $\tilde{h}$. Equivalent assumptions could be stated in terms of g instead of $\tilde{h}$, but we focus on $\tilde{h}$ because that is the functional of g and $\mathbf{C}$ that we model in our estimating procedure. Assume that the TMLE update $\hat{m}_{\hat{\epsilon}} \in \mathcal{M}_m$ with probability 1 and assume that $\hat{\bar{h}}_{x^*}/\hat{\bar{h}} \in \mathcal{M}_{\tilde{h}}$ with probability 1. Finally, define the following dissimilarity measure on the cartesian product of $\mathcal{F} \equiv \mathcal{M}_m \times \mathcal{M}_{\tilde{h}}$:

$$d\left((h,m),\left(\tilde{h},\tilde{m}\right)\right) = \max\left(\sup_{v\in\mathcal{V}} \mid h-\tilde{h} \mid (v), \sup_{v\in\mathcal{V}} \mid m-\tilde{m} \mid (v)\right).$$

The following are the regularity conditions required for Theorem 1, i.e. for asymptotic normality of the TMLE $\hat{\psi}^*$.

Uniform consistency: Assume that

$$d\left(\left(\hat{\bar{h}}_{x^*}/\hat{\bar{h}}, \hat{m}_{\hat{\epsilon}}\right), \left(\bar{h}_{x^*}/\bar{h}, m\right)\right) \to 0$$

in probability as $n \to \infty$. Note that this assumption is only needed for proving the asymptotic equicontinuity of our process; it is not needed for proofs of relevant convergence rates for the second order terms.

Bounded entropy integral: Assume that there exists some $\eta > 0$, so that $\int_0^\eta \sqrt{\log\left(N(\epsilon,\mathcal{F},d)\right)}d\epsilon < \infty$, where $N(\epsilon,\mathcal{F},d)$ is the number of balls of size ϵ w.r.t. metric d needed to cover $\mathcal{F}$.

Universal bound: Assume $\sup_{f\in\mathcal{F},\mathbf{O}} \mid f \mid (\mathbf{O}) < \infty$, where the supremum of $\mathbf{O}$ is over a set that contains $\mathbf{O}$ with probability one. This assumption will typically be a consequence of the choosing a specific function class $\mathcal{F}$ that satisfies the above entropy condition.

Positivity: Assume

$$\sup_{v\in\mathcal{V}} \frac{\bar{h}_{x^*}(v)}{\bar{h}(v)} < \infty.$$

Consistency and rates for estimators of nuisance parameters: Assume that $\|\hat{m} - m\| \left\|\hat{\bar{h}} - \bar{h}\right\| = o_P\left((C_n)^{-1/2}\right)$. Note that this rate is achievable if, for example, estimation of $\bar{h}$ relies on some pre-specified parametric model, or if both $\bar{h}$ and m are estimated at rate $C_n^{-1/4}$.

Rate of the second order term: Assume that

$$R_{n1} \equiv -\int_v \left\{ \left(\frac{\hat{\bar{h}}_{x^*}}{\hat{\bar{h}}} - \frac{\bar{h}_{x^*}}{\bar{h}} \right) (\hat{m}_{\hat{\epsilon}} - m)(v)\bar{h}(v)d\mu(v) \right\} = o_P\left(1/\sqrt{C_n}\right).$$

Note that this condition is provided here purely for the sake of completeness, since it will satisfied based on the previously assumed rates of convergence for $\|\hat{m} - m\| \left\|\hat{\bar{h}} - \bar{h}\right\|$. This follows from the fact that the parametric TMLE update step $\hat{m}_{\hat{\epsilon}}$ of $\hat{m}$ will have a negligible effect on the rate of convergence of the initial estimator $\hat{m}$, that is, $\hat{m}_{\hat{\epsilon}}$ will converge at "nearly" the same rate as $\hat{m}$.

Limited connectivity and limited dependence of Y,X and C: Let $K_{max,n} = max_i\{K_i\}$ for a fixed network with n nodes. Assume that $K^2_{max,n}/n$ converges to 0 in probability as $n \to \infty$.

A key condition is *consistency and rates for estimators of nuisance parameters.* This condition will be satisfied, for example, if both models converge to the truth at rate $C_n^{1/4}$. It can in fact be weakened, but for a more general discussion and the corresponding technical conditions we refer to the Appendix of van der Laan (2014). With the exception of the rates of convergence, the more general conditions for asymptotic normality of the TMLE presented in that paper apply to our setting as well.

B. Overview of the proof of Theorem 1

We want to show that $\sqrt{C_n}(\hat{\psi}-\psi)$ converges in law to a Normal limit as n goes to infinity for some rate $\sqrt{C_n}$ such that $\sqrt{n/\left(K_{max}(n)\right)^2} \leq \sqrt{C}_n \leq \sqrt{n}$, where the rate $\sqrt{C_n}$ is the order of the variance of the sum of the first-order linear approximation of $(\hat{\psi}-\psi)$.

Broadly, the proof has two parts: First, we require that the second order terms in the expansion of $\hat{\psi}-\psi$ are stochastically less than $1/\sqrt{C_n}$, that is that

$$\hat{\psi}_n - \psi \;=\; \frac{1}{n}\sum_{i=1}^{n}\left\{f_i(\mathbf{O}) - E[f_i(\mathbf{O})]\right\} + o_p\left(1/\sqrt{C_n}\right),$$

where $f_i(\mathbf{O})$ is the contribution of the ith observation to the estimator. Specifically, for our influence function

$$D(\mathbf{o}) \;=\; \sum_{j=1}^{n}\frac{1}{n}\sum_{i=1}^{n} E\left[m\left(V_i^*\right) \mid C_j = c_j\right] - \psi + \frac{1}{n}\sum_{i=1}^{n}\frac{\bar{h}_{x^*}(v_i)}{\bar{h}(v_i)}\left\{y_i - m\left(v_i\right)\right\},$$

the contribution of the ith observation is

$$f_i(\mathbf{o}) \;=\; \sum_{j=1}^{n} E\left[m\left(V_i^*\right) \mid C_j = c_j\right] + \frac{\bar{h}_{x^*}(v_i)}{\bar{h}(v_i)}\left\{y_i - m\left(v_i\right)\right\}.$$

Then proving asymptotic normality of the TMLE amounts to the asymptotic analysis of the sum $\frac{1}{n}\sum_{i=1}^{n}\left\{f_i(\mathbf{O}) - E[f_i(\mathbf{O})]\right\}$, and the second part of the proof establishes that the first order terms converge to a normal distribution when scaled by $\sqrt{C_n}$, that is that $\sqrt{C_n}\frac{1}{n}\sum_{i=1}^{n}\left\{f_i(\mathbf{O}) - E[f_i(\mathbf{O})]\right\} \rightarrow_d N(0,\sigma^2)$ for some finite σ^2.

The proof that the second order terms are stochastically less than $1/\sqrt{C_n}$ is an extension of the empirical process theory of Van Der Vaart and Wellner (1996) and follows the same format as the proof in van der Laan (2014). Indeed, the proof offered by van der Laan (2014) holds immediately after replacing the rate or scaling factor $\sqrt{n}$ with $\sqrt{C_n}$ throughout. Only one step in the van der Laan (2014) proof relies on the network structure, which is the major difference between the setting in that paper, where the number of network connections is fixed and bounded as n goes to infinity, and the present setting: the proof requires bounding the Orlicz norms of several empirical processes corresponding to components of the influence function for ψ, and a key step is bounding the expectation of $E\left[|X_n(f)|^p\right]$, where $X_n(f)$ is the stochastic process that describes the difference between the empirical (indexed by n) and the true distribution functions of a component of the influence function for ψ. This step relies on a combinatorial argument about nature of overlapping friend groups in the underlying network, and the argument for the case of growing K_i is subsumed by the argument for fixed K in van der Laan (2014).

The proof that the first order terms converge to a normal distribution requires a central limit theorem for dependent data with growing and possibly irregularly sized dependency neighborhoods, where a dependency neighborhood for unit i is a collection of observations on which the observations for unit i may be dependent. We prove such a CLT in Lemmas 1 and 2. In the next section we use the CLT for growing and irregular dependency neighborhoods, along with an orthogonal decomposition of the first order terms, to prove the remainder of Theorem 1.

C. Central limit theorem for first order terms

Proving asymptotic normality of the TMLE amounts to the asymptotic analysis of the sum $\frac{1}{n}\sum_{i=1}^{n}\{f_i(\mathbf{O}) - E[f_i(\mathbf{O})]\}$. As a start, decompose $\sum_{i=1}^{n}\{f_i(\mathbf{O}) - E[f_i(\mathbf{O})]\}$ into a sum of three orthogonal components:

$$\begin{aligned} f_{\mathbf{Y},i}(\mathbf{Y},\mathbf{X},\mathbf{C}) &= f_i(\mathbf{O}) - E\left[f_i(\mathbf{O}) \mid \mathbf{X},\mathbf{C}\right], \\ f_{\mathbf{X},i}(\mathbf{X},\mathbf{C}) &= E[f_i(\mathbf{O}) \mid \mathbf{X},\mathbf{C}] - E[f_i(\mathbf{O}) \mid \mathbf{C}], \text{ and} \\ f_{\mathbf{C},i}(\mathbf{C}) &= E[f_i(\mathbf{O}) \mid \mathbf{C}] - E[f_i(\mathbf{O})]. \end{aligned}$$

Note that

$$f_i(\mathbf{O}) - E[f_i(\mathbf{O})] = f_{\mathbf{Y},i}(\mathbf{Y},\mathbf{X},\mathbf{C}) + f_{\mathbf{X},i}(\mathbf{X},\mathbf{C}) + f_{\mathbf{C},i}(\mathbf{C})$$

and with slight abuse of notation we will also write $f_{\mathbf{Y},i}(\mathbf{O})$, $f_{\mathbf{X},i}(\mathbf{O})$ and $f_{\mathbf{C},i}(\mathbf{O})$. Let $f_{\mathbf{Y}}(\mathbf{O}) = \sum_{i=1}^{n} f_{\mathbf{Y},i}(\mathbf{O})$, $f_{\mathbf{X}}(\mathbf{O}) = \sum_{i=1}^{n} f_{\mathbf{X},i}(\mathbf{O})$ and $f_{\mathbf{C}}(\mathbf{O}) = \sum_{i=1}^{n} f_{\mathbf{C},i}(\mathbf{O})$. For $i = 1, \ldots, n$, let

$$\begin{aligned} Z_{Y,i} &= \frac{f_{\mathbf{Y},i}(\mathbf{Y},\mathbf{X},\mathbf{C})}{\sqrt{Var(\sum_{i=1}^{n} f_{\mathbf{Y},i}(\mathbf{Y},\mathbf{X},\mathbf{C}))}} \\ Z_{X,i} &= \frac{f_{\mathbf{X},i}(\mathbf{X},\mathbf{C})}{\sqrt{Var(\sum_{i=1}^{n} f_{\mathbf{X},i}(\mathbf{X},\mathbf{C}))}} \\ Z_{C,i} &= \frac{f_{\mathbf{C},i}(\mathbf{C})}{\sqrt{Var(\sum_{i=1}^{n} f_{\mathbf{C},i}(\mathbf{C}))}}. \end{aligned}$$

and

$$\begin{aligned} Z'_{Y,i} &= \frac{f_{\mathbf{Y},i}(\mathbf{Y},\mathbf{X},\mathbf{C})\,|(\mathbf{X},\mathbf{C})}{\sqrt{Var(\sum_{i=1}^{n} f_{\mathbf{Y},i}(\mathbf{Y},\mathbf{X},\mathbf{C})\,|(\mathbf{X},\mathbf{C})\,)}} \\ Z'_{X,i} &= \frac{f_{\mathbf{X},i}(\mathbf{X},\mathbf{C})\,|\mathbf{C}}{\sqrt{Var(\sum_{i=1}^{n} f_{\mathbf{X},i}(\mathbf{X},\mathbf{C})\,|\mathbf{C}\,)}} \end{aligned}$$

We use the prime to denote conditional random variables: $Z'_{Y,i}$ conditions $f_{\mathbf{Y},i}(\mathbf{O})$ on $(\mathbf{X},\mathbf{C})$ and rescales it by the standard error of $f_{\mathbf{Y}}(\mathbf{O})|\,(\mathbf{X},\mathbf{C})$. Similarly, $Z'_{X,i}$

conditions $f_{\mathbf{X},i}(\mathbf{O})$ on $\mathbf{C}$ and rescales it by the standard error of $f_{\mathbf{X}}(\mathbf{O})|\mathbf{C}$. Let

$$\sigma^2_{nY}(\mathbf{x},\mathbf{c}) = Var\left(\sum_{i=1}^{n} f_{\mathbf{Y},i}(\mathbf{Y},\mathbf{x},\mathbf{c})\,|(\mathbf{X}=\mathbf{x},\mathbf{C}=\mathbf{c})\right)$$
$$\sigma^2_{nY} = E_{P_{\mathbf{X},\mathbf{C}}}\left[\sigma^2_{nY}(\mathbf{X},\mathbf{C})\right],$$

$$\sigma^2_{nX}(\mathbf{c}) = Var\left(\sum_{i=1}^{n} f_{\mathbf{X},i}(\mathbf{X},\mathbf{c})\,|\mathbf{C}=\mathbf{c}\right)$$
$$\sigma^2_{nX} = E_{P_{\mathbf{C}}}\left[\sigma^2_{nX}(\mathbf{C})\right],$$

and

$$\sigma^2_{nC} = Var\left(\sum_{i=1}^{n} f_{\mathbf{C},i}(\mathbf{C})\right).$$

Note that by the law of total variance $\sigma^2_{nX} = Var(\sum_{i=1}^n f_{\mathbf{X},i}(\mathbf{X},\mathbf{C}))$ and $\sigma^2_{nY} = Var\,(\sum_{i=1}^n f_{\mathbf{Y},i}(\mathbf{Y},\mathbf{X},\mathbf{C}))$.

Let Z'_{nY} denote $\sum_{i=1}^n Z'_{Y,i}$, Z'_{nX} denote $\sum_{i=1}^n Z'_{X,i}$, Z_{nY} denote $\sum_{i=1}^n Z_{Y,i}$, Z_{nX} denote $\sum_{i=1}^n Z_{X,i}$, and Z_{nC} denote $\sum_{i=1}^n Z_{C,i}$. We will establish convergence in distribution of each of the three terms separately. Because Z'_{nY} and Z'_{nX} converge to distributions that do not depend on their conditioning events, conditional convergence in distribution implies convergence of Z_{nY} and Z_{nX} to the same limiting distributions. Since $f_Y(\mathbf{O})$,$f_X(\mathbf{O})$, and $f_C(\mathbf{O})$ are orthogonal by construction, the variance of the limiting distribution of their sum is the sum of their marginal variances. If the three processes converge at the same rate the limiting variance will be the sum of the variances of the three processes. However, the three terms may converge at different rates, in which case the limiting distribution of $\hat{\psi}-\psi$ will be given by the limiting distribution of the term(s) with the slowest rate of convergence.

In order to show that Z'_{nX}, Z'_{nY}, and Z_{nC} all converge in distribution to a $N(0,1)$ random variable, we can use three separate applications of the central limit theorem given in Lemma 1, which is based on Stein's method.

Stein's method (Stein, 1972) quantifies the error in approximating a sample average with a normal distribution. (For an introduction to Stein's method see Ross, 2011.) Stein's method has been used to prove CLTs for dependent data with dependence structure given by *dependency neighborhoods* (Chen and Shao, 2004): the dependency neighborhood for observation i is a set of indices D_i such that observation i is independent of observation j, for any $j \notin D_i$. Conditionally on $\mathbf{C}$, $f_{X,i}$ and $f_{X,j}$ are independent for any nodes i and j such that $A_{ij}=0$ and there is no k with $A_{ik}=A_{jk}=1$, that is for any nodes that do not share a tie or have any mutual network contacts. The same is true for $f_{Y,i}$ and $f_{Y,j}$ conditional on $\mathbf{X}$ and $\mathbf{C}$ and for $f_{C,i}$ and $f_{C,j}$. Thus the three collections of random variables $Z'_{X,1},...,Z'_{X,n}$, $Z'_{Y,1},...,Z'_{Y,n}$, and $Z_{C,1},...,Z_{C,n}$ each has a dependency neighborhood structure with $D_i = i \cup \{j : A_{ij}=1\} \cup \{k : A_{jk}=1 \text{ for } j : A_{ij}=1\}$, that is the "friends" and "friends of friends" of node i. Define the indicators $R(i,j)$ for any $(i,j) \in \{1,\ldots,n\}^2$ to be an indicator of dependence between $Z_{X,i}$ and $Z_{X,j}$, $R(i,j) = 1$ iff $j \in D_i$ or, equivalently, if $i \in D_j$. For any $i \in \{1,\ldots,n\}$ the set $\{Z'_{X,j} : (R(i,j) = 1, j \in \{1,\ldots,n\})\}$ forms the dependency neighborhood of $Z'_{X,i}$ and the collection $\{Z'_{X,j} : (R(i,j) = 0, j \in \{1,\ldots,n\})\}$ is independent of $Z'_{X,i}$. The same logic applies to defining the dependency neighborhoods for $Z'_{Y,1}, ..., Z'_{Y,n}$ conditional on $\mathbf{X}$ and $\mathbf{C}$, and for $Z_{C,1}, ..., Z_{C,n}$ based on (unconditional) independence of each $f_{C,i}(\mathbf{O})$ and $f_{C,j}(\mathbf{O})$, as determined by the network structure and the distributional assumptions made for the baseline covariates $\mathbf{C}$.

Applied to Z'_{nX} , Stein's method provides the following upper bound

$$\begin{aligned} d(Z'_{nX}, Z) \quad &\leq \quad \sum_{i=1}^{n} \sum_{j,k \in D_i} E\left|Z'_{X,i} Z'_{X,j} Z'_{X,k}\right| \\ &\quad + \sqrt{\frac{2}{\pi}} \sqrt{Var\left(\sum_{i=1}^{n} \sum_{j \in D_i} Z'_{X,i} Z'_{X,j}\right)}, \end{aligned}$$

where $Z \sim N(0,1)$ and $d(\cdot,\cdot)$ is the Wasserstein distance metric (Vallender, 1974).

In order to show that Z'_{nX} converges in distribution to Z, we must show that the righthand side of the inequality converges to zero as n goes to infinity. We will first show that this convergence holds when $K_i = |F_i| = K_{max}(n)$ for all i, that is when all nodes have the same number of ties. We will then show that removing any tie from the network preserves an upper bound on the righthand side of the inequality. This completes our proof that for any network such that $K_i \leq K_{max}(n)$ for all i and $\frac{K^2_{max}(n)}{n}$ converges to zero as n goes to infinity, Z'_{nX} converges in distribution to a standard normal distribution. The same argument applied to Z_{nC} proves that it has a Normal limiting distributions as well.

Lemma 1 (Applying Stein's Method to the dependent sum)**.** *Consider a network of nodes given by adjacency matrix A. Let $U_1, ..., U_n$ be bounded mean-zero random variables with finite fourth moments and with dependency neighborhoods $D_i = i \cup \{j : A_{ij} = 1\} \cup \{k : A_{jk} = 1 \text{ for } j : A_{ij} = 1\}$, and let K_i be the degree of node i. If $K_i = K_{max}(n)$ for all i and $K_{max}(n)^2/n \to 0$, then $\frac{\sum U_i}{\sqrt{var(\sum U_i)}} \xrightarrow{D} N(0,1)$.*

Proof (Proof of Lemma 1). Let $U'_i = \frac{U_i}{\sqrt{var(\sum U_i)}}$. Application of Stein's method often involves defining the so-called "Stein coupling" (W, W', G) (Fang, 2011; Fang et al., 2015). Consider the following sum of dependent variables $W = \sum_{i=1}^{n} U'_i$. Define a discrete random variable I distributed uniformly over $\{1,\ldots,n\}$ and define another random variable $W' = (W - \sum_{j=1}^{n} R(I,j) U'_j)$. Finally, define $G = -nU'_I$ and note that (W, W', G) forms a Stein coupling (Fang, 2011; Fang et al., 2015). We also let $D = (W' - W) = -\sum_{j=1}^{N} R(I,j)U'_j$. This Stein coupling allows us then to derive the upper bound

$$d(W,Z) \leq \sum_{i=1}^{n} \sum_{j,k \in D_i} E\left|U'_i U'_j U'_k\right| + \sqrt{\frac{2}{\pi}} \sqrt{Var\left(\sum_{i=1}^{n} \sum_{j \in D_i} U'_i U'_j\right)}, \qquad (1)$$

as shown in Ross (2011). We will now show that, for any network structure,

$$\sum_{i=1}^{n} \sum_{j,k \in D_i} E\left|U'_i U'_j U'_k\right| + \sqrt{\frac{2}{\pi}} \sqrt{Var\left(\sum_{i=1}^{n} \sum_{j \in D_i} U'_i U'_j\right)} = O\left(\frac{\sum_{i,j,k} R(i,j)R(i,k)}{\left[\sum_{i,j} R(i,j)\right]^{3/2}}\right). \qquad (2)$$

The righthand side of the above equation is equal to $\sqrt{\frac{(K_{max}(n))^2}{n}}$ under the assumption of $K_{max}(n)$ ties for each node $i = \{1, \ldots, n\}$. By assumption, we also have that $\frac{K_{max}(n)}{\sqrt{n}}$ converges to zero as n goes to infinity, and therefore if we can show equation (2) we have proved that $\frac{\sum U_i}{\sqrt{var(\sum U_i)}} \xrightarrow{D} N(0,1)$.

Consider the term

$$\sum_{i=1}^{n} \sum_{j,k \in D_i} E\left|U'_i U'_j U'_k\right| = \frac{1}{var(\sum U_i)^{3/2}} \sum_{i=1}^{n} E\left\{\left|U_i \left(\sum_{j \in D_i} U_k\right)^2\right|\right\}.$$

By the assumption of bounded 4th moments, $var(\sum U_i)^{3/2} = O\left(\left[\sum_{i,j} R(i,j)\right]^{3/2}\right)$, that is, $var(\sum U_i)$ stabilizes to a constant when scaled by $\sum_{i,j} R(i,j)$. Using the fact that each $|U_i|$ is bounded we get

$$\begin{aligned} & \sum_{i=1}^{N} E\left\{\left|U_i \left(\sum_{j \in D_i} U_j\right)^2\right|\right\} \\ \leq \; & M \sum_{i=1}^{n} \left\{\sum_{j,k} R(i,j)R(i,k)\right\} \\ = \; & M \sum_{i,j,k} R(i,j)R(i,k), \end{aligned}$$

for some positive constant $M < \infty$. Combining the above expressions, we get

$$\sum_{i=1}^{n} \sum_{j,k \in D_i} E\left|U'_i U'_j U'_k\right| = O\left(\frac{\sum_{i,j,k} R(i,j)R(i,k)}{\left[\sum_{i,j} R(i,j)\right]^{3/2}}\right).$$

Now consider the second term:

$$\sqrt{Var\left(\sum_{i=1}^{n}\sum_{j\in D_i} U_i'U_j'\right)} = \frac{\sqrt{Var\left(\sum_{i=1}^{n}\sum_{j\in D_i} U_iU_j\right)}}{var(\sum U_i)^2}.$$

There are $\sum_{i,j} R(i,j)$ terms in $\sum_{i=1}^{n}\sum_{j\in D_i} U_iU_j$, and the number of terms U_kU_l with which U_iU_j has non-zero covariance is $|D_i \cup D_j| \leq \sum_k R(i,k) + \sum_k R(i,k)$, so $Var\left(\sum_{i=1}^{n}\sum_{j\in D_i} U_iU_j\right) \leq M\sum_{i,j} R(i,j)\sum_k R(i,k)$ for some finite M. Therefore $Var\left(\sum_{i=1}^{n}\sum_{j\in D_i} U_iU_j\right) = O\left(\sum_{i,j,k} R(i,j)R(i,k)\right)$. $Var(\sum U_i)^2 = O\left(\left[\sum_{i,j} R(i,j)\right]^2\right)$, so the second term is of smaller order than the first term. Therefore we have only to consider the first term and we have completed the proof.

Lemma 2 (Bound goes to zero when $K_i \leq K_{max}(n)$ for all i)**.** *Convergence to zero of the righthand side of Equation (1) is preserved under the removal of ties and holds as long as* $K_i \leq K_{max}(n)$ *for all* i *and* $\frac{K_{max}^2(n)}{n}$ *converges to zero as* n *goes to infinity.*

Proof (Proof of Lemma 2). Consider a sequence of networks with n going to infinity such that the righthand side of Equation (1) converges to 0, i.e.

$$\sum_{i=1}^{n}\sum_{j,k\in D_i} E\left|U_i'U_j'U_k'\right| + \sqrt{\frac{2}{\pi}}\sqrt{Var\left(\sum_{i=1}^{n}\sum_{j\in D_i} U_i'U_j'\right)} \rightarrow 0.$$

Because the second term is of the same or smaller order than the first, we only have to consider the first term. For this sequence of networks, define $A_n = \sum_{i=1}^{n}\sum_{j,k\in D_i} E\left|U_i'U_j'U_k'\right|$. Removing a single tie from the underlying network has the effect of rendering independent some pairs that were previously dependent; We now consider the effect of rendering a single dependent pair independent but otherwise leaving the distributions of the random variables the same. Suppose the pair rendered independent is (l,m). Define a new sequence of networks with n going to infinity to be identical to the previous sequence but with pair (l,m) independent, and let A_n' be the first term in the righthand side of Equation (1) for this new sequence. Then

$$A_n' = A_n - 2\sum_{k\in D_l\cup D_m} E\left|U_l'U_m'U_k'\right|$$

which is bounded above by A_n.

This completes the proof that Z_{nX}', Z_{nY}', and Z_{nC} have Normal limiting distributions.

Lemma 3 (Conditional CLT implies marginal CLT). *Z'_{nX} converges to Normal distribution after marginalizing over $\mathbf{C}$ (but conditioning on the network as captured by the adjacency matrix $\mathbf{A}$) and Z'_{nY} converges to Normal distribution after marginalizing over $(\mathbf{X}, \mathbf{C})$. That is, Z_{nX} and Z_{nY} both converge to Normal distributions.*

Proof (Proof of Lemma 3). For illustration consider $Z'_{nX} = \sum_{i=1}^{n} Z'_{2,i}$, where

$$Z'_{X,i} = \left(f_{\mathbf{X},i}(\mathbf{X}, \mathbf{C})\,|\mathbf{C}\right) / \sqrt{\sigma^2_{nX}(\mathbf{C})}$$

and note that the proof of the convergence of Z_{nY} is nearly identical. The conditional CLT results from Lemma 1 show that

$$P\left[Z'_{nX} \leq x\,|\mathbf{C} = \mathbf{c}\right] = P\left[\left(\sum_{i=1}^{N} \frac{f_{\mathbf{X},i}(\mathbf{X}, \mathbf{c})}{\sqrt{\sigma^2_{nX}(\mathbf{c})}} \leq x\right) |\mathbf{C} = \mathbf{c}\right]$$

converges to $\Phi(x)$ for each x and almost every $\mathbf{c}$, where Φ is the cumulative distribution function of the standard Normal random variable and $\mathbf{C}$ is a given sequence $(C_i : i = 1, \ldots, n)$. Let $P_{\mathbf{C}}$ denote the distribution of $\mathbf{C}$. Then

$$\begin{aligned} P(Z_{nX} \leq x) &\equiv P\left[\left(\sum_{i=1}^{N} \frac{f_{\mathbf{X},i}(\mathbf{X}, \mathbf{C})}{\sqrt{\sigma^2_{nX}}} \leq x\right)\right] \\ &= \int_{\mathbf{c}} P(Z'_{nX} \leq x|\mathbf{C} = \mathbf{c}) dP_{\mathbf{C}}(\mathbf{c}). \end{aligned}$$

For a given x, the dominated convergence theorem is now applied with $f_n(\mathbf{c}) = P(Z'_{nX} \leq x|\mathbf{C} = \mathbf{c})$ and the limit given by $f(\mathbf{c}) = \Phi(x) = m$, where m is some constant that doesn't depend on $\mathbf{c}$. From the previous conditional CLT result it follows that $f_n(\mathbf{c})$ converges to $f(\mathbf{c})$ pointwise for each $\mathbf{c}$. The next step is to find an integrable function g, such that $f_n < g$ and $\int g(\mathbf{c}) dP_{\mathbf{C}}(\mathbf{c}) < \infty$. The proof is then completed by choosing $g = 1$.

We have now shown that Z_{nY}, Z_{nX}, and Z_{nC} are asymptotically normally distributed. We now show that the sum of the three processes converges in distribution to a Normal random variable. Consider three cases: (1) the three processes have the same rate of marginal convergence in distribution, (2) one of the three processes converges faster than the other two, and (3) two of the processes converge faster than the third. In all three cases the rate of convergence for the sum will be the slowest of the three marginal rates. In case (3), the limiting distribution of the sum is determined entirely by the one process that converges with a slower rate than the other two: the other two processes will converge to constants (specifically to their expected values of 0) when standardized by the slower rate; Slutsky's theorem concludes the proof. We focus on case (1) below; case (2) follows immediately by applying the proof below to the two processes that converge at the same slower rate and applying Slutsky's to the third, faster converging process.

For convenience, in order to show that the sum of the three dependent processes also converges to Normal, define

$$C_n^* := \sigma_{nY}^2 + \sigma_{nX}^2 + \sigma_{nC}^2.$$

Note that C_n^* is related to C_n as follows: $C_n = O(n^2/C_n^*)$.

Lemma 4 (CLT for the sum of the three orthogonal processes)**.** *If all three processes have the same marginal rate of convergence, then*

$$\frac{1}{\sqrt{C_n^*}}\left(f_{\mathbf{Y}}(\mathbf{Y},\mathbf{X},\mathbf{C}) + f_{\mathbf{X}}(\mathbf{X},\mathbf{C}) + f_{\mathbf{C}}(\mathbf{C})\right) \to N(0,1).$$

Proof (Proof of Lemma 4). Without the loss of generality, we prove that $Z_{nX} + Z_{nC} \to N(0,2)$ and note that the general result for $(Z_{nY} + Z_{nX} + Z_{nC})$ follows by applying a similar set of arguments.

Consider the following random vector (Z_{nX}, Z_{nC}) taking values in $\mathbb{R}^2$. Let $F_n(x_1,x_2) \equiv P(Z_{nX} \le x_1, Z_{nC} \le x_2)$, where $(x_1,x_2) \in \mathbb{R}^2$. Let $\Phi^2(x_1,x_2) \equiv P(Z_X \le x_1)P(Z_C \le x_2)$, for $Z_X \sim N(0,1)$ and $Z_C \sim N(0,1)$, that is, $\Phi^2(x_1,x_2)$ defines the CDF of the bivariate standard normal distribution, for $(x_1,x_2) \in \mathbb{R}^2$. The goal is to show that $F_n(x_1,x_2) \to \Phi^2(x_1,x_2)$, for any $(x_1,x_2) \in \mathbb{R}^2$. The convergence in distribution for $Z_{nX} + Z_{nC}$ will follow by applying the Cramer and Wold Theorem (1936).

Note that

$$\begin{aligned} &P(Z_{nX} \le x_1, Z_{nC} \le x_2) \\ =&P(Z_{nX} \le x_1 \,|Z_{nC} \le x_2)P(Z_{nC} \le x_2). \end{aligned}$$

First, from the previous application of Stein's method, we have that

$$P(Z_{nC} \le x_2) \to \Phi(x_2),$$

where $\Phi(x_2) \equiv P(Z_C \le x_2)$, $Z_C \sim N(0,1)$ and $x_2 \in \mathbb{R}^2$. Also note that

$$\begin{aligned} &P(Z_{nX} \le x_1 \,|Z_{nC} \le x_2\,) \\ = &\sum_{\mathbf{c}\in\mathcal{C}} P(Z_{nX} \le x_1 \,|\mathbf{C} = \mathbf{c}\,)P(\mathbf{C} = \mathbf{c}\,|Z_{nC} \le x_2\,), \end{aligned}$$

where $\mathcal{C}$ denotes the support of $\mathbf{C}$, $Z_{nX} = \frac{1}{\sqrt{C_n^*}} f_{\mathbf{X}}(\mathbf{X},\mathbf{C})$, $Z_{nC} = \frac{1}{\sqrt{C_n^*}} f_{\mathbf{C}}(\mathbf{C})$ and

$$P(\mathbf{C} = \mathbf{c}\,|Z_{nC} \le x_2\,) = \frac{P(\mathbf{C} = \mathbf{c})I\big(\big(1/\sqrt{C_n^*}\big)\, f_{\mathbf{C}}(\mathbf{c}) \le x_2\big)}{P\big(\big(1/\sqrt{C_n^*}\big)\, f_{\mathbf{C}}(\mathbf{c}) \le x_2\big)}.$$

By another application of Stein's method, it was shown that

$$P(Z_{nX} \le x_1 \,|\mathbf{C} = \mathbf{c}\,) \to \Phi(x_2),$$

for any realization of $\mathbf{c} \in \mathcal{C}$. That is, we've shown that the limiting distribution of Z_{nX} conditional on $\mathbf{C} = \mathbf{c}$, does not itself depend on the conditioning event $\mathbf{C} = \mathbf{c}$. Applying Lemma 3, we finally conclude that $F_n(x_1,x_2) \to \Phi^2(x_1,x_2)$, for any $(x_1,x_2) \in \mathbb{R}^2$ and the result follows.

D. Variance estimation

The estimate of the variance of the TMLE $\hat{\psi}$ can be obtained from the sum, scaled by $1/n^2$, of the three plug-in estimators of

$$\begin{aligned} \sigma^2_{nY} &= \sum_{i,j} E(f_{\mathbf{Y},i}(\mathbf{O})f_{\mathbf{Y},j}(\mathbf{O})) \\ \sigma^2_{nX} &= \sum_{i,j} E(f_{\mathbf{X},i}(\mathbf{O})f_{\mathbf{X},j}(\mathbf{O})) \\ \sigma^2_{nC} &= \sum_{i,j} E(f_{\mathbf{C},i}(\mathbf{O})f_{\mathbf{C},j}(\mathbf{O})). \end{aligned}$$

Alternatively, one can estimate the variance from a single plug-in estimator

$$\frac{1}{n^2}\sum_{i,j} E(f_i(\mathbf{O})f_j(\mathbf{O})).$$

Note that contribution to these variances of any pair i, j not in each others dependency neighborhoods will be 0. Therefore, it is acceptable to sum only over pairs i, j sharing a tie or a mutual contact in the underlying network. Finally, note that we do not need to know the true rate of convergence $\sqrt{C_n}$ to obtain a valid estimate of the C.I. for ψ; this rate is captured by the number of non-zero terms in the variance sums.

Appendix III : Simulations

All simulation and estimation was carried out in **R** language (R Core Team, 2015) with packages `simcausal` (Sofrygin et al., 2015) and `tmlenet` (Sofrygin and van der Laan, 2015). The full **R** code for this simulation study is available in a separate github repository (`github.com/osofr/Ogburn_etal_simulations`). Sofrygin and van der Laan (2015); Sofrygin et al. (2017, 2018) provide additional details on implementation, computation, and simulations for asymptotic regimes with a bounded number of ties per node and with no latent variable dependence.

The simulations were repeated for community sizes of $n = 500$, $n = 1,000$ and $n = 10,000$. The estimation was repeated by sampling $1,000$ such datasets, conditional on the same network (sampled only once for each sample size). For the simulations with dependence due to direct transmission, the baseline covariates were independently and identically distributed. The probability of success for each Y_i was a logit-linear function of i's exposure X_i (indicator of receiving the economic incentive), the baseline covariates C_i and the three summary measures of i's friends exposures and baseline covariates. In particular, we also assumed that the probability of maintaining gym membership increased on a logit-linear scale as a function of the following network summaries: the total number of i's friends who were exposed ($\sum_{j:A_{ij}=1} X_j$), the total number of i's friends who were physically active at baseline ($\sum_{j:A_{ij}=1} PA_j$) and the product of the two summaries ($\sum_{j:A_{ij}=1} X_j \times \sum_{j:A_{ij}=1} PA_j$). The summary measures

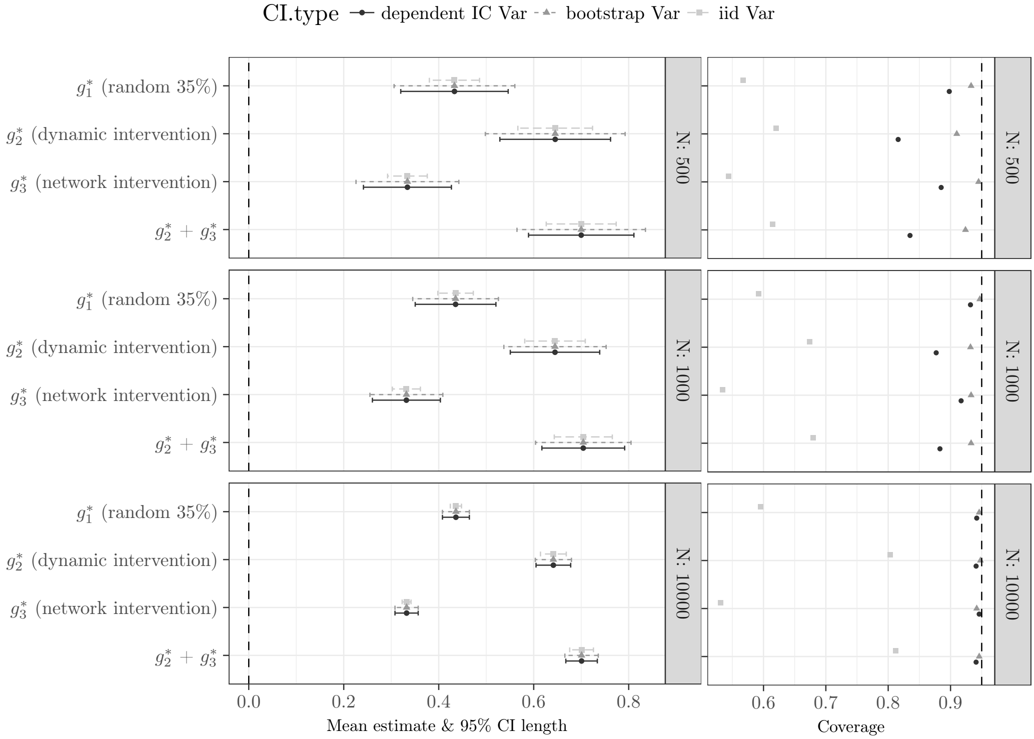

Fig. 1. Mean 95% CI length (left panel) and coverage (right panel) for the TMLE in small world network with dependence due to direct transmission, by sample size, intervention and CI type. Results are shown for the estimates of the average expected outcome under four hypothetical interventions (g_1^*, g_2^*, g_3^* and $g_2^* + g_3^*$).

and the outcome regression model were correctly specified, but we do not know (and therefore did not a priori correctly specify a model for) the true density of h. The economic incentive to attend local gym had a small direct effect on each individual who was not physically active at baseline and no direct effect on those who were already physically active. However, physically active individuals were more likely to maintain gym membership over the follow-up period if they had at least one physically active friend at baseline. We repeated these simulations with the addition of latent variable dependence, which we introduced by generating unobserved latent variables for each node which affected the node's own outcome as well as the outcomes of its friends.

In addition to the preferential attachment network model (results in main text), we also simulated social networks from the small world network model (Watts and Strogatz, 1998) with a rewiring probability of 0.1. The results of these simulations are in Figures 1 and 2.

We examined the empirical distribution of the transformed TMLEs, comparing their histogram estimates to the predicted normal limiting distribution, with the results shown in Figure 3, where the histogram plots are displayed by

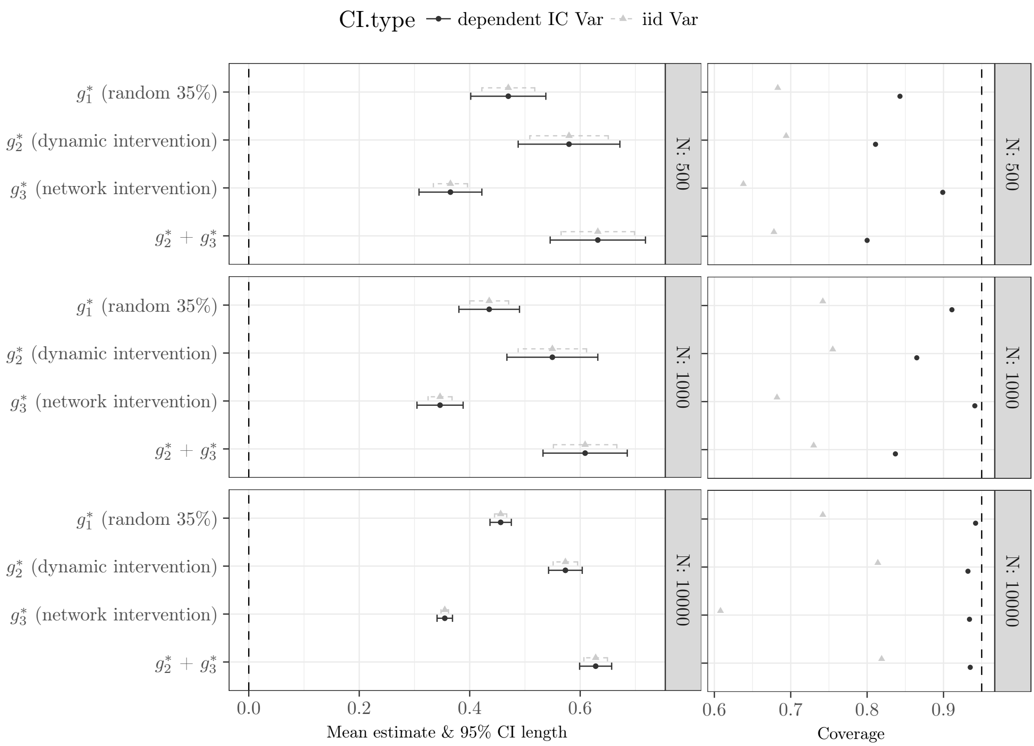

Fig. 2. Mean 95% CI length (left panel) and coverage (right panel) for the TMLE in small world network with latent variable dependence, by sample size, intervention and CI type. Results are shown for the estimates of the average expected outcome under four hypothetical interventions (g_1^*, g_2^*, g_3^* and $g_2^* + g_3^*$).

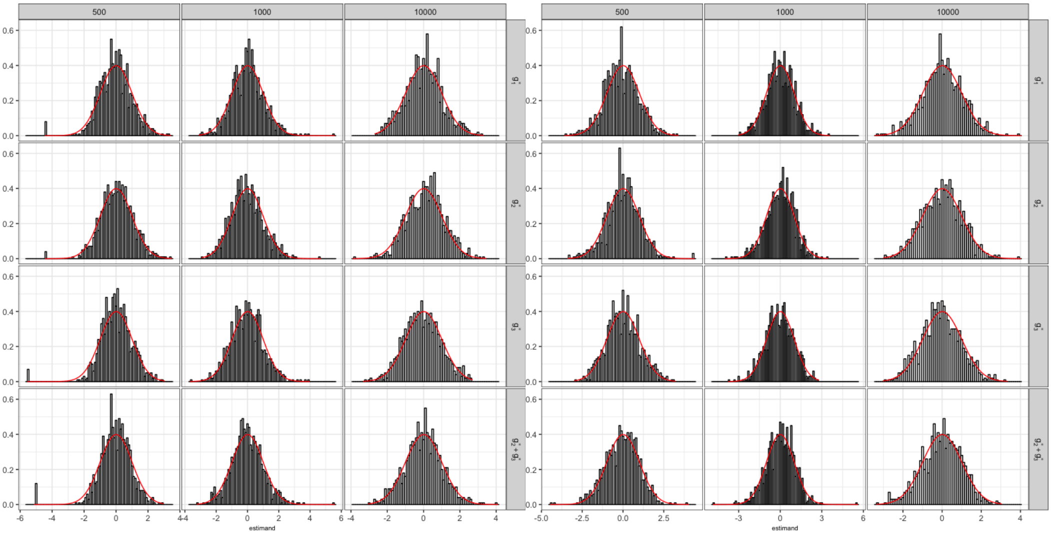

Fig. 3. Comparing re-scaled empirical TMLE distributions (black) to their theoretical normal limit (red) with varying sample size (x-axis) and intervention type (y-axis). TMLEs were centered at the truth and then re-scaled by true SD. Results shown for the preferential attachment network (left) and the small world network (right).

sample size (horizontal axis) and the intervention type (vertical axis). The estimates were first centered at the corresponding true parameter values and then re-scaled by their corresponding true standard deviation (SD). We note that our results indicate that the estimators converge to their normal theoretical limiting distribution, even in networks with power law node degree distribution, such as the preferential attachment network model, as well as in the densely connected networks obtained under the small world network model. The results shown in Figure 3 were generated from simulations with dependence due to direct transmission; simulations with latent variable dependence (not shown) evinced similar approximate normality.

Appendix IV : Comparison of Estimands

Table 1 summarizes the relationships among the two sets of assumptions (with and without latent variable dependence) and the two classes of estimands (marginal over $\mathbf{C}$ and conditional on $\mathbf{C}$) according to their properties and according to the limitations of our proposed methods.

Table 1. Properties of marginal estimands and of estimands conditional on $\mathbf{C}$

Properties that we have demonstrated for the two classes of estimands	Estimand class	
	Marginal	Conditional
nonparametrically identified with or without latent variable (LV) dependence	yes	yes
estimator is CAN with or without LV dependence	yes	yes
efficient estimator is available with LV dependence	no	no
efficient estimator is available without LV dependence	yes	yes
consistent and tractable variance estimation with LV dependence	no	yes
consistent and tractable variance estimation without LV dependence	yes	yes

Appendix V : Glossary of Notation

$\mathbf{A}$ with entries $A_{ij} \equiv I\{\text{subjects } i \text{ and } j \text{ share a tie}\}$ is the adjacency matrix for the network.

$K_i = \sum_{j=1}^{n} A_{ij}$, that is, K_i is the degree of node i, or the number of individuals sharing a tie with individual i.

$F_i = j : A_{ij} = 1$ is the set of nodes with with node i shares a tie (node i's "friends").

C_i is covariates

X_i is exposure

Y_i is outcome

s_X is a summary function of $\mathbf{C}$ upon which X depends.

s_Y is a summary function of $\mathbf{C}, \mathbf{X}$ upon which Y depends.

$W_i = s_{X,i}\left(\{C_j : A_{ij} = 1\}\right)$

$V_i = s_{Y,i}\left(\{C_j : A_{ij} = 1\}, \{X_j : A_{ij} = 1\}\right)$

$O_i = (C_i, W_i, X_i, V_i, Y_i)$

x_i^* represents a user-specified intervention value of X_i.

$Y_i(\mathbf{x}^*)$, shorthand Y_i^*, denotes the potential or counterfactual outcome of individual i in a hypothetical world in which $P(\mathbf{X} = \mathbf{x}^*) = 1$.

$V_i(\mathbf{x}^*)$, shorthand V_i^*, is equal to $s_{Y,i}(\mathbf{C}, \mathbf{x}^*)$ and is a counterfactual random variable in a hypothetical world in which $P(\mathbf{X} = \mathbf{x}^*) = 1$.

$\bar{Y}_n^* = \frac{1}{n} \sum_{i=1}^{n} Y_i^*$.

$p_C(\mathbf{c}) = P\left(\mathbf{C} = \mathbf{c}\right)$

$g(\mathbf{x}|\mathbf{w}) = P\left(\mathbf{X} = \mathbf{x} \mid \mathbf{W} = \mathbf{w}\right)$

$g_i(x|w) = P(X_i = x|W_i = w)$

$p_Y(\mathbf{y}|\mathbf{v}) = P(\mathbf{Y} = \mathbf{y}|\mathbf{V} = \mathbf{v})$

$p_{Y,i}(y|v) = P(Y_i = y|V_i = v)$

$h_i(v) = P(V_i = v)$

$h_{i,x^*}(v) = P(V_i^* = v)$

$m(v) = \sum_y y\, p_Y(y|v)$ is the conditional expectation of Y given $V = v$.

$\bar{h}(v_i) = \frac{1}{n}\sum_{j=1}^n h_j(v_i)$

$\bar{h}_{x^*}(v_i) = \frac{1}{n}\sum_{j=1}^n h_{j,x^*}(v_i)$

$v_i = s_{Y,i}(\mathbf{c}, \mathbf{x})$

$V_i^* = s_{Y,i}(\mathbf{C}, \mathbf{x}^*)$

$D(\mathbf{o})$ is the efficient influence function under assumptions (A1), (A4) and (A5).

$D'(\mathbf{o})$ is the efficient influence function under assumptions (A1) and (A4).

$D_c(\mathbf{o})$ is an influence function conditional on $\mathbf{C} = \mathbf{c}$.

$K_{max,n} = max_i\{K_i\}$

$\sqrt{C_n}$ is the rate of convergence in Theorem 1.

References

Benkeser, D., C. Ju, S. Lendle, and M. van der Laan (2018). Online cross-validation-based ensemble learning. *Statistics in medicine 37*(2), 249–260.

Chen, L. H. and Q.-M. Shao (2004). Normal approximation under local dependence. *The Annals of Probability 32*(3), 1985–2028.

Fang, X. (2011). *Multivariate, combinatorial and discretized normal approximations by Stein's method.* Ph. D. thesis.

Fang, X., A. Röllin, et al. (2015). Rates of convergence for multivariate normal approximation with applications to dense graphs and doubly indexed permutation statistics. *Bernoulli 21*(4), 2157–2189.

Munoz, I. D. and M. J. van der Laan (2011). Super learner based conditional density estimation with application to marginal structural models. *The International Journal of Biostatistics 7*(1), 1–20.

R Core Team (2015). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Ross, N. F. (2011). Fundamentals of stein's method. *Probability Surveys 8*, 210–293.

Sofrygin, O., R. Neugebauer, and M. J. van der Laan (2017). Conducting simulations in causal inference with networks-based structural equation models. *arXiv preprint arXiv:1705.10376*.

Sofrygin, O., E. L. Ogburn, and M. J. van der Laan (2018). Single time point interventions in network-dependent data. In *Targeted Learning in Data Science*, pp. 373–396. Springer.

Sofrygin, O. and M. J. van der Laan (2015). Semi-Parametric Estimation and Inference for the Mean Outcome of the Single Time-Point Intervention in a Causally Connected Population. *U.C. Berkeley Division of Biostatistics Working Paper Series* (Working Paper 344).

Sofrygin, O. and M. J. van der Laan (2015). *tmlenet: Targeted Maximum Likelihood Estimation for Network Data.* R package version 0.1.0.

Sofrygin, O., M. J. van der Laan, and R. Neugebauer (2015). *simcausal: Simulating Longitudinal Data with Causal Inference Applications.* R package version 0.5.0.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Stat. Prob.*, pp. 583–602.

Vallender, S. (1974). Calculation of the wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications 18*(4), 784–786.

van der Laan, M. J. (2014). Causal inference for a population of causally connected units. *Journal of Causal Inference 0*(0), 2193–3677.

Van der Laan, M. J. and S. Rose (2011). *Targeted learning: causal inference for observational and experimental data.* Springer Science & Business Media.

van der Laan Mark, J., C. Polley Eric, et al. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology 6*(1), 1–23.

Van Der Vaart, A. W. and J. A. Wellner (1996). Weak convergence. In *Weak Convergence and Empirical Processes*, pp. 16–28. Springer.

Watts, D. J. and S. H. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature 393*(6684), 440–442.