Using Embeddings to Estimate Peer Influence on Social Networks

Irina Cristali
Department of Statistics
The University of Chicago
icristali@uchicago.edu.edu

Victor Veitch
Department of Statistics
The University of Chicago
victorveitch@gmail.com

Abstract

We address the problem of using observational data to estimate peer contagion effects, the influence of treatments applied to individuals in a network on the outcomes of their neighbours. A main challenge to such estimation is that *homophily*—the tendency of connected units to share similar latent traits—acts as an unobserved confounder for contagion effects. Informally, it's hard to tell whether your friends have similar outcomes because they were influenced by your treatment, or whether it's due to some common trait that caused you to be friends in the first place. Because these common causes are not usually directly observed, they cannot be simply adjusted for. We describe an approach to perform the required adjustment using node embeddings learned from the network itself. The main aim is to perform this adjustment non-parametrically, without functional form assumptions on either the process that generated the network or the treatment assignment and outcome processes. The key questions we address are: How should the causal effect be formalized? And, when can embedding methods yield causal identification?

1 Introduction

Consider the following example.

Example. We want to infer the effect that social pressure has on vaccination. Suppose we observe networked data from a population where each unit i is a person in an interconnected social network, and for each unit we know whether they were vaccinated at the beginning of the study period, T_i, and whether they were vaccinated at the end of the study period, Y_i. We are interested in estimating the effects of the treatment T_i of person i on the outcome Y_j of person j. In addition to their vaccination status, each unit has attributes C_i that act as (proxies for) causes of both the particular network ties they form, and their vaccination behavior. For instance, C_i may include age, race, education status, income level, political affiliation, and so forth.

The core challenge here is that we want to estimate a causal effect (e.g., what would happen if we intervened by vaccinating popular people?), but the variables C_i, C_j act as confounders between the treatment T_j and outcome Y_i. The reason is that when we define a contagion effect from j to i we must condition on the presence of an edge between (i, j). The edge is causally influenced by both C_i and C_j, so the conditioning creates a dependency between these variables (it acts as a collider). For example, if we learn that Alice and Bob are friends then we can infer that they likely share a political affiliation. Accordingly, the association between T_j and Y_i may be either due to the causal influence of T_j on Y_i (Bob got vaccinated because Alice did), or due to the common cause (Alice and Bob got vaccinated because they are democrats, democrats get vaccinated at higher rates, and it's only due to chance that Alice got vaccinated before Bob). In

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

general, homophily—the tendency of similar people to cluster in a network—is confounded with contagion. [ST11]

Now, if we observed the attributes C, we could formalize the contagion effect using standard causal tools and then identify the effect from observational data by adjusting for C [EB17; EKB16; BDF20]. However, such detailed knowledge is often unavailable. Intuitively, we could instead make use of the following observation. The pattern of network ties itself carries information about each C_i, so we can estimate $\{\hat{C}_i\}_i$ from the network. Then, we can adjust for the estimated $\hat{C}_i$ in some suitable causal estimation procedure. The aim of this paper is to clarify this type of procedure.

? provide an estimation strategy of this kind. They first assume that the network is generated by either a stochastic block model [HLL83] or a latent space model [RFR16; HRH02]. The attributes C correspond to the latent community identities or latent space positions. They then further assume that the outcome of each unit is defined by a particular linear structural equation, which includes a term for both the average treatment of that node's neighbours, and a term for the effect of the attributes C_i. Their procedure is to estimate C using the assumed network model, and then use the estimated $\hat{C}$ in a linear regression to determine the coefficient of the average-neighbour-treatment term.

The limitation of this approach is that it relies critically on the assumed parametric form of both the network model and the outcome model. Indeed, even the target of estimation is defined as a parameter of this assumed model. Our aim here is to develop a non-parametric version of this procedure. This consists of three steps.

1. Formalize the target causal effect non-parametrically. The challenge here is that the estimand must depend on the network we are working with (because contagion requires knowing who is friends with whom) and the network must itself be modeled as a random variable (to accommodate homophily).
2. Establish sufficient conditions for estimated attributes to yield causal identification. The idea is that it is not necessary to exactly reconstruct C, but only extract the minimum information that will identify the causal effect—this turns out to be (plausibly) a much easier task.
3. A concrete method for contagion estimation using node embedding techniques.

This is preliminary work and we do not yet report experimental results from the method.

2 Setup

Consider a network G_n of n individuals, where connections between people are encoded through undirected edges between nodes. We define the degree of a node as the number of connections it has. The neighbors of node i are the nodes with which i has ties. Each such link is captured by the network adjacency matrix A, where $A_{ij} = \mathbb{1}_{\{\text{i and j share a tie}\}}$. To simplify notation, we take $A_{ii} = 1$ for all i—the treatment of unit i will influence its own outcome Y_i. We also consider the following vector of variables associated with each node:

$$O_i = (Y_i, C_i, T_i),$$

where Y_i is the observed outcome, T_i is the treatment, and C_i are (unobserved) attributes that may causally influence Y, T and A. Each unit is sampled from some known distribution, $O_i \overset{\text{iid}}{\sim} P$. Then, the network A is sampled from some unknown process that depends on $\{C_i\}_i$.

More precisely, we consider the following structural equation model.

$$C_i \leftarrow f_C[\varepsilon_{C_i}] \tag{2.1}$$

$$A_{ij} \leftarrow f_A[\{C_i\}_i, \varepsilon_{ij}] \tag{2.2}$$

$$T_i \leftarrow f_T[C_i, \varepsilon_{T_i}] \tag{2.3}$$

$$Y_i \leftarrow f_Y[S_Y(\{T_j : A_{ij} = 1\}), C_i, \varepsilon_{Y_i}]. \tag{2.4}$$

The variables ε are exogenous noise, which we take to be identically distributed and independent of the network and of each other. The function S_Y summarizes the neighbours' treatment—e.g.,

S_Y might be the average function. In words: the treatment assignment for each node i depends only on its attributes C_i, the outcome assignment depends on the node attributes C_i and the treatments of all its neighbours, and the network structure depends on all C. The structural functions f are fixed but unknown.

This formalization is adapted from [Ogb+17]. It is modified to treat the network as a random variable, and to simplify the dependence of the treatment and outcome on covariates (they allow for dependency on covariates of neighbours as well).

3 Formalizing the causal estimand

With the assumed structural equation model in hand, we turn to formalizing the target causal effect. Consider setting $T_i \leftarrow t_i^*$ for each unit i. We define the causal estimand to be

$$\psi_n := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i | \text{do}(T = t^*), \{C_i\}_i, G_n]. \tag{3.1}$$

That is, the estimand is the average outcome we would have seen had the treatment assignment been set to $T \leftarrow t^*$. Critically, we condition on both the node attributes $\{C_i\}$ and the network G_n. The interpretation of the causal effect is then: the average outcome under the hypothetical treatment, applied to the *same set of people* connected by the *same link structure*. This is certainly not the only possible way of formalizing our goal. For instance, we might have conditioned on the graph alone (marginalizing over attributes of people not fixed by the link structure), or we might have marginalized out both the graph and the attributes (targeting a more generic notion of contagion effect). However, this particular estimand has the advantage of being readily interpretable without parametric assumptions, and, as we will see, plausibly identifiable from observational data.

There is one apparent significant drawback of this formalization: the estimand is fundamentally tied to the particular sample available in our study. When we ask questions such as, "what is the social contagion effect of vaccination?" it's usually with the goal of guiding policy. If we have a precise estimate of the effect among a particular sample of people in, e.g., Wyoming, it's unclear how that informs our policy about a new network of people in Illinois. In ordinary practice, the estimand (e.g., average treatment effect) is a parameter of some population, and judgements about transportability of effects reduce to judgements about whether populations are similar. Ideally, we'd like conditions under which the estimand approximates some population parameter, so that the requisite judgements reduce to only the ordinary required ones (e.g., "Are social dynamics in Illinois similar enough to social dynamics in Wyoming that results can be expected to translate?").

Happily, such a result is possible. In Theorem 1 we justify the validity of the causal estimand ψ_n introduced in Equation 3.1, by showing it converges to a fixed quantity as the size of the network G_n increases. The result heavily relies upon Stein's method ([Ros11]) and we include details of the background and proof in the Appendix.

Before stating the theorem, we first note that the causal effect of peer influence

$$\psi_n := \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[Y_i | \text{do}(T = t^*), \{C_i\}_i, G_n],$$

represents a random variable since we are conditioning on the unknown confounders C. This observation is a central point of the following theorem

Theorem 1 (Validity of causal estimand ψ_n)**.** *Consider an observed social network G_n, and let D_n be the maximum degree of the nodes in G_n. If $D_n = O(n^{1/4})$, and if $\mathbb{E}|Y_i|^4 < \infty$ for all nodes i, then $\psi_n \to \psi$ in probability, for some real number ψ.*

Proof. See Appendix. □

This theorem thus establishes that, provided the maximum network node degree grows no faster than $O(n^{1/4})$, the sequence of causal estimands $\{\psi_n\}_n$ converges to a fixed quantity which does not depend on the particular network G_n. This confirms the validity of using the same measure ψ_n to estimate social contagion in various different data samples.

4 Causal inference of peer effects using node embedding methods

Having formalized and motivated the target causal effect of interest ψ_n, we now turn to sufficient conditions for causally identifying ψ_n. Recall that ψ_n depends on the graph structure and the unobserved confounders C_i. Exactly reconstructing the latent features C_i from the network may be prohibitively difficult. However, we do not need to have access to the full covariate information, but rather only to a proxy that suffices for causal identification. We thus establish sufficient conditions for the inferred proxy to enable identification. The idea is to determine a condition that is plausibly satisfied by already established node embedding methods. We state our main theorem below.

Theorem 2 (Causal identification)**.** *Let G be the network and A be its adjacency matrix. Let $V_i := S_Y(\{T_j : A_{ij} = 1\})$ be the aggregated treatment at node i, and v^* its value under the hypothetical treatment intervention $T = t^*$. Suppose that for each node i we have an embedding $\lambda_i \in \mathbb{R}^k$ that satisfies the following conditions.*

i $A_{ij} \perp\!\!\!\perp Y_i | (\lambda_i, T_j)$*;*

ii $P(V_i = v^* \mid \lambda_i(C_i)) > 0$ *for all* v^**;*

iii λ_i *is* C_i*-measurable.*

Letting $m_{G_n}(t^*, \lambda_i) := \mathbb{E}[Y_i | V_i = S_Y(\{t_j^* : A_{ij} = 1\}), \lambda_i, G_n]$*, we have that*

$$\psi_n = \frac{1}{n} \sum_{i=1}^{n} m_{G_n}(t^*, \lambda_i). \tag{4.1}$$

Proof. Consider Fig. 1 illustrating the issues that can arise due to the unobserved confounders C. To estimate ψ_n, it is unavoidable to condition on the network structure, and implicitly on the edges A_{ij}. Doing so, however, introduces a collider bias which prevents accurate estimation. From Fig. 1 we can see that C_i blocks the backdoor path between T_j and Y_i that's opened by conditioning on A_{ij}. Accordingly,

$$\begin{aligned}\mathbb{E}[Y_i \mid do(T = t^*), C_i, G_n] &= \mathbb{E}[Y_i \mid T = t^*, C_i, G_n] \\ &= \mathbb{E}[Y_i \mid T = t^*, \lambda_i(C_i), G_n] \text{ (due to conditions (i) and (ii)).}\end{aligned}$$

Accordingly,

$$\psi_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[Y_i | do(T = t^*), C_i, G_n]$$

can be identified by computing $m(t^*, \lambda_i)$ for each node i of the graph and summing these quantities up:

$$\psi_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[Y_i | do(T = t^*), C_i, G_n] = \frac{1}{n} \sum_{i=1}^{n} m_{G_n}(t^*, \lambda_i). \tag{4.2}$$

□

In Theorem 2, we expect condition (i) to hold true because embedding methods are tools which decouple the properties of the unit and the network structure, and have shown good empirical performance at explaining the local network structure ([HYL17]). Condition (ii) represents the standard positivity assumption required for performing causal inference, while assumption (iii) reflects the requirement that the node embeddings should be functions of the latent node attributes C_i.

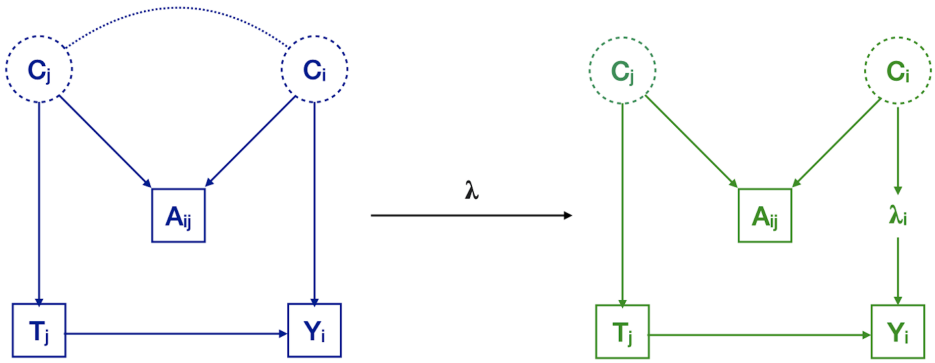

Figure 1: Identification of causal effects using embeddings in a semi-supervised learning problem. This diagram shows a potential backdoor path passing through the unobserved confounders C_j and C_i, while conditioning on the network structure (represented via the adjacency matrix variables A_{ij}) when attempting to estimate the effect of the treatment T_j on the response Y_i. This problematic backdoor path is blocked if A_{ij} is conditionally independent from the response Y_i given the embedding λ_i and the treatment T_i, allowing for successful causal adjustment.

5 Estimation of causal peer influence effects

Having established how to identify causal effects using black-box network embeddings, we now discuss how to estimate these effects from data. Recall that, by theorem 2, the causal parameter of interest is identified as:

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} m_{G_n}(t^*, \lambda_i), \tag{5.1}$$

where $m_{G_n}(t^*, \lambda_i) = \mathbb{E}[Y_i \mid T = t^*, \lambda_i(C_i), G_n]$. The goal is then to use the observed data and the network link structure to produce estimates $\hat{m}_{G_n}(t^*, \hat{\lambda}_i)$ of $m_{G_n}(t^*, \lambda_i(C_i))$, where $\hat{\lambda}_i$ is the output of an embedding method that we hope converges to some λ_i satisfying the conditions of theorem 2. We can then use the plug-in estimator:

$$\hat{\psi}_n = \frac{1}{n}\sum_{i=1}^{n} \hat{m}_{G_n}(t^*, \hat{\lambda}_i). \tag{5.2}$$

We use embedding based semi-supervised prediction models to learn $\hat{\lambda}$ and $\hat{m}_{G_n}(t^*, \hat{\lambda}_i)$. For concreteness, we describe a particular approach based on [Vei+19; VWB19]. The estimation procedure follows three main steps:

Step 1. We train a model using relational empirical risk minimization [Vei+19] to learn embeddings $\hat{\lambda}_i$ and $\hat{m}$. Let Sample(G_n, k) be a sampling algorithm that returns a random subgraph of size k from G_n (e.g., the subgraph induced by a random walk of length k). Randomly divide the nodes into a subset I_0 with labels that will not be used during training, and the remaining nodes $I \backslash I_0$. For each vertex i, define $v_i = S_Y(\{t_j : A_{ij} = 1\})$. Then, define the loss function

$$L(G_k, \lambda, \gamma) = \sum_{i \in I \backslash I_0} (y_i - m(v_i, \lambda_i; \gamma))^2 + \sum_{i,j \in I \times I} \text{CrossEntropy}(A_{ij}, \sigma(\lambda_i^T \lambda_j)),$$

where σ is the sigmoid function. The second term is a network reconstruction term that extracts node-level convariate information from the network by requiring the embeddings to be predictive of the edge structure. The first term learns a predictor m, parameterized by γ, that predicts y_i from the aggregated treatment v_i and the embedding λ_i. For instance, m could be a linear model or a neural network.

Then, we train the model by fitting

$$\hat{\lambda}, \hat{\gamma} = \underset{\lambda, \gamma}{\text{argmin}}\, \mathbb{E}_{G_k = Sample(G_n, k)}[L(G_k, \lambda, \gamma)]. \tag{5.3}$$

Step 2. We define $\hat{m}_{G_n}(t^*, \lambda_i) = m(S_Y(\{t_j^* : A_{ij} = 1\}), \lambda_i; \hat{\gamma})$. Then, for each unit $i \in I_0$ we compute $\hat{m}_{G_n}(t^*, \hat{\lambda}_i)$.

Step 3. Finally, we compute the estimate as:

$$\hat{\psi}_n(I_0) = \frac{1}{|I_0|} \sum_{i \in I_0} \hat{m}_{G_n}(t^*, \hat{\lambda}_i). \tag{5.4}$$

6 Related work

This section offers a brief overview of some directions of work related to our paper. Firstly, there have been a number of recent studies [Laa14; Ogb+17; TFS17; Leu21] which also tackle the issue of causally estimating social contagion in observational studies, however they differ from this work by assuming no unobserved sources of network confounding, such as homophily.

Other works which did account for latent confounding [SM16; ST11] used parametric models to represent the network generating processes, whereas this paper aims to examine more general cases of networks by using a nonparametric model.

Another important line of work focusing on estimating causal social influence in networks relies on randomized experiments where treatments are randomly assigned to the network units, in order to study how the responses of the neighbors are affected under various scenarios [EKB16; TK13; Van10; Rub80]. However, since randomized experiments may not always be feasible, it is convenient to also benefit from causal inference techniques for purely observational data.

Finally, some of the few recent studies which also address nonparametric peer influence estimation from observational data propose different approaches which complement the node-embedding based method proposed in this paper. [EB17] involves using high-dimensional adjustments for covariates, whereas this work assumes that covariates are unobserved in addition to the network. [ET21], on the other hand, does tackle the situation of unobserved confounding, yet it describes using negative control outcome and exposure variables to estimate contagion effects, which represents a different technique from the methods described in this paper.

7 Conclusions, limitations, and future directions

This paper tackled the problem of estimating social contagion from observational data in the presence of unobserved confounding. The main contributions of this work at its current stage are

1. formalizing and justifying a causal estimand capturing contagion in the presence of homophily;
2. giving sufficient conditions for network embeddings to enable causal identification.

This paper justified the validity of the proposed peer influence measure and embedding-based prediction approach from a theoretical standpoint (theorems 1 and 2) and described a practical method of obtaining an estimate of peer contagion (Section 5).

This work is however still in its preliminary stages. A future step involves illustrating the performance of our estimation technique on a real social network dataset. Furthermore, one current methodological limitation is the lack of an asymptotic normality result concerning the peer influence estimator $\hat{\psi}_n$ defined in Section 5. We seek to obtain such asymptotics in the future which will enable us to conduct statistical inference in finite samples (and not just asymptotically), by using normal approximations and constructing confidence intervals around the estimated values $\hat{\psi}_n$.

References

[BDF20] Y. Bramoullé, H. Djebbari, and B. Fortin. "Peer effects in networks: a survey". In: *Annual Review of Economics* (2020).

[EB17] D. Eckles and E. Bakshy. "Bias and high-dimensional adjustment in observational studies of peer effects". In: *Journal of the American Statistical Association* (2017).

[EKB16] D. Eckles, R. Kizilcec, and E. Bakshy. "Estimating peer effects in networks with peer encouragement designs". In: *Proceedings of the National Academy of Sciences* (2016).

[ET21] N. Egami and E. Tchetgen. "Identification and estimation of causal peer effects using double negative controls for unmeasured network confounding". In: (2021).

[HYL17] W. Hamilton, R. Ying, and J. Leskovec. "Representation learning on graphs: methods and applications". In: (2017).

[HRH02] P. D. Hoff, A. Raftery, and M. Handcock. "Latent space approaches to social network analysis". In: *Journal of the American Statistical Association* (2002).

[HLL83] P. Holland, K. B. Laskey, and S. Leinhardt. "Stochastic blockmodels: first steps". In: *Social Networks* (1983).

[Laa14] M. Laan. "Causal inference for a population of causally connected units". In: *Journal of Causal Inference* (2014).

[Leu21] M. Leung. "Network cluster-robust inference". In: *SSRN Electronic Journal* (2021).

[Ogb+17] E. L. Ogburn, O. Sofrygin, I. Diaz, and M. J. van der Laan. "Causal inference for social network data". In: *arXiv e-prints* (2017).

[PZ18] V. M. Panaretos and Y. Zemel. "Statistical aspects of wasserstein distances." In: *arXiv: Methodology* (2018).

[RFR16] R. Rastelli, N. Friel, and A. Raftery. "Properties of latent variable network models". In: *Network Science* (2016).

[Ros11] N. Ross. "Fundamentals of stein's method". In: *Probability Surveys* (2011).

[Rub80] D. B. Rubin. "Randomization analysis of experimental data: the Fisher randomization test comment". In: *Journal of the American Statistical Association* 371 (1980).

[SM16] C. Shalizi and E. McFowland III. "Estimating causal peer influence in homophilous social networks by inferring latent locations". In: *arXiv e-prints* (2016).

[ST11] C. R. Shalizi and A. C. Thomas. "Homophily and contagion are generically confounded in observational social network studies". In: *Sociol. Methods Res.* 2 (2011).

[TFS17] E. Tchetgen, I. Fulcher, and I. Shpitser. "Auto-g-computation of causal effects on a network". In: *Journal of the American Statistical Association* (2017).

[TK13] P. Toulis and E. Kao. "Estimation of causal peer influence effects". In: *Proceedings of the 30th International Conference on Machine Learning*. 3. 2013.

[Van10] T. VanderWeele. "Direct and indirect effects for neighborhood-based clustered and longitudinal data". In: *Sociological methods research* (2010).

[Vei+19] V. Veitch, M. Austern, W. Zhou, D. M. Blei, and P. Orbanz. "Empirical risk minimization and stochastic gradient descent for relational data". In: *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*. 2019.

[VWB19] V. Veitch, Y. Wang, and D. M. Blei. "Using embeddings to correct for unobserved confounding in networks". In: *Advances in Neural Information Processing Systems*. 2019.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] Please refer to section 7.
   (c) Did you discuss any potential negative societal impacts of your work? [No] This paper mainly focuses on proposing a new causal estimand for peer influence, and describing a method for its estimation; this is a purely methodological paper which does not have any negative societal impacts.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] Please refer to section 2 and to the assumptions made in Theorem 1.
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Appendix

Below we include the proof of Theorem 1, which shows the validity of our main measure, the causal estimand ψ_n for peer contagion on the network G_n. Since this theorem heavily relies on Stein's method ([Ros11]), we first detail the necessary background and results below.

We first characterize the network influence of the response variable Y_i for each node via the notion of *dependency neighborhoods* [Ros11].

Definition 3 (Dependency neighborhoods.). *An n-tuple* $(Y_1, Y_2, \ldots, Y_n)$ *of random variables has dependency neighborhoods* $N_i \subseteq \{1, 2, \ldots, n\}$, $i = 1, 2, \ldots, n$, *if* $i \in N_i$ *and* Y_i *is independent of* $\{Y_j\}_{j \notin N_i}$.

We show how Theorem 1 arises as a direct application of the following law of large numbers for relational variables.

Theorem 4 (Law of large numbers for dependent random variables. Modified from [Ros11].)**.** *Let $Y_1, \dots, Y_n$ be random variables such that $\mathbb{E}[Y_i^4] < \infty$, $\mathbb{E}[Y_i] = \mu_i$, $\sigma^2 = Var(\sum_i Y_i)$, and define $W = \sum_i (Y_i - \mu_i)/\sigma$. Assume that the variables $(Y_1, \dots, Y_n)$ have dependency neighborhoods N_i, $i = 1, 2, \dots, n$, respectively, and also define $D := \max_{1 \le i \le n} |N_i|$. Let Z be a standard normal random variable. Let d_W be the Wasserstein metric. The following inequality holds*

$$d_W(W, Z) \le \frac{D^2}{\sigma^3} \sum_{i=1}^{n} \mathbb{E}|Y_i|^3 + \frac{\sqrt{28} D^{3/2}}{\sqrt{\pi}\sigma^2} \sqrt{\sum_{i=1}^{n} \mathbb{E}[Y_i]^4}. \tag{A.1}$$

Proof. The proof of this result follows immediately from Theorem 3.6 in [Ros11]. □

We are now ready to return to the proof of the main theorem 1. For the network G_n, let us consider the following n-tuple of random variables $(\tilde{Y}_1, \dots, \tilde{Y}_n)$, where

$$\tilde{Y}_i = \frac{\mathbb{E}[Y_i | \text{do}(T = t^*), \{C_i\}_i, G_n]}{n}.$$

Let $\sigma^2 = \text{Var}(\sum_{i=1}^n \tilde{Y}_i)$ and $\mu_i = \mathbb{E}[\tilde{Y}_i] = \mathbb{E}\Big[\frac{\mathbb{E}[Y_i | \text{do}(T = t^*), \{C_i\}_i, G_n]}{n}\Big]$. Consider then the random variable $W_n = \sum_{i=1}^n (\tilde{Y}_i - \mu_i)/\sigma$.

Furthermore, note that, due to the structural equation model defined in section 2, and its assumption that the errors ε_{Y_i} have an i.i.d. structure, it follows that, for each node i, the size of the dependency neighborhoods N_i of $\mathbb{E}[Y_i | \text{do}(T = t^*), \{C_i\}_i, G_n]$, and implicitly $\tilde{Y}_i$, is the same as the degree of the node. Therefore, according to Theorem 4, it follows that for Z a standard normal variable, it holds that

$$d_W(W_n, Z) \le \frac{D_n^2}{\sigma^3} \sum_{i=1}^{n} \mathbb{E}|\tilde{Y}_i|^3 + \frac{\sqrt{28} D^{3/2}}{\sqrt{\pi}\sigma^2} \sqrt{\sum_{i=1}^{n} \mathbb{E}|\tilde{Y}_i|^4}, \tag{A.2}$$

where D_n is the maximal degree of the network G_n.

By the assumptions of Theorem 1, since $\mathbb{E}|\tilde{Y}_i|^4$ is finite, it follows via Jensen's inequality that all lower moments must be finite as well, and in particular, $\mathbb{E}|\tilde{Y}_i|^3 < \infty$ and $\text{Var}\mathbb{E}[\tilde{Y}_i]$. It therefore follows that $\sigma = O(\sqrt{n})$, $\sum_{i=1}^n \mathbb{E}|\tilde{Y}_i|^3 = O(n)$, and $\sum_{i=1}^n \mathbb{E}|\tilde{Y}_i|^4 = O(n)$. To obtain the convergence of ψ_n, it suffices to do so for its scaled version, W_n. In turn, it suffices to show that the right hand side of inequality A.2 is finite. For that, one needs to ensure that

$$\max \left\{ \frac{D_n^2}{\sigma^3} \sum_{i=1}^{n} \mathbb{E}|\tilde{Y}_i|^3, \frac{\sqrt{28} D^{3/2}}{\sqrt{\pi}\sigma^2} \sqrt{\sum_{i=1}^{n} \mathbb{E}|\tilde{Y}_i|^4} \right\}$$

is finite as $n \to \infty$. A little calculation shows that the right hand side becomes $O(1)$ if $D_n = O(n^{1/4})$.

Since random variable convergence in the Wasserstein distance implies convergence in distribution [PZ18], it follows that for some appropriately chosen sequences of real numbers $a_n \in \mathbb{R}$ and $b_n > 0$ we obtain that $b_n(\psi_n - a_n) \to_{\mathscr{D}} Z$. This further implies that there exists a real number ψ such that $\psi_n \to \psi$ in probability. This concludes the proof of the theorem and confirms the statistical validity of the estimand ψ_n.