Improving control over unobservables with network data

Vincent Starck[*][†]
LMU München

September 5, 2025

Abstract

This paper develops a method to conduct causal inference in the presence of unobserved confounders by leveraging networks with homophily, a frequently observed tendency to form edges with similar nodes. I introduce a concept of *asymptotic homophily*, according to which individuals' selectivity scales with the size of the potential connection pool. It contributes to the network formation literature with a model that can accommodate common empirical features such as homophily, degree heterogeneity, sparsity, and clustering, and provides a framework to obtain consistent estimators of treatment effects that are robust to selection on unobservables. I also consider an alternative setting that accommodates dense networks and show how selecting linked individuals whose observed characteristics made such a connection less likely delivers an estimator with similar properties. In an application, I recover an estimate of the effect of parental involvement on students' test scores that is greater than that of OLS, arguably due to the estimator's ability to account for unobserved ability.

Keywords: Causal Inference, Networks, Selection on unobservables, Homophily.

[*] I am very grateful to Susanne M. Schennach for her helpful comments and continuous support throughout this project. I also thank Peter Hull, Toru Kitagawa, Soonwoo Kwon, and Jonathan Roth for their valuable advice and numerous suggestions. This research has also benefited from the input of seminar participants at Brown, Michigan State, University of Pennsylvania, University of Alberta, University of Melbourne, University of Sydney, LMU München, and University of Amsterdam, as well as participants to the Microeconometrics class of 2022-2023 conference at Duke and the Encounters in Econometrics conference at Oxford. I gratefully acknowledge financial support from the European Research Council (Starting Grant No. 852332)

[†] V.Starck@lmu.de

1 Introduction

Estimating the effect of a treatment is a frequent goal in economics and social sciences. A common challenge is the presence of unobserved confounders that threaten the validity of the unconfoundedness assumption, which is typically necessary to perform inference with standard methods. Tools that strengthen control over variables that affect outcomes but are hard to measure, such as ability, culture, work ethic, tastes, *etc.*, are thus particularly valuable.

Recently, networks and datasets with a spatial structure have become increasingly available to researchers, providing new avenues for research. Homophily or assortative matching is an ubiquitous feature of empirical networks: nodes tend to associate with similar nodes (Lazarsfeld, Merton et al., 1954; Clark and Ayers, 1992; Case, Rosen and Hines Jr, 1993; McPherson, Smith-Lovin and Cook, 2001; Moody, 2001; Currarini, Jackson and Pin, 2009; Boucher and Mourifié, 2017; Dzemski, 2019). As Zeleneev (2020) note, homophily is likely to also operate through unobserved factors. An example is the tendency for people to form friendship ties based on ability (Clark and Ayers, 1992; Burgess et al., 2011; Boutwell, Meldrum and Petkovsek, 2017), a variable typically unavailable to the researcher.

Homophily then generates opportunities to create unobservable-adjusted comparison groups. For instance, if we are interested in the effect of parental involvement on student test scores, we may be concerned about unobserved confounding from differences in student ability. However, if students of similar ability are likely to be friends (Clark and Ayers, 1992; Burgess et al., 2011; Boutwell, Meldrum and Petkovsek, 2017), the omitted variable bias can be reduced by comparing connected students.

The paper develops the idea that homophilic networks can be exploited to derive consistent estimators of treatment effects in the presence of unobserved confounders. This is formally done under two main frameworks: either the network is asymptotically homophilic – homophily captures the essence of link formation – or the network at least features homophily in unobservables.

In the former case, I let the link formation probability vary with the size of the network: people become pickier to limit the number of connections or improve their average quality as the network expands. This is consistent with the common view that the average degree should not increase proportionally to the size of the network and that most networks are sparse. As people are able to form increasingly better matches with a larger pool of potential neighbors, they become more selective because of decreasing benefits per additional match, preference for quality of matches, or limited resources to devote to additional connections.

This accomplishes two things. First, the approach provides an asymptotic approximation that does not render the mechanism of network formation negligible in the limit: selectivity scales with the size of the connection pool, in the spirit of drifting sequences. As such, it contributes to the network formation literature with a model that can accommodate common features of empirical networks such as homophily, sparsity, clustering, etc. Second, it is sufficient to establish consistency and asymptotic normality for estimators that use m^{th}-order connections or people with more than c connections in common as comparison groups.

In an alternative framework, I only assume some preference for homophilic matching in the unobserved variables. This allows for any functional form in the observed covariates involved in link formation and possibly dense networks. In this scenario, comparison groups can be derived by comparing people whose observables differ but nevertheless connect. This hinges on the following intuition: if there is no observed rationale for two people being friends, the reason for their friendship likely lies in the unobserved world. If two people are connected despite their observables indicating that such a link was unlikely, they are more likely to be close in terms of unobservables. By suitably manipulating a discrepancy in observables and letting it grow with sample size, one can recover consistent estimators.

I provide results that allow for the estimation of the Conditional Average Treatment Effect (CATE), which provides a way to describe the heterogeneity of the treatment effect for sampled individuals. The conditional average effect may be the end goal of the analysis (when a specific unit is targeted for treatment or policy) or may be a prelude to aggregation to the Average Treatment Effect (ATE).

I define a general form of CATE estimator as a function of a group of counterfactual observations to be determined, then propose different choices to deal with different empirical issues. In all cases, estimators isolate increasingly better counterfactuals as to recover the CATE asymptotically. I show that the proposed estimators of the (C)ATE are asymptotically normal, enabling statistical inference.

Although results pertain to nonparametric estimators, the intuition is valid in parametric specifications and similar results are achievable under similar or weaker conditions. Propensity score analysis – and then possibly doubly robust estimators – could also be developed.

Finally, I demonstrate the feasibility and effectiveness of the method through both simulations and an empirical application. In the application, I obtain an estimate of the effect of parental involvement on students' test scores that suggests a greater impact than OLS does, arguably due to the estimator's ability to account for unobserved ability and motivation.

Related literature The paper is at the intersection of the literature on networks (Jackson, 2010; Graham, 2015; De Paula, 2017; Newman, 2018), in particular those featuring homophilic network formation (Boucher, 2015; Graham, 2016, 2017; Demirer, 2019; Gao, 2020; Mele, 2022), and estimation of treatment effects (Imbens, 2004; Imbens and Wooldridge, 2009; Imbens and Rubin, 2015), both of which considerably grew in size over the last decades.

In a related paper, Auerbach (2022) considers a partially linear outcome regression where the nonlinear term depends on an unobserved variable. Using information from a network whose formation hinges on the unobserved variable, he is able to recover consistent estimates of regression coefficients under general assumptions. See also Goldsmith-Pinkham and Imbens (2013); Hsieh and Lee (2016); Johnsson and Moon (2021), who use a related frameworks and provide a way to analyze peer effects.

Through the help of a pseudo-distance, Zeleneev (2020) devises a method to identify agents with similar values of latent fixed effects, which allows him to estimate parameters of interest while controlling for unobserved heterogeneity. Demirer (2019) provides partial identification results in linear models under homophilic behavior and proposes a comprehensive nomenclature for homophily.

The present paper considers general outcome equations in a causal inference framework, at the expense of some generality in the network formation process. Specifically, I consider a nonparametric potential outcome setup, but I impose structure on network formation, especially homophily in the unobservables. The potential outcome framework is suitable to discuss causality issues and the approach explicitely deals with the common concerns of treatment effect heterogeneity and nonlinearities. In addition, the method circumvents the need to define and estimate equivalent classes and focuses on the common case of sparse networks, in contrast to previous papers. Finally, homophilic structures allow for the use of higher-order neighbors or friends in common through triangular inequality relationships, which leads to a class of intuitive estimators that are easy to implement.

2 Improving control over unobservables using network data

2.1 Notation and assumptions

The sample is a cross-section of n individuals. The treatment status of individual i, $T_i \in \{0,1\}$, and the corresponding outcome, $Y_i = Y_i(T_i)$ with the potential outcome notation (Neyman, 1923; Rubin, 1974), are observed. As the notation for the outcome suggests, the Stable Unit Treatment Value Assumption (SUTVA) is maintained throughout.

The covariates, $X = (X^o, X^u) \in \mathcal{X}^o \times \mathcal{X}^u \overset{\text{def}}{=} \mathcal{X} \subset \mathbb{R}^d$, are divided into observed variables, X^o, and unobserved variables, X^u. There is a norm $\|\cdot\|$ on $\mathbb{R}^d$ (with some abuse of notation, this will be used to represent the norm on $\mathcal{X}^o$ or $\mathcal{X}^u$), typically Euclidean. I focus on continuously distributed covariates X, though discrete variables can be accommodated – typically under weaker conditions since concerns such as asymptotic bias disappear. To avoid technical difficulties with vanishing denominators, it will be convenient to assume that covariates have a smooth density bounded from below. I thus make the following assumption throughout the analysis:

Assumption 2.1 (Existence of bounded densities)**.**
The joint distribution of the covariates admits a density f with respect to Lebesgue measure. On the compact $\mathcal{X}$, the density is continuously differentiable and satisfies $f \geq \underline{f}$ for some positive $\underline{f}$.

Draws of (Y_i, T_i, X_i) are i.i.d. and realizations of a random variable are denoted by the corresponding lower-case letter. $B_r(x)$ denotes a ball of radius r centered at x. C represents a generic (positive) constant.

A network is given through a (binary) weighting/link matrix W, of size $(n \times n)$. The neighborhood $\mathcal{N}(i)$ refers to the links, friends, or connections of the node or individual i, *i.e* $\mathcal{N}(i) \overset{\text{def}}{=} \{j \in \{1, \ldots, n\} | W_{ij} = 1\}$, $\mathcal{N}_t(i)$ denotes neighbors with a specific treatment status t, *i.e* $\mathcal{N}_t(i) \overset{\text{def}}{=} \{j \in \{1, \ldots, n\} | W_{ij} = 1, T_j = t\}$. These definitions extend to higher-order neighbors, say of order m, which are denoted by $\mathcal{N}_t^m(i)$. Connections in common are given by $\mathcal{N}_t(i; j) \overset{\text{def}}{=} \mathcal{N}_t(i) \cap \mathcal{N}_t(j)$.

The goal is to conduct inference about treatment effects. In particular, I develop inference methods for the Conditional Average Treatment Effect (CATE), $\text{CATE}(x_i) \overset{\text{def}}{=} \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x_i]$, and then for the Average Treatment Effect (ATE), $\text{ATE} \overset{\text{def}}{=} \mathbb{E}[Y_i(1) - Y_i(0)]$. Although I focus on average treatment effects, the insights can be exploited to obtain *, e.g.,* quantiles of treatment effects or the average effect on the treated. The usual statement about omitting 'almost surely' qualifiers, in particular pertaining to conditional expectations, applies.

The following core assumptions are maintained throughout the paper:

Assumption 2.2 (Causal Inference)**.**
a) Unconfoundedness: $(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i | X_i$
b) Overlap: $0 < C < \mathbb{P}[T_i = 1 | X_i] < 1 - C < 1$

These two assumptions are ubiquitous in the treatment effect literature, although this version of unconfoundedness conditions on X instead of X^o. It is thus only assumed that treatment is independent of potential outcomes when conditioned on individual characteristics, including unobserved ones. Since covariates that may influence selection into treatment, such as ability, work ethic, or personal preferences, are typically unobserved, this is often a valuable relaxation: selection on some unobservables is allowed.

2.2 Network formation

Let i, j be two individuals and $i \neq j$. I focus on link-formation models of the type

$$W_{ij} = 1 \iff \eta_{ij} \leq w_n(h(X_i^o; X_j^o) + \|X_i^u - X_j^u\|) \tag{1}$$

where $w_n : \mathbb{R}^+ \to [0; 1]$ is a decreasing function that satisfies $\lim_{x \to \infty} w_n(x) = 0$. Typically, w_n would decrease with n to accommodate network sparsity, *e.g.*, $w_n(x) =$

$\max\{1 - s_n x, 0\}$ or $e^{-s_n \frac{1}{2} x^2}$. The function h is arbitrary but known, and $\eta_{ij} = \eta_{ji}$ are independent uniform[1] shocks, drawn independently of $(X_i, X_j, T_i, T_j, Y_i, Y_j)$. The dimensionality of the unobserved variables is arbitrary and matters only for rates of convergence.

Dyadic network formation processes are common in the literature*, e.g.*, Graham (2017); Gao (2020); Zeleneev (2020); Auerbach (2022); Johnsson and Moon (2021). Compared to more general specifications (for example, Auerbach (2022) posits that links are formed whenever $\eta_{ij} \leq w(X_i, X_j)$ and only imposes a weak continuity assumption on w), the model (1) adds some separability and homophily in the unobservables.

The model can be seen as a variant of Graham (2017)'s by considering $w_n \left(\frac{1}{a_i a_j} (h(X_i^o; X_j^o) + \|X_i^u - X_j^u\|) \right)$. This makes two main adjustments compared to his models of the form $w(X_{ij}'\theta + A_i + A_j)$: (i) the unobserved heterogeneity appears in multiplicative form, instead of additive terms, and (ii) there is an unobserved part of X_{ij} that satisfies homophilic restrictions. The analysis extends to such models under the assumption that expansiveness does not directly affect the outcome of interest.

Although homophily puts more structure on the network formation model, it often matches empirical observations. In addition, the multiplicative form is sometimes easier to interpret. In the limit of the asymptotic homophily model to be developed in the next Section, it would correspond to a scaling factor for probabilities. A person with characteristics $(x_i, 2a_i)$ is twice as expansive and has twice the probability of forming a link than a person with characteristics (x_i, a_i) as $n \to \infty$.

The other main feature of this network formation is the explicit dependence of w on network size, allowing for sparse networks. This is often the empirically relevant setup since the mean degree of a node is rarely expected to scale with the size of the network (Jackson, 2010; Newman, 2018).

The function h may feature homophily as well[2] as in Subsection 2.3[3] in which case

[1] Since one can apply an inverse cumulative distribution function on both sides to generate any distribution, the uniform assumption is made without loss of generality.

[2] In this case in particular, it may make sense to consider variables whose variance has been normalized to put them on the same scale. Nevertheless, the results hold if the norms weight each dimension differently as to reflect stronger selection in some covariates.

[3] More precisely, this Subsection combines observables and unobservables under a single norm, *i.e.*, $\|X_i - X_j\|$ which can, but need not, correspond to the sum of two norms.

$h(X_i^o; X_j^o) = \|X_i^o - X_j^o\|$. In some applications, it may be of interest to let h be an arbitrary function, as in Subsection 2.4. For instance, some work relationships may warrant skill complementarity, in which case there is non-(possibly anti-) homophilic selection in a covariate.

The model can be given the usual interpretation of 'link creation under a mutual positive utility of forming a link' ($w - \eta$ then reflecting utility; see, *e.g.*, Jackson (2010)), where people derive more utility from interacting with similar individuals, or rationalizes the idea that people with similar characteristics are more likely to meet and thus to form a connection. Nevertheless, since the network is primarily seen as information to draw from, this rationale may not be necessary. For instance, if individuals end up developing similar characteristics after randomly forming connections, a researcher that observes the network after covariates have evolved could use the present framework. In other words, (1) need not be the structural equation for network formation but should approximate the relationship between the links and the covariates relevant to selection at the time of observation.

I first explore the case of *asymptotic homophily*, *i.e.*, homophilic behavior is the core mechanism of network formation and individuals' selectivity is tied to the size of their potential matching pool. This provides an asymptotic theory when homophilic behavior is pronounced relative to network size and formalizes the intuition that connections among individuals can be used to form comparison groups. It also shows the identifying power of homophilic restrictions under simpler conditions than the results of the later sections and provides a network formation model compatible with many empirically-relevant features.

A second framework is discussed in Subsection 2.4, possibly letting the link formation be independent of sample size.

2.3 Asymptotic homophily

2.3.1 The asymptotic homophily framework

Suppose associations are captured by homophily, so that the probability of a connection is decreasing in $\|X_i - X_j\|$, and the network is sparse: the average degree of a node increases arbitrarily slowly. As people are able to form increasingly better matches with a larger pool of potential neighbors, they become more selective because of decreasing benefits per additional match, preference for quality of matches, or limited resources to devote to additional connections.

To reflect this behavior, the sequence of functions w_n must satisfy two conditions. First, the sequence must be decreasing in order to decrease the probability of forming connections as n rises. Homophily further suggests that people penalize dissimilar individuals increasingly more harshly so that the average match quality (in terms of homophilic preferences) increases.

Functions of the form $w_n(x) \geq g(s_n x)$ are consistent with such behavior – homophily becomes more prevalent as n rises – irrespective of the exact form of w_n (or g). I adopt the following definitions:

Definition 2.1 (Asymptotic homophily)**.** a) Network formation is asymptotically homophilic if $W_{ij} = \mathbb{1}_{\eta_{ij} \leq w_n(\|X_i - X_j\|)}$, $w_n(x) \geq g(s_n x)$, where $g : [0; \infty[\to [0; 1]$ is decreasing and $\lim_{n\to\infty} s_n = \infty$.
b) Network formation is regularly asymptotically homophilic if $W_{ij} = 1$ whenever $w_n(\|X_i - X_j\|) \geq \eta_{ij}$, $w_n(x) = g(s_n x)$, where $g : [0; \infty[\to [0; 1]$ is a decreasing function such that $0 < \int_{\mathbb{R}^d} g(\|y\|)\, dy < \infty$, and $\lim_{n\to\infty} n {s_n}^{-d} = \infty$.

Part a) formalizes the idea of *asymptotic homophily*; part b) provides regularity conditions for asymptotic results. The definition is new and the resulting formation mechanism is consistent with the empirical regularities of social networks (Jackson, 2010): sparsity[4], transitivity/clustering[5], degree heterogeneity[6], and homophily.

[4] It is natural to let the average degree of a node be constant or grow only slowly for most applications. For instance, the average number of friends is typically viewed as constant or slowly increasing as the network expands, requiring the probability of forming a link to decrease with the size of the network. Letting the degree increase, albeit slowly, allows one to take advantage of asymptotic approximations.

[5] Intuitively, clustering occurs because groups of similar individuals tend to form connections. Moreover, as shown in the appendix, the clustering coefficient does not vanish asymptotically, in contrast to Poisson random graphs (Erdős, Rényi et al., 1960) or configuration models (Bender, Canfield and McKay, 1990).

[6] Because of the influence of covariates, the expected degree varies across individuals. The underlying density affects the degree distribution because people with common characteristics have an easier time forming connections. The sequence s_n can also be refined using pair-specific rates – say, $s_n/(a_i a_j)$ as previously mentioned – to account for popularity/expansiveness-type of behaviors, although this is not pursued here for simplicity.

Asymptotic homophily formalizes the notion that homophilic behavior is pronounced relative to sample size in the spirit of a drifting sequence[7]; the degree to which individuals are selective is, in a sense, preserved as we proceed to an asymptotic approximation.

2.3.2 Comparison groups

Under an asymptotically homophilic network formation process, it is possible to derive estimators whose bias is asymptotically negligible, even if some confounders are unobserved. Given a comparison group $\mathcal{C}_i$ for individual i, I define a CATE[8] estimator

$$\widehat{\text{CATE}}(x_i; \mathcal{C}_i) \overset{\text{def}}{=} \frac{1}{|\mathcal{C}_{i1}|} \sum_{j \in \mathcal{C}_{i1}} Y_j - \frac{1}{|\mathcal{C}_{i0}|} \sum_{j \in \mathcal{C}_{i0}} Y_j \tag{2}$$

where $\mathcal{C}_{it} \overset{\text{def}}{=} \mathcal{C}_i \cap \{j | T_j = t\}$. Note that the comparison group will vary with the sample size, although the dependence is left implicit in the notation.

Remark: In what follows, the vector X from the unconfoundedness condition is assumed to be part of the network formation for ease of exposition. In practice, the binding restriction is that X^u belongs to it: If some observed covariate (x_k) that affects the outcome does not influence network formation, it can be controlled for nonparametrically by multiplying each weight $\mathbb{1}(j \in \mathcal{C}_{it})$ with kernel weights $K_b(\|X^o_{ik} - X^o_{jk}\|)$ throughout.[9] Because the weights based on the network act as (noisy versions of) kernel weights, the use of such hybrid weights is naturally nested in the current framework.

The first idea is to rely on friends to construct a comparison group, , *i.e.*, set $\mathcal{C}_i = \mathcal{N}(i)$. This generates a CATE estimator based on the difference between treated friends and non-treated friends. Thanks to triangular inequality relationships and

[7] Similarly to Bekker (1994), who analyzes the behavior of IV estimators with many instruments, or Borusyak, Hull and Jaravel (2022), who analyze shift-share instruments with a growing number of shocks. "The sequence is designed to make the asymptotic distribution fit the finite sample distribution better. It is completely irrelevant whether or not further sampling will lead to samples conforming to this sequence" (Bekker, 1994).

[8] Note that x_i is not fully observed. The CATE of a given individual is identified, but not the underlying function of x. For this reason, the CATE is mainly interesting on its own when one cares about the treatment effect of a specific unit.

[9] Alternatively, it is possible to make linear adjustments for observed covariates, as in Appendix B.

the nature of homophily, however, one can extract additional information from the friendship network. For instance, one can consider friends of friends or higher-order friendships:

$$\begin{aligned}\widehat{\text{CATE}}(x_i; \cup_{m=1}^{M} \mathcal{N}^m(i)) = & \frac{1}{|\cup_{m=1}^{M} \mathcal{N}_1^m(i)|} \sum_{j \in \cup_{m=1}^{M} \mathcal{N}_1^m(i)} Y_j \\ & - \frac{1}{|\cup_{m=1}^{M} \mathcal{N}_0^m(i)|} \sum_{j \in \cup_{m=1}^{M} \mathcal{N}_0^m(i)} Y_j\end{aligned} \tag{3}$$

for some upper order of friendship M. For $M = 1$, this is the simple estimator that compares treated friends and non-treated friends. Although the estimator averages more observations as M increases, the comparison group increasingly selects observations whose characteristics differ from those of i. $M = 1$ can be a reasonable choice if the ATE is the target, but higher values can be useful, *e.g.*, to estimate a specific CATE.

An alternative estimator relies on having at least c friends in common:

$$\begin{aligned}\widehat{\text{CATE}}(x_i; \{j|\ |\mathcal{N}(i;j)| > c\}) = & \frac{1}{|\{j|\ |\mathcal{N}_1(i;j)| \geq c\}|} \sum_{j \in \{j|\ |\mathcal{N}_1(i;j)| \geq c\}} Y_j \\ & - \frac{1}{|\{j|\ |\mathcal{N}_0(i;j)| \geq c\}|} \sum_{j \in \{j|\ |\mathcal{N}_0(i;j)| \geq c\}} Y_j\end{aligned} \tag{4}$$

Although both estimators allow for consistent estimation of treatment effects, the composition of the underlying comparison groups can differ significantly. Therefore, it can be a useful robustness check to compare their treatment effect estimates, for instance if one is concerned about peer effects or related issues that would likely affect these estimators in different ways.

As an illustration, consider Figure 1 where the comparison group for individual i with unobserved characteristics $x_i \in \mathbb{R}^2$ is shown in red and the remaining observations in black. The leftmost picture is the unknown[10] group that consists of all observations below a certain distance. Next, on the right, friends are used as the comparison group, providing a noisy version of (i): selected observations tend to fall

[10] Except in the extreme network formation process in which individuals select friends deterministically conditional on covariates: $w(x) = \mathbb{1}_{[0,C]}(x)$. Then, the first two pictures become identical.

close to x_i, but some close observations are ignored while observations farther away may be selected nonetheless.

In the third picture, one looks at friends of friends. This allows us to make use of more observations, which will reduce the variance of the estimator, but there is also a tendency to grab more observations outside the sphere. Finally, the last picture selects individuals who have at least two friends in common with i. This typically reduces the bias compared to using the previous groups, but selects fewer observations.

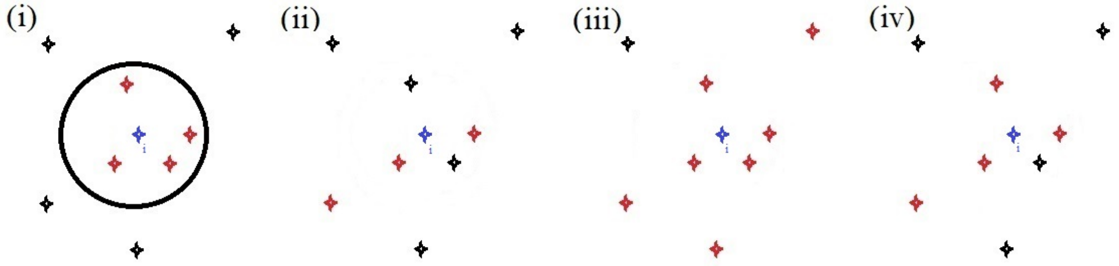

Figure 1: Comparison among possible comparison groups for individual i. Observations are represented by stars (blue = i; red = included in $\mathcal{C}_i$; black = not included). From left to right, (i) (Infeasible) individuals within a given distance of x_i, (ii) individuals who are friends with i, (iii) individuals who are friend with i or friend of a friend of i, (iv) individuals who have two friends in common with i.

2.3.3 CATE inference

Under *asymptotic homophily*, standard asymptotics are achievable with estimators such as (3) or (4), whose bias asymptotically disappears. As for the variance, its collapse is based on averaging over increasingly many observations. The asymptotic behavior of the size of a comparison group depends on the speed of convergence of w_n, of which the following lemma provides a formal analysis.

Lemma 2.1. *The number of connections for i exceeds any real number with probability approaching one if* $\lim_{n\to\infty} n \int_{\mathbb{R}^d} w_n(\|x_j - x_i\|) f(x_j)\ dx_j = \infty$.
Under regular asymptotic homophily, this holds, and moreover
a) The probability of forming a connection of order up to M is $O\left(s_n{}^{-d}\left(ns_n{}^{-d}\right)^{M-1}\right)$.
The number of connections for i of order up to M then exceeds any real number with

probability approaching one.
b) If ${ns_n}^{-2d} \to 0$*, the probability of* i *and* j *having at least* c *friends in common is* $O\left({s_n}^{-2d}(n{s_n}^{-2d})^{c-1}\right)$.

The lemma, proven in the appendix, provides conditions under which the sizes of potential comparison groups grow to infinity. It also specifies the rates at which the probabilities of forming a connection up to the M-th or having more than c friends in common decrease under *asymptotic homophily*, and when all corresponding counts go to infinity. With overlap, the lemma ensures the number of treated and untreated connections both grow to infinity.

The condition in the lemma states that w_n must not vanish too fast to ensure that connections are still being formed. The formal criterion analyses the integral $\int_{\mathbb{R}^d} w_n(\|x\|)f(x+x_i)\,dx$, suggesting that link functions that do not depend on network size or vanish uniformly slowly enough such as $w_n(x) = {s_n}^{-1}g(x)$ induce unbounded friend counts.

Basically, individuals must not become too selective too quickly to ensure that they keep forming connections. This is natural in many networks (*e.g.*, friendship network) as the expected degree is often viewed as only slowly increasing. In the *asymptotic homophily* framework, this means $n{s_n}^{-d}$ increases at a low rate.

Asymptotic homophily is not only consistent with a growing count, but also implies an improvement in matching quality that is absent when w_n is constant or grows uniformly. Specifically, a sequence such as $w_n(x) = {s_n}^{-1}g(x)$ would stabilize the "posterior" distribution $f_{X_j|j\in\mathcal{N}(i)}$; it does not imply that people improve their average match in larger networks.

As a result, *regular asymptotic homophily* will be key in securing consistency properties. An important part in establishing these is the analysis of the bias

$$\begin{aligned}\mathbb{B}_i \overset{\text{def}}{=} \,&\mathbb{E}[Y_j(1)|j \in \mathcal{C}_i, T_j = 1]] - \mathbb{E}[Y_j(1)|X_j = x_i]\\ &- (\mathbb{E}[Y_j(0)|j \in \mathcal{C}_i, T_j = 0] - \mathbb{E}[Y_j(0)|X_j = x_i])\end{aligned}$$

which will be shown to disappear under various conditions. Specifically, the following conditions on w_n will ensure that the bias vanishes.

Assumption 2.3 (Hölder continuity of CATE and convergence of link function)**.** a) CATE(x) is Hölder continuous with exponent α on a neighborhood of x_i, *i.e.* for any x, y in the neighborhood $\|\text{CATE}(y) - \text{CATE}(x)\| \leq C\|y - x\|^\alpha$ for some $\alpha > 0$.
b) For some $\varepsilon_n \downarrow 0$, either $\mathcal{C}_i = \cup_{m=1}^M \mathcal{N}^m(i)$ and $\sum_{m=1}^M (w_n(\frac{\varepsilon_n}{m}))^m = o({s_n}^{-d})$, or $\mathcal{C}_i = \{j| \, |\mathcal{N}(i;j)| \geq c\}$ and $(w_n(\frac{\varepsilon_n}{2}))^c = o({s_n}^{-d})$.

Hölder continuity is a standard assumption that imposes a mild degree of smoothness in the CATE. Part b) of the assumption restricts the way w_n converges to 1; it requires a sufficiently fast convergence away from the origin.

Although consistency can be achieved under very weak conditions, the rates of convergence may be low and the conditions on w_n may be hard to interpret. Assuming that the underlying functions of X are sufficiently smooth implies a clean rate of $O({s_n}^{-2})$ for the bias under regular asymptotic homophily. In practice, I require existence of first-order derivatives:

Assumption 2.4 (Existence of Derivatives)**.** The conditional expectation $\mathbb{E}[Y_i(t)|X_i = x]$ and the propensity score $p(x) \overset{\text{def}}{=} \mathbb{P}[T_i = 1|X_i = x]$ are continuously differentiable.

Now, the consistency theorem reads

Theorem 2.1 (Consistency)**.** *a) Suppose* $\lim_{n\to\infty} n \int_{\mathbb{R}^d} w_n(\|x\|) f(x + x_i) \; dx = \infty$*, and* $\mathbb{E}[Y_j(t)^2] < \infty$ *for* $t = 0, 1$.
Then, $\widehat{\text{CATE}}(x_i; \mathcal{C}_i)$ *is consistent for* CATE(x_i) *under Assumption 2.2. Moreover, the bias satisfies* $\mathbb{B}_i = O(\varepsilon_n^\alpha + {s_n}^d R)$ *with* $R = \sum_{m=1}^M w_n(\frac{\varepsilon_n}{m})^m$ *and* $R = w_n(\varepsilon_n/2)^c$*, respectively, for* $\varepsilon_n \downarrow 0$ *as in Assumption 2.2.*
b) If the network formation is regularly asymptotically homophilic, Existence of derivatives holds, and the covariate density has bounded second-order derivatives, then the estimators are consistent with bias $\mathbb{B}_i = O({s_n}^{-2})$.

The theorem is proven in the appendix. The main difficulty in part a) is to derive an expression for the bias that can subsequently be bounded via homophilic assumptions and triangular inequalities. In the second part, the existence of derivatives allows the use of Taylor expansions so that the derivation shares similarities with nonparametric kernel analysis, though the noisy matching through w_n makes the problem non-standard.

To sum up, consistency is secured provided homophilic behavior is preserved as sample size grows and w_n does not drop too fast with sample size. If w_n does not decrease to zero (so that there is still an asymptotic bias due to different covariates) or does so too quickly (so that the friend count shrinks), the estimator is no longer consistent. Because mean degrees are typically not increasing quickly with network size, these situation are seldom empirically relevant. When the network is dense – the probability of forming a link does not drop to 0 – inference results can be obtained using the method of Subsection 2.4.

I finalize the analysis with an asymptotic normality result: CATE estimators are asymptotically normal at x_i. Formally,

Theorem 2.2 (Asymptotic Normality)**.** *Suppose that Consistency assumptions hold and the potential outcomes have $2 + \delta$ moments for some $\delta > 0$ (conditional on $X = x_i$). Then, for a bias $\mathbb{B}_i$ at location x_i, the (conditional) asymptotic distribution reads*

$$\sqrt{|\mathcal{C}_i|}(\widehat{\mathrm{CATE}}(x_i; \mathcal{C}_i) - \mathrm{CATE}(x_i) - \mathbb{B}_i) \xrightarrow{d} \mathcal{N}(0; V)$$

where $V = \frac{\mathbb{V}[Y_j(1)|X_j=x_i]}{\mathbb{P}[T_j=1|X_j=x_i]} + \frac{\mathbb{V}[Y_j(0)|X_j=x_i]}{\mathbb{P}[T_j=0|X_j=x_i]}$.
Moreover, $\mathbb{B}_i$ is asymptotically negligible if $\mathrm{CATE}(x)$ is Hölder continuous with exponent α on a neighborhood of x_i and one of the following holds:
(i) $\mathcal{C}_i = \cup_{m=1}^{M} \mathcal{N}^m(i)$ with $\sum_{m=1}^{M} w_n(\frac{n^{-\gamma}n}{m})^m = o({\lambda_n}^2 \frac{M}{n})$ for $\gamma > \frac{1}{2\alpha}$ or
(ii) $\mathcal{C}_i = \{j |\ |\mathcal{N}(i;j)| \geq c\}$ and ${s_n}^d w_n(n^{-\gamma})^c = o({\lambda_n}^{-1/2})$ and $\gamma > \frac{1}{2\alpha}$ or
(iii) Network formation is regularly asymptotically homophilic, Existence of Derivatives holds, the covariate density has bounded second-order derivatives, and $\frac{\sqrt{\lambda_n}}{{s_n}^2} \to 0$ for $\lambda_n \overset{\text{def}}{=} n{s_n}^{-d}$.

When the CATE itself is of particular interest, the theorem provides a way to perform standard inference. The variance can be estimated with the same kind of truncation methods and one would generally increase M, as long as the bias remains negligible, in order to decrease the variance. Given that the size of the comparison group increases with powers of M, including neighbors of up to the fourth order, but not much higher, usually makes sense.

The size of the comparison group grows at rate λ_n. If λ_n grows too fast compared to s_n, the estimator is still consistent though inference requires handling the bias. In social networks such as friendship networks, it is often reasonable to let the average degree grow only slowly (Newman, 2018) so that a slow growth of λ_n such as $\ln(n)$ can be justified.

Finally, consistent estimation of treatment effects at given x_i's suggests that inference about average effects is possible. This is the next result: the average treatment effect can be estimated under unobservable-robust unconfoundedness.

2.3.4 ATE inference

Given the last two theorems, one obtains an estimator $\widehat{\text{ATE}}$ by averaging over a CATE estimator at all x_i. The resulting ATE estimator is then consistent and asymptotically normal under the theorems' conditions and regularity conditions.

Specifically, consider $\frac{1}{n}\sum_{i=1}^n \widehat{\text{CATE}}(x_i, \mathcal{C}_i)$ and collect the terms involving Y_i for each i to obtain $\frac{1}{n}\sum_{i=1}^n \left(T_i \sum_{j\in\mathcal{C}_i} \frac{1}{|\mathcal{C}_{j1}|} - (1-T_i)\sum_{j\in\mathcal{C}_i}\frac{1}{|\mathcal{C}_{j0}|}\right) Y_i$.[11][12]

Formally, this leads to the estimator

$$\widehat{\text{ATE}} \overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n (2T_i - 1)\theta_i Y_i \tag{5}$$

where $\theta_i = \sum_{j\in\mathcal{C}_i} \frac{1}{|\mathcal{C}_{jT_i}|}$. Its asymptotic distribution is described by the following theorem:

Theorem 2.3. *Suppose that $\mathbb{E}[Y_j(t)^2] < \infty$, that condition (iii) of Theorem 2.2 holds, and that $\frac{\sqrt{n}}{{s_n}^2} \to 0$. Then,*

$$\sqrt{n}(\widehat{\text{ATE}} - \text{ATE}) \xrightarrow{d} N\left(0; \mathbb{E}\left[\frac{\mathbb{V}[Y_i(1)|X_i]}{p(X_i)} + \frac{\mathbb{V}[Y_i(0)|X_i]}{1-p(X_i)}\right] + \mathbb{V}[\text{CATE}(X_i)]\right)$$

The theorem is proven in the appendix. It can be seen that the asymptotic variance reaches the semiparametric efficiency bound for the estimation of the ATE (Hahn (1998); Hirano, Imbens and Ridder (2003)). The theorem enables inference about the average treatment effect under (unobservable-robust) unconfoundedness in an asymptotically efficient way.

[11] This last expression holds as long as $i \in \mathcal{C}_i$ implies $j \in \mathcal{C}_i$, which is true for the comparison groups previously discussed.

[12] Although the weights are well defined with probability approaching one, it may be useful to regularize them in finite sample by adding a small offset to the denominators.

2.4 Dense networks with homophily in the unobservables

This section considers a link formation model that only assumes homophilic behavior in unobservables and is suitable for dense networks. The method provides insight about how to create sub-groups that are increasingly close in terms of unobservables and can be adapted to deal with empirical concerns such as matching on treatment status. Although dense networks are less frequent, this approach is thus valuable as it covers additional network structures and applications of interest.

Now, a link exists between i and j if $\eta_{ij} \leq w(h(X_i^o; X_j^o) + \|X_i^u - X_j^u\|)$. The function w does not explicitly depend on n anymore, although some dependence could be accounted for. The function h need not be homophilic nor separable in the observed X^o but is assumed known. Most results would apply with minor modifications if a lower bound with the relevant properties can be obtained. The dimension of $\mathcal{X}^u$ is denoted by d_u.

When observables affect the outcome, they can be controlled for using standard methods.[13] For ease of exposition, I focus on presenting how to control for the unobserved components.

I consider again $\frac{1}{|\mathcal{C}_{i1}(\kappa)|}\sum_{j\in\mathcal{C}_{i1}(\kappa)} Y_j - \frac{1}{|\mathcal{C}_{i0}(\kappa)|}\sum_{j\in\mathcal{C}_{i0}(\kappa)} Y_j$, where the comparison group now depends on a truncation parameter κ. It will play a key role by placing a lower bound on h, inducing a closer distribution of unobservables among friends.

The main idea is that if there is no observed rationale for two people being friends, it becomes more likely that there is a unobserved reason for their friendship. Then, people that are friends despite a high value of h are less likely to differ strongly on unobservables. To see this, consider for illustration the case where people reject friendship with anyone whose quality of match falls below a certain threshold (*i.e.*, $w(x) = 1$ for x large enough). Then, two friends with an h close to the boundary must have close unobservables since a high discrepancy in unobservables would have brought $h + \|X_i^u - X_j^u\|$ above the threshold.

This suggests using comparison groups of the form $\mathcal{C}_i(\kappa) \overset{\text{def}}{=} \{j \in \mathcal{N}(i) | h(x_i^o, x_j^o) > \kappa\}$. The estimator then truncates the sums to select individuals whose observed characteristics make them unlikely to be friends. κ is viewed as a sequence converging

[13] For instance, by constructing again products with kernel weights or performing regression adjustments.

to ∞ at a rate to be determined. Using this strategy, a counterpart to Theorem 2.2 can be established with κ-truncation replacing *asymptotic homophily.*

Theorem 2.4. *Suppose* $\mathrm{CATE}(x)$ *is Hölder continuous with exponent* α *on a neighborhood of* x_i*,* w *has bounded support, and there exist a sequence* $\lambda_n \to \infty$ *and a sequence* b_n *such that* κ*-truncation satisfies* $nw(\kappa + b_n)b_n^{d_u+1} = O(\lambda_n)$ *and a sequence* $\varepsilon_n \downarrow 0$ *satisfying* $\kappa + \varepsilon_n > \sup\{supp\{w\}\}$ *eventually. Then, the estimator satisfies*

$$\sqrt{|\mathcal{C}_i|}(\widehat{\mathrm{CATE}}(x_i; \mathcal{C}_i(\kappa)) - \mathrm{CATE}(x_i) - \mathbb{B}_i) \xrightarrow{d} \mathcal{N}(0; V)$$

and the bias is negligible if $\sqrt{\lambda_n}{\varepsilon_n}^{\alpha} \to 0$*.*

The condition on $w(\kappa + b_n)b_n^{d_u+1}$ restricts the speed at which κ can increase so that the number of observations used in estimating the CATE keeps growing. The first part pertains to the behavior of the w function; the second term pertains to the space in which unobservables live.

The term $w(\kappa + b_n)$ comes from the increasing cost of truncating as potential connections are accepted at decreasing rates. In the presence of a discontinuity at the end of the support, *i.e.*, $w(x) = a\mathbb{1}_{x\leq D}$ for $a \in]0; 1]$ and $D \in \mathbb{R}^+$, this term disappears. The second term, ${b_n}^{d_u+1}$, is the result of forcing unobservables in a b_n-ball using values of h lying between κ and $\kappa + b_n$.

A natural estimator of the end of support, if unknown, is the highest value of h among all i, j satisfying $W_{ij} = 1$. Unbounded support of w can be accommodated, provided the function vanishes sufficiently quickly (the threshold is exponential: functions decaying faster than $w = e^{-x}$ provide a sufficiently fast decay). In this case, a factor of $\rho_h(\kappa + b_n/2)$, where ρ_h is (a lower bound on) the tail decay of h conditional on $X_j^u \in B_r(x_i^u)$ for some r, has to be added. The reason is the tail decay of the density while one seeks increasingly larger values of h.

The conditions on ε_n ensure that the bias disappears sufficiently fast to enable standard inference. A more primitive statement is $w(\kappa + \varepsilon_n) = o(\lambda_n/n)$, which mirrors conditions such as those in Assumption 2.3, part b); this boils down to $\kappa + \varepsilon_n$ eventually crossing the end of the support of w when it is finite.

Finally, one can again consider alternative comparison groups – for instance, by including friends of friends or people with sufficiently many common friends – or construct the truncated group differently – for instance, considering triangle of friends can give more leeway to vary h.

Overall, these results show that suitably refining a comparison group such as friends using the observed covariates allows one to isolate increasingly good matches in terms of unobservables. Although the levels of the unobserved variables is not identified, groups with increasingly similar values can be recovered. The rates are slower than those arising from asymptotically homophilic networks, though it should be noted that they focus directly on the unobserved components while the observed variables can be adjusted in more traditional ways, *e.g.*, regression adjustments. This is sufficient to recover consistent estimators of treatment effects under unobserved confounders with dense networks and to learn who is comparable to whom in terms of unobservables.

3 Simulations

I assess the performance of the estimators through simulations. I consider various outcome equations and measure the resulting root mean square error (RMSE). The RMSE of standard estimators making use of observed variables is provided for comparison.

The variables are generated as follows: a random vector V, whose components are uniform, triangular, and sum of three uniforms, is used to construct

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 0.7 & 0.3 & 0 \\ -0.1 & 1 & 0.4 \\ 0 & -0.6 & 0.7 \end{pmatrix} \begin{pmatrix} V_1 \\ V_2 \\ V_3 \end{pmatrix}$$

so that there is a non-trivial correlation structure among the components of X. In the baseline, $\mathrm{corr}(X_1, X_2) = 0.28$, $\mathrm{corr}(X_1, X_3) = -0.26$, and $\mathrm{corr}(X_2, X_3) = -0.32$, though the results exhibit similar patterns for weaker or stronger correlations. The variances are normalized to one. The first two variables are observed, but the last one is not, *i.e.*, $X^o = (X_1, X_2)$ and $X^u = X_3$.

The propensity score follows a logistic distribution with argument $X\beta$, where $\beta = \begin{pmatrix} 1 & 1 & \beta_3 \end{pmatrix}'$, and the treatment status is then drawn conditional on X. The parameter β_3 controls selection on unobservables and takes value in $\{0, 0.5, 1\}$. In the first case, the probability of being treated does not change with X_3; in the last case, the unobserved variable is on a similar footing as each of the observed ones. The performance of traditional methods that cannot account for the unobserved component is expected to deteriorate as β_3 increases.

The outcome equation is given by $y = 5 + \text{CATE}(x)T + g(x) + \varepsilon$. I explore three specifications:

- Homogeneous treatments effects ($\text{CATE}(x) \equiv 1$) with linear impact of unobservables (g is linear in x)
- Heterogeneous treatments effects ($\text{CATE}(x) = 2\Phi(-x_1 + x_2 + x_3)$) with linear impact of unobservables (g is linear in x)
- Heterogeneous treatments effects ($\text{CATE}(x) = 2\Phi(-x_1 + x_2 + x_3)$) with quadratic impact of unobservables (g contains both a linear term in x and a quadratic term ${x_3}^2 + x_2 x_3$)

The network formation process uses $w(x) = e^{-\frac{1}{2}x^2}$.[14] The baseline sample size is $n = 500$ and s_n is calibrated so that the average number of friends is roughly five-six, which is around the number of close friends people often report, *e.g.*, in the application. As n rises, s_n evolves at the rate $n/\ln(n)$).

The researcher controls for observables using kernel weights that multiply network weights, as previously described. People form friendship links based on X_2 and X_3. As a result, only X_1 is unaccounted for in network weights, but a researcher observing X_2 may want to further include it in the kernel weights. I consider both possibilities (referred to as base- and over-controlling case below). Alternative methods (*e.g.*, OLS) always make use of all observables.

The results for all estimators and sample sizes $n = 500, 2000$ are reported in Appendix C. For a simple and representative review, the results for the specifications $M = 1, c = 2$ (orange and yellow lines, respectively) and OLS (blue line) are graphically depicted below for the 3 types of outcome equations and $n = 500$.

[14] Other specifications such as $w(x) = \mathbb{1}_{x<1}$ or $w(x) = \max\{1 - x, 0\}$ deliver similar results.

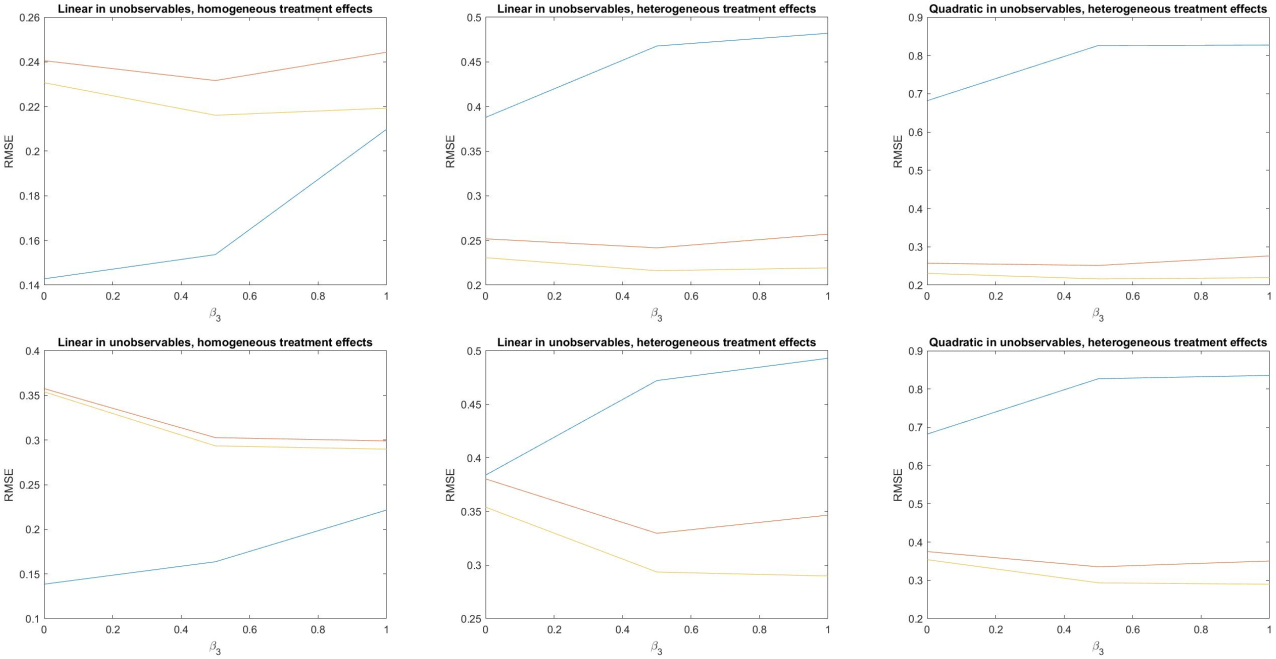

Figure 2: RMSE of OLS, ATE estimators using $M = 1, c = 2$. The first row refers to base-control case; the second row refers to over-control-case.

In the linear homogeneous case, OLS performs well[15] when there is no selection on unobservables, but its performance quickly deteriorates as β_3 increases. If the selection on the unobserved variable is on par with the level of selection on observables, it reaches an RMSE similar to that of the proposed estimators. Furthermore, it behaves very poorly when the unobservables become more prevalent in the functional form or the treatment response.

In contrast, estimators that leverage network information are able to perform similarly regardless of the strength of selection in unobservables and dominate OLS across specifications unless we force homogeneity and linearity. In the presence of unobserved confounding, the properties of OLS become quickly unappealing, especially when the unobserved variable has nonlinear impacts on the outcome or the treatment effect. Other methods based on estimating the propensity score also fail to deliver reliable estimates of the ATE as they can only accommodate heterogeneity and nonlinearities that arise from observed confounders. As a result, they are even

[15] Note that even in the linear case with homogeneous treatment with no selection, OLS could theoretically exhibit a higher RMSE. Although unbiased, the OLS estimator using observables could have a higher variance than an estimator that accounts for the unobserved variable.

dominated by OLS and *a fortiori* by the proposed estimators.

Finally, the simulations provide some guidance for empirical decisions. The choice of M or c does not appear to have a strong influence on the performance of the estimator. Over-controlling leads to a general increase in RMSE, but the estimators still perform well and improve substantially over OLS in nonlinear or heterogeneous settings. Therefore, it is advisable to separately control for observables whose relevance to network formation is uncertain.

4 Application

I provide an application of the method to the estimation of the effect of parental involvement on students' test scores. I use the dataset from the project "Attitudes and Relationships among Primary and High School Students", see Portella and Kirschbaum (2022). The dataset contains information about 4409 Brazilian high-school students, their beliefs, and friendship ties among them.

The outcome of interest is the average grade, ranging from 0 to 10, and the treatment is the level of parental support. The average is taken over math, Portuguese, English, history, geography, and art grades.[16] For comparability across school years, I normalize the grade by subtracting the mean over students in a given year. The dataset contains a (self-reported) score for parent support which ranges from around -3 to 1; the treatment is dichotomized by truncating around the mean of 0 and the goal is to estimate the ATE.

Although some covariates are available – age, gender, race, religion, class dummies, whether parents are employed, poverty, and importance of study for the child, they are at best proxies for the underlying causes. Thus, omitted variable bias remains a concern, especially due to the usual unobserved ability.

The sign of the omitted variable bias is unclear in this context, even assuming that the score for the importance of studying controls adequately for motivation. Intuitively, children with lower ability may require more attention but are also more likely through heredity to have less able parents, who may be less inclined or able to help.

[16] Although additional subjects are available, including them would require dropping a substantial number of observations because of missing data.

Formally, a possible model for the individual's grade, y_i, would be $y_i = F(a_i; m_i; s_i)$ where F is an unknown function, a_i is ability, m_i is the level of (intrinsic) motivation[17], and s_i is parental support. Denoting parents' ability by A_i, one could specify $a_i = A_i + \text{noise}_i$ and $s_i = \delta_1 a_i + \delta_2 A_i + \text{noise}_i$ with independent noises. This would imply $s_i = (\delta_1 + \delta_2)a_i + \text{noise}_i$, where the signs of the deltas are likely to be negative and positive, respectively, leading to an indeterminate correlation sign. Hence, it is not obvious whether parents pay more attention to children with higher or lower ability. Moreover, the function F is unlikely to be linear since, *e.g.*, ability may increase the return to motivation and parent support and there may be nonlinear returns to motivation and/or ability.

As a result, not only is the OLS estimate likely to be unreliable, but it is also difficult to figure out the direction of the bias. This motivates the use of alternative methods that can account for the presence of unobservables.

There is evidence that people, teenagers in particular, form friendship links based on intelligence. Some studies (Clark and Ayers, 1992; Burgess et al., 2011; Boutwell, Meldrum and Petkovsek, 2017) have documented homophilic matching on various measures of intelligence among teenagers. According to Boutwell, Meldrum and Petkovsek (2017), "preadolescent friendship dyads are robustly correlated on measures of general intelligence".

It is thus plausible that friendship ties[18] account for ability, suggesting an avenue for correcting the ability bias through the estimator developed in this paper. For each individual, I use friends as a comparison group[19] and compute $\widehat{\text{ATE}}$.

According to OLS, the treatment has a small effect of 0.04. The estimator developed in the paper, however, suggests a much higher effect of 0.17. Because of

[17] Grades may be affected directly or indirectly through increased effort. The target is the total effect that includes the mediated effect, conditioning on intrinsic motivation and ability.

[18] There is also some evidence for homophily in some of the covariates. The ratio of the average distance in age and gender among friends to the average distance between any two individuals is about one fourth and one half, respectively.

[19] The number of people with no reported friend is relatively high, but most of these students are also missing covariates. This suggests zero friend counts are more indicative of a missing data problem than of a general asocial behavior. Consequently, I restrict the sample to the sample used for OLS, which also facilitates comparability. The effective sample size is then 2777 and people have an average number of about 5 friends, as calibrated for the simulation study. This relatively low number is consistent with situations in which people report close friends, which is generally the preferred target for the type of matching exercises that the method exploits.

the control for unobserved ability and its potential nonlinear interactions, the latter estimate may be more reasonable.

Table 1: Estimates and standard errors

	Estimated ATE	standard error
$\widehat{\text{ATE}}$	0.17	0.05
$\hat{\beta}_{OLS}$	0.04	0.01

Remark 1: running a regression without the parent's employment status and the poverty score increases the treatment effect estimate from OLS to 0.11. Going back to the model $s_i = \delta_1 a_i + \delta_2 \text{parents' ability}_i + \text{noise}_i$, this may indicate that OLS may be *more* biased upon controlling for parent's employment and poverty because these variables may act as proxies for parent ability and thus increase the conditional correlation between ability and parent support.

Remark 2: if one believes that the importance of study score accurately reflects motivation level, then they may want to properly control on this variable. A simple way[20] to address this concern is to estimate the effect on high and low score (dichotomizing at the mean of 0) students and then aggregate. This results in a slightly lower average treatment effect estimate of $0.21\ 1447/2777 + 0.07\ 1333/2777 = 0.15$.

Remark 3: Switching to $M = 2$ or using people with at least two friends in common as a comparison group delivers similar results: the effect is estimated to be 0.20 or 0.16. Because these versions of the estimator rely on comparison groups with different relationships to the individual, this can alleviate concerns that the effect is driven by other factors.

References

Auerbach, Eric. 2022. "Identification and estimation of a partially linear regression model using network data." *Econometrica*, 90(1): 347–365.

[20] If one wants to incorporate all controls, a possibility is to run a regression of y_i on the treatment and relevant controls for each i using observations in $\mathcal{C}_i$, then to form an optimal weighted average of the treatment effect estimates. Using all controls included in the OLS regression, this delivers an even higher treatment effect estimate of 0.31, though less precisely estimated (standard error 0.16)

Bekker, Paul A. 1994. "Alternative approximations to the distributions of instrumental variable estimators." *Econometrica: Journal of the Econometric Society*, 657–681.

Bender, Edward A, E Rodney Canfield, and Brendan D McKay. 1990. "The asymptotic number of labeled connected graphs with a given number of vertices and edges." *Random Structures & Algorithms*, 1(2): 127–169.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. "Quasi-experimental shift-share research designs." *The Review of economic studies*, 89(1): 181–213.

Boucher, Vincent. 2015. "Structural homophily." *International Economic Review*, 56(1): 235–264.

Boucher, Vincent, and Ismael Mourifié. 2017. "My friend far, far away: a random field approach to exponential random graph models." *The econometrics journal*, 20(3): S14–S46.

Boutwell, Brian B, Ryan C Meldrum, and Melissa A Petkovsek. 2017. "General intelligence in friendship selection: A study of preadolescent best friend dyads." *Intelligence*, 64: 30–35.

Burgess, Simon, Eleanor Sanderson, Marcela Umaña-Aponte, et al. 2011. *School ties: An analysis of homophily in an adolescent friendship network.* Centre for Market and Public Organisation.

Case, Anne C, Harvey S Rosen, and James R Hines Jr. 1993. "Budget spillovers and fiscal policy interdependence: Evidence from the states." *Journal of public economics*, 52(3): 285–307.

Clark, ML, and Marla Ayers. 1992. "Friendship similarity during early adolescence: Gender and racial patterns." *The journal of psychology*, 126(4): 393–405.

Currarini, Sergio, Matthew O Jackson, and Paolo Pin. 2009. "An economic model of friendship: Homophily, minorities, and segregation." *Econometrica*, 77(4): 1003–1045.

Demirer, Mert. 2019. "Partial Identification of Linear Models Using Homophily in Network Data." *Working paper.*

De Paula, Aureo. 2017. “Econometrics of network models.” Vol. 1, 268–323, Cambridge University Press Cambridge.

Dzemski, Andreas. 2019. “An empirical model of dyadic link formation in a network with unobserved heterogeneity.” *Review of Economics and Statistics*, 101(5): 763–776.

Erdős, Paul, Alfréd Rényi, et al. 1960. “On the evolution of random graphs.” *Publ. math. inst. hung. acad. sci*, 5(1): 17–60.

Gao, Wayne Yuan. 2020. “Nonparametric identification in index models of link formation.” *Journal of econometrics*, 215(2): 399–413.

Goldsmith-Pinkham, Paul, and Guido W Imbens. 2013. “Social networks and the identification of peer effects.” *Journal of Business & Economic Statistics*, 31(3): 253–264.

Graham, Bryan S. 2015. “Methods of identification in social networks.” *Annu. Rev. Econ.*, 7(1): 465–485.

Graham, Bryan S. 2016. “Homophily and transitivity in dynamic network formation.” National Bureau of Economic Research.

Graham, Bryan S. 2017. “An econometric model of network formation with degree heterogeneity.” *Econometrica*, 85(4): 1033–1063.

Hahn, Jinyong. 1998. “On the role of the propensity score in efficient semiparametric estimation of average treatment effects.” *Econometrica*, 315–331.

Hirano, Keisuke, Guido W Imbens, and Geert Ridder. 2003. “Efficient estimation of average treatment effects using the estimated propensity score.” *Econometrica*, 71(4): 1161–1189.

Hsieh, Chih-Sheng, and Lung Fei Lee. 2016. “A social interactions model with endogenous friendship formation and selectivity.” *Journal of Applied Econometrics*, 31(2): 301–319.

Imbens, Guido W. 2004. “Nonparametric estimation of average treatment effects under exogeneity: A review.” *Review of Economics and statistics*, 86(1): 4–29.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Imbens, Guido W, and Jeffrey M Wooldridge. 2009. "Recent developments in the econometrics of program evaluation." *Journal of economic literature*, 47(1): 5–86.

Jackson, Matthew O. 2010. *Social and economic networks.* Princeton university press.

Johnsson, Ida, and Hyungsik Roger Moon. 2021. "Estimation of peer effects in endogenous social networks: Control function approach." *Review of Economics and Statistics*, 103(2): 328–345.

Lazarsfeld, Paul F, Robert K Merton, et al. 1954. "Friendship as a social process: A substantive and methodological analysis." *Freedom and control in modern society*, 18(1): 18–66.

Lin, Zhexiao, and Fang Han. 2025. "On regression-adjusted imputation estimators of average treatment effects." *Journal of Econometrics*, 251: 106080.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual review of sociology*, 27(1): 415–444.

Mele, Angelo. 2022. "A structural model of homophily and clustering in social networks." *Journal of Business & Economic Statistics*, 40(3): 1377–1389.

Moody, James. 2001. "Race, school integration, and friendship segregation in America." *American journal of Sociology*, 107(3): 679–716.

Newman, Mark. 2018. *Networks.* Oxford university press.

Neyman, Jerzy. 1923. "On the application of probability theory to agricultural experiments. Essay on principles." *Ann. Agricultural Sciences*, 1–51.

Portella, Alysson, and Charles Kirschbaum. 2022. "Replication Data for: Racial Social Norms among Brazilian Students: Academic Performance, Social Status and Racial Identification."

Rubin, Donald B. 1974. “Estimating causal effects of treatments in randomized and nonrandomized studies.” *Journal of educational Psychology*, 66(5): 688.

Zeleneev, Andrei. 2020. “Identification and estimation of network models with nonparametric unobserved heterogeneity.” *Department of Economics, Princeton University.*

Appendix A: Proofs

4.1 Lemma 2.1

Proof. Consider a sample of $(n+1)$ observations, including i. By independence, the degree of individual i satisfies $d_i \overset{\text{def}}{=} |\mathcal{N}(i)|= \sum_{j\neq i,j=1}^{n+1} \mathbb{1}_{\eta_{ij}\leq w_n(\|x_i - X_j\|)} \sim \mathcal{B}(n,\pi_n)$ with $\pi_n \overset{\text{def}}{=} \mathbb{E}[w_n(\|x_i - X_j\|)] = \int_{\mathbb{R}^d} w_n(\|x_j - x_i\|)f(x_j)\, dx_j$ by the law of iterated expectation and distributional properties of η. If $n\pi_n \to C > 0$, the friend count asymptotically follows a Poisson distribution with parameter C, while slower sequences π_n induce an unbounded friend count: by Chebyshev inequality we have for any N and large sufficiently large n

$$
\begin{aligned}
\mathbb{P}[d_i \leq N] &\leq \mathbb{P}[|d_i - n\pi_n| \geq n\pi_n - N] \\
&\leq \frac{n\pi_n(1-\pi_n)}{(n\pi_n - N)^2} \\
&= \frac{1-\pi_n}{n\pi_n(1-\frac{N}{n\pi_n})^2}
\end{aligned}
$$

and thus $\mathbb{P}[d_i \leq N] \to 0$ as long as $n\pi_n = n\int_{\mathbb{R}^d} w_n(\|x_j - x_i\|)f(x_j)\, dx_j \to \infty$.

This is the case when the network formation is regularly asymptotically homophilic. Indeed, consider

$$\begin{aligned}\int_{\mathbb{R}^d} w_n(\|x-x_i\|)f(x)\ dx &= \int_{\mathbb{R}^d} g(s_n\|x-x_i\|)f(x)\ dx \\ &\geq \underline{f}\int_{B_{\underline{r}}(x_i)} g(s_n\|x-x_i\|)\ dx \\ &\geq \frac{\underline{f}}{{s_n}^d}\int_{B_{s_n\underline{r}}(0)} g(\|y\|)\ dy \\ &\geq \frac{\underline{f}}{{s_n}^d}\int_{B_{\underline{r}}(0)} g(\|y\|)\ dy\end{aligned}$$

for sufficiently large n. Since $\int_{\mathbb{R}^d} g(\|y\|)\ dy > 0$ and g is decreasing, $\int_{B_{\underline{r}}(0)} g(\|y\|)\ dy > 0$ (otherwise, $\int_{B_{\underline{r}s_n}(0)} g(\|y\|)\ dy = 0$ implies $g \equiv 0$, contradicting the former assertion). Thus, *regular asymptotic homophily* suffices to establish $\lim_{n\to\infty} n\int_{\mathbb{R}^d} w_n(\|x_j - x_i\|)f(x_j)\ dx_j = \infty$.

For part a), write $w_n(\|x_i - x_j\|)$ as w_{ij},

Since the covariate density is continuous (and then bounded so that the dominated convergence theorem applies below with dominating function $Cg(\|y\|)$),

$$\begin{aligned}\mathbb{E}[|\mathcal{N}_i^m|] &= \frac{n!}{(n-m)!}\int\prod_{k=0}^{m-1} g(s_n\|x_{j_k} - x_{j_{k+1}}\|)\prod_{k=1}^{m} f(x_{j_k})\prod_{l>k+1}(1-w_{j_kj_l})\ d\left(\prod_{k=1}^{m} x_{j_k}\right) \\ &= \frac{n!}{(n-m)!}{s_n}^{-d}\int\prod_{k=1}^{m} g(\|y_k\|)\prod_{k=1}^{m} f\left(x_i + \frac{y_k + y_{k-1} + \dots}{s_n}\right)\prod_{l>k+1}(1-w_{kl})\ d\left(\prod_{k=1}^{m} y_k\right) \\ &= \frac{n!}{(n-m)!}\frac{1}{{s_n}^d}O(1)\end{aligned}$$

which establishes that $\mathbb{E}[|\mathcal{C}_i|] = O\left(\frac{n!}{(n-m)!{s_n}^d}\right)$, and thus $|\mathcal{C}_i|= O((n{s_n}^{-d})^M)$ by Markov inequality.

The second part of the statement follows directly from $|\cup_{m=1}^{M}\mathcal{N}^m(i)|\geq d_i > N$ for any N with probability approaching one.

Finally, for part b), the number of friends in common between i and j follows a binomial distribution with parameters $(n-2,\pi_c)$. The probability π_c is of order ${s_n}^{-2d}$ and then the order of the comparison group is $n(n{s_n}^{-2d})^c$ while the probability is $(n{s_n}^{-2d})^c$.

□

4.2 Theorem 2.1

Proof. **Part a)** Lemma 2.1 ensures asymptotics apply. By the overlap assumption, the number of treated friends and untreated friends both grow to infinity.

Consider the treated group. Letting $\tilde{Y}_{j;n}$ denote i.i.d. draws from $Y(1)|\mathcal{C}_{i1}$, we have $\sup_n \mathbb{E}[(\tilde{Y}_{j;n})^2] < \infty$ since $\mathbb{E}[Y_j(1)^2|\mathcal{C}_{i1}] \to \mathbb{E}[Y_j(1)^2|X_j = x_i, T_j = 1] = \mathbb{E}[Y_j(1)^2|X_j = x_i]$ by continuity (this is shown in details for $Y_j(1)$ below) and unconfoundedness. Then, $\frac{1}{|\mathcal{C}_{i1}|}\sum_{j\in\mathcal{C}_{i1}}(Y_j(T_j) - \mathbb{E}[Y_j(1)|\mathcal{C}_{i1}]) + \mathbb{E}[Y_j(1)|\mathcal{C}_{i1}] - \mathbb{E}[Y_j(1)|X_j = x_i] \xrightarrow{p} 0$ by the law of large numbers for triangular arrays and continuity, which is formally established by bounding $\mathbb{E}[Y_j(1)|\mathcal{C}_{i1}] - \mathbb{E}[Y_j(1)|X_j = x_i]$ as follows.

Divide the expectation into an integral over a ε-ball centered at x_i and an integral over its complement. Within the ball,

$$\begin{aligned}
&\left|\int_{B_\varepsilon(x_i)} (\mathbb{E}[Y_j(1)|X = x] - \mathbb{E}[Y_j(1)|X = x_i])f_{X|\mathcal{C}_i,T}\ dx\right| \\
&\leq \sup_{B_\varepsilon}|\mathbb{E}[Y_j(1)|X = x] - \mathbb{E}[Y_j(1)|X = x_i]| \\
&= O(\varepsilon^\alpha)
\end{aligned}$$

by Hölder continuity.

The integral outside the ball,
$\int_{B^c_\varepsilon(x_i)}(\mathbb{E}[Y_j(1)|X = x] - \mathbb{E}[Y_j(1)|X = x_i])f_{X|\mathcal{C}_i,T}\ dx$, can be bounded using information about the composition of $\mathcal{C}_i$. I proceed with the two groups separately.
For the first choice of comparison group, *i.e.*, $\mathcal{C}_i = \cup_{m=1}^M \mathcal{N}^m(i)$, note that $\varepsilon < \|x_j - x_i\| = \left\|\sum_{k=0}^{m-1} x_{j_{k+1}} - x_{j_k}\right\| \leq \sum_{k=0}^{m-1}\|x_{j_{k+1}} - x_{j_k}\|$ and thus outside the ball

$$\begin{aligned}
f_{\mathcal{C}_i|X_j,T_j} &= \sum_{m=1}^{M} \mathbb{P}[W_{ij}^m = 1, W_{ij}^{m'} = 0 \; m' < m|X,T] \\
&= \sum_{m=1}^{M} \mathbb{E}[\mathbb{P}[W_{ij}^m = 1, W_{ij}^{m'} = 0 \; m' < m|\{X_{j_k}\},T]|X_j,T_j] \\
&\leq \sum_{m=1}^{M} \int \prod_{k=0}^{m-1} w_{j_{k+1}j_k} \prod_{l\neq k}(1 - w_{j_k j_l}) f_{X_{-j}|X_j} \\
&\leq \sum_{m=1}^{M} w_n \left(\frac{\varepsilon}{m}\right)^m
\end{aligned}$$

It follows that

$$\begin{aligned}
&\left|\int_{B_\varepsilon^c(x_i)} (\mathbb{E}[Y_j(1)|X=x] - \mathbb{E}[Y_j(1)|X=x_i]) f_{X|\mathcal{C}_i,T} \; dx\right| \\
&\leq \int_{B_\varepsilon^c(x_i)} |\mathbb{E}[Y_j(1)|X=x] - \mathbb{E}[Y_j(1)|X=x_i]| \, \frac{\mathbb{P}[\mathcal{C}_i|X=x] f_{X,T}(x,1)}{\mathbb{P}[\mathcal{C}_i, T=1]} \; dx \\
&\leq \frac{\sum_{m=1}^{M} \left(w_n(\frac{\varepsilon}{m})\right)^m (\mathbb{E}[|\mathbb{E}[Y_j(1)|X]|] + \mathbb{E}[Y_j(1)|X=x_i])}{\left(\sum_{m=1}^{M} \sum_{j_0,\cdots,j_m} \mathbb{P}[W_{j_k j_{k+1}} = 1, W_{j_k j_l} = 0 \; l \neq k+1]\right)} \\
&\leq C \frac{n}{\lambda_n M} \sum_{m=1}^{M} \left(w_n\left(\frac{\varepsilon}{m}\right)\right)^m
\end{aligned}$$

Hence, the integrals are $O({\varepsilon_n}^\alpha)$ and $O(\frac{n}{\lambda_n M}\sum_{m=1}^{M}(1 - w_n(\frac{\varepsilon_n}{m}))^m)$, respectively, and the bias disappears provided $\sum_{m=1}^{M}(1 - w_n(\frac{\varepsilon_n}{m}))^m = o(\lambda_n \frac{M}{n})$. Applying the same reasoning to the non-treated gives the result for the first choice of comparison group.

For the second comparison group, the conditional distribution (leaving the conditioning on $T = t$ implicit) satisfies

$$\begin{aligned}
f_{X_j|\mathcal{C}_i} &= \frac{\mathbb{P}[\mathcal{C}_i|X_j]f(x)}{\int \mathbb{P}[\mathcal{C}_i|X_j]f(x)\ dx} \\
&= \frac{\mathbb{E}[\mathbb{P}[W_{i1}=W_{j1}=\cdots=W_{im}=W_{jm}=1|X_1,\cdots,X_m,X_j]]f(x)}{\int \mathbb{E}[\mathbb{P}[W_{i1}=W_{j1}=\cdots=W_{im}=W_{jm}=1|X_1,\cdots,X_m,X_j]]f(x)\ dx} \\
&= \frac{\prod_{c=1}^{m} \int w_n(\|x_i-x_k\|)w_n(\|x_j-x_k\|)f(x_k)dx_k f(x)}{\prod_{c=1}^{m} \int w_n(\|x_i-x_k\|)w_n(\|x_j-x_k\|)f(x_k)dx_k f(x)\ dx} \\
&\leq C{s_n}^d \sum_{m=c}^{n-2} \left(w_n\left(\frac{\varepsilon_n}{2}\right)\right)^m \\
&= C{s_n}^d \left(w_n\left(\frac{\varepsilon_n}{2}\right)\right)^c\ O(1)
\end{aligned}$$

so that we want $\left(w_n(\frac{\varepsilon_n}{2})\right)^c = o({s_n}^{-d})$

Part b) Consider the case $M=1$ for simplicity. One can compute

$$\begin{aligned}
\mathbb{E}[Y_j(t)|\mathcal{C}_i, T_j=t] &= \mathbb{E}[\mathbb{E}[Y_j(t)|X_j]|\mathcal{C}_i, T_j=t] \\
&= \int_{\mathbb{R}^d} \mathbb{E}[Y_j(t)|x_j] f_{x_j|\mathcal{C}_i,x_i,t} dx_j \\
&= \int_{\mathbb{R}^d} \mathbb{E}[Y_j(t)|x_j] f_{x_j|\mathcal{C}_i,x_i,t} dx_j \\
&= \int_{\mathbb{R}^d} \mathbb{E}[Y_j(t)|x_j] \frac{w_{ij}p_t(x_j)f_{x_j}}{\int_{\mathbb{R}^d} w_{ik}p_t(x_k)f_{x_k}dx_k} dx_j \\
&= \int_{\mathbb{R}^d} \mathbb{E}[Y_j(t)|x_i+y/s_n] \frac{g(\|y\|)p_t(x_i+y/s_n)f_{x_i+y/s_n}}{\int_{\mathbb{R}^d} g(\|z\|)p_t(x_i-z/s_n)f_{x_i-z/s_n}dz} dy \\
&= \mathbb{E}[Y_j(t)|x_i] \int_{\mathbb{R}^d} \frac{g(\|y\|)}{\int_{\mathbb{R}^d} g(\|z\|)dz} dy + O({s_n}^{-1}) \int_{\mathbb{R}^d} g(\|y\|)ydy + O({s_n}^{-2}) \\
&= \mathbb{E}[Y_j(t)|x_i] + O({s_n}^{-2})
\end{aligned}$$

using the changes of variable $y = s_n(x_j - x_i)$ and $z = s_n(x_k - x_i)$ and a first-order expansion in densities, expectation, and propensity scores (noting that $\frac{A+a_n}{B+b_n} = \frac{A}{B}\left(1+\frac{a_n}{A}\right)\left(1-\frac{b_n}{B+b_n}\right) = \frac{A}{B} + O(a_n) + O(b_n) + O(a_n b_n)$) and that second-order derivatives are bounded.

□

4.3 Theorem 2.2

Proof. By the Law of Large Numbers,

$$\left(\frac{|\mathcal{C}_{it}|}{n\mathbb{P}[j \in \mathcal{C}_{it}]} - 1\right) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\mathbb{1}(j \in \mathcal{C}_{it})}{\mathbb{P}[j \in \mathcal{C}_{it}]} - 1\right) \xrightarrow{p} 0 \tag{6}$$

for $t \in \{0, 1\}$. Next, by Lindeberg-Feller's central limit theorem where Lindeberg's condition follows from the dominated convergence theorem under $2 + \delta$-moment existence,

$$\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\mathbb{1}(j\in\mathcal{C}_{i1})}{\sqrt{\mathbb{P}[j\in\mathcal{C}_{i1}]}}\left(Y_j - \mathbb{E}[Y_j(1)|\mathcal{C}_i, T_j = 1]\right) \\ \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\mathbb{1}(j\in\mathcal{C}_{i0})}{\sqrt{\mathbb{P}[j\in\mathcal{C}_{i0}]}}\left(Y_j - \mathbb{E}[Y_j(1)|\mathcal{C}_i, T_j = 1]\right) \end{pmatrix}$$
$$\xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \mathbb{V}[Y_j(1)|X_j = x_i] & 0 \\ 0 & \mathbb{V}[Y_j(0)|X_j = x_i] \end{pmatrix}\right)$$

Combining these two results and noting that $|\mathcal{C}_i| = |\mathcal{C}_{i1}| + |\mathcal{C}_{i0}|$, it follows that

$$\sqrt{|\mathcal{C}_i|}\left(\widehat{\text{CATE}}(x_i; \mathcal{C}_i) - \text{CATE}(x_i) - \mathbb{B}_i\right) \xrightarrow{d} \mathcal{N}(0; V) \tag{7}$$

Finally, the bias is negligible when $\sqrt{n}(\mathbb{E}[Y_j(t)|\mathcal{C}_i, T_j = t] - \mathbb{E}[Y_j(t)|X_j = x_i]) \to 0$. From the consistency proof, it is seen that this occurs if $\varepsilon_n = n^{-\gamma}$ under the conditions of the theorem. □

4.4 Theorem 2.3

The main preliminary result, which establishes a variant of Assumption 3.3 (ii) in Lin and Han (2025), is the mean-square convergence of the weights to inverse-propensity score quantities:

Lemma 4.1. *Let $\omega_{ijt} \overset{\text{def}}{=} \mathbb{1}(j \in \mathcal{C}_i)\frac{1}{|\mathcal{C}_{jt}|}$ for $t \in \{0, 1\}$. We have*

$$\mathbb{E}\left[\left(\sum_j \omega_{ijT_i} - \frac{T_i}{p(X_i)} - \frac{1 - T_i}{1 - p(X_i)}\right)^2\right] \to 0 \tag{8}$$

Proof. The proof follows steps similar to Lemma B.3. in Lin and Han (2025). In what follows, $\boldsymbol{T}$ denotes the vector of treatment statuses and $\boldsymbol{X}$ the collection of covariates, f_t is the density of X conditional on treatment status $t \in \{0,1\}$, and it is understood that $j \neq i$ when conditioning on both X_i and X_j.

First, decompose

$$\mathbb{E}\left[\left(\sum_j \omega_{ij1} - \frac{T_i}{p(X_i)} - \frac{1-T_i}{1-p(X_i)}\right)^2\right] = \mathbb{E}\left[\mathbb{E}\left[\left(\sum_j \omega_{ij} - \frac{1}{p(X_i)}\right)^2 \middle| T, T_i = 1\right] T_i\right] + \mathbb{E}\left[\mathbb{E}\left[\left(\sum_j \omega_{ij0} - \frac{1}{1-p(X_i)}\right)^2 \middle| T, T_i = 1\right](1-T_i)\right]$$

The two terms can be handled analogously, so it suffices to consider the case where i is treated. The corresponding term can be decomposed further into

$$\sum_j \omega_{ij} - \frac{1}{p(X_i)} = \frac{n(R_4+R_3+R_2)}{n_1} + \frac{n}{n_1}\frac{f(X_i)}{f_1(X_i)} - \frac{1}{p(X_i)} \tag{9}$$

where

$$R_2 \overset{\mathsf{def}}{=} f_1^{-1}(X_i)\left[\frac{1}{n}\sum_j \frac{\mathbb{1}(j \in \mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y} - f(X_i)\right] \tag{10}$$

$$R_3 \overset{\mathsf{def}}{=} \frac{1}{n}\sum_j (f_1^{-1}(X_j) - f_1^{-1}(X_i))\,\frac{\mathbb{1}(j \in \mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y} \tag{11}$$

$$R_4 \overset{\mathsf{def}}{=} \frac{1}{n}\sum_j \left[\left(\frac{1}{n_1}\sum_{k:T_k=1}\frac{\mathbb{1}(k \in \mathcal{C}_j){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right)^{-1} - f_1^{-1}(X_j)\right]\frac{\mathbb{1}(j \in \mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y} \tag{12}$$

From the law of large numbers,

$$\mathbb{E}\left[\mathbb{E}\left[\left(\frac{n}{n_1}\frac{f(X_i)}{f_1(X_i)} - \frac{1}{p(X_i)}\right)^2 \middle| \boldsymbol{T}, T_i = 1\right] T_i\right] \to 0 \tag{13}$$

So it remains to show that

$$\mathbb{E}\left[\mathbb{E}\left[\left(\frac{nR_m}{n_1}\right)^2 \middle| \boldsymbol{T}, T_i = 1\right] T_i\right] \to 0 \tag{14}$$

for $m = 2, 3, 4$.

R_2

Using iterated expectations to condition on X_i, the second moment can be decomposed into the bias

$$
\begin{aligned}
&\mathbb{E}\left[\left\{\mathbb{E}\left[\frac{\mathbb{1}(j \in \mathcal{C}_i){s_n}^d)}{\int g(\|y\|)\ \mathrm{d}y}\middle| X_i, T_i = 1\right] - f(X_i)\right\}^2\right] \\
&= \mathbb{E}\left[\left\{\mathbb{E}\left[\mathbb{E}\left[\frac{\mathbb{1}(j \in \mathcal{C}_i){s_n}^d)}{\int g(\|y\|)\ \mathrm{d}y}\middle| X_i, X_j, T_i = 1\right]\middle| X_i, T_i = 1\right] - f(X_i)\right\}^2\right] \\
&= \mathbb{E}\left[\left\{\int \frac{g(s_n\|x_j - X_i\|)}{{s_n}^{-d}\int g(\|y\|)\ \mathrm{d}y} f(x_j)\ \mathrm{d}x_j - f(X_i)\right\}^2\right] \\
&= \mathbb{E}\left[\left\{\int \frac{g(\|z\|)}{\int g(\|y\|)\ \mathrm{d}y}(f(X_i + z/s_n) - f(X_i))\ \mathrm{d}z\right\}^2\right] \\
&= O({s_n}^{-4d})
\end{aligned}
$$

and the variance

$$
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n}\mathbb{V}\left[\frac{\mathbb{1}(j \in \mathcal{C}_i){s_n}^d}{\int g(\|y\|)\ \mathrm{d}y}\middle| X_i, T_i = 1\right]\right] \\
&\leq \mathbb{E}\left[\frac{1}{n}{s_n}^{2d}\mathbb{E}\left[\left(\frac{\mathbb{1}(j \in \mathcal{C}_i)}{\int g(\|y\|)\mathrm{d}y}\right)^2\middle| X_i, T_i = 1\right]\right] \\
&\leq \frac{1}{n}{s_n}^{2d}\frac{O(1)}{{s_n}^d} \\
&\leq O((n{s_n}^{-d})^{-1})
\end{aligned}
$$

which implies

$$
\mathbb{E}\left[\mathbb{E}\left[\left(\frac{nR_2}{n_1}\right)^2\middle| T, T_i = 1\right] T_i\right] \to 0 \tag{15}
$$

R_3

After iterated expectations, the bias equals (the expectation of)

$$
\begin{aligned}
&\left\{\mathbb{E}\left[(f_1^{-1}(X_j) - f_1^{-1}(X_i))\,\frac{\mathbb{1}(j\in\mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\middle|X_i, T_i=1, T_j=1\right]\right\}^2 \\
&= \left\{\int (f_1^{-1}(x_j) - f_1^{-1}(X_i))\,\frac{g(s_n\|x_j - X_i\|){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\, f_1(x_j)\mathrm{d}x_j\right\}^2 \\
&= \iint (f_1^{-1}(X_i + z/s_n) - f_1^{-1}(X_i))(f_1^{-1}(X_i + \tilde{z}/s_n) - f_1^{-1}(X_i))\,\frac{g(\|z\|)g(\|\tilde{z}\|)}{(\int g(\|y\|)\,\mathrm{d}y)^2} \\
&\times f_1(X_i + z/s_n)f_1(X_i + \tilde{z}/s_n)\,\mathrm{d}z\mathrm{d}\tilde{z} \to 0
\end{aligned}
$$

while the variance term corresponds to

$$
\begin{aligned}
&\frac{1}{n}\mathbb{V}\left[(f_1^{-1}(X_j) - f_1^{-1}(X_i))\,\frac{\mathbb{1}(j\in\mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\middle|X_i, T_i=1, T_j=1\right] \\
&\le \frac{1}{n{s_n}^{-2d}}\mathbb{E}\left[\frac{(f_1^{-1}(X_j) - f_1^{-1}(X_i))^2\,\mathbb{1}(j\in\mathcal{C}_i)}{(\int g(\|y\|)\,\mathrm{d}y)^2}\right] \to 0
\end{aligned}
$$

R_4

R_4 can be rewritten as $\frac{1}{n}\sum_j f_1^{-2}(X_j)\left[f_1(X_j) - \frac{1}{n_1}\sum_{k\neq j, T_k=1}\frac{\mathbb{1}(k\in\mathcal{C}_j){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right]\frac{\mathbb{1}(j\in\mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}$, up to lower-order terms. Then, by Jensen's inequality and iterated expectations,

$$
\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{n}\sum_j f_1^{-2}(X_j)\left[f_1(X_j) - \frac{1}{n_1}\sum_{k\neq j, T_k=1}\frac{\mathbb{1}(k\in\mathcal{C}_j){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right]\frac{\mathbb{1}(j\in\mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right)^2\right] \\
&\le \mathbb{E}\left[\frac{1}{n}\sum_j\left(f_1^{-2}(X_j)\left[f_1(X_j) - \frac{1}{n_1}\sum_{k\neq j, T_k=1}\frac{\mathbb{1}(k\in\mathcal{C}_j){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right]\frac{\mathbb{1}(j\in\mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right)^2\right] \\
&\le \mathbb{E}\left[\left(f_1^{-2}(X_j)\left[f_1(X_j) - \frac{1}{n_1}\sum_{k\neq j, T_k=1}\frac{\mathbb{1}(k\in\mathcal{C}_j){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right]\frac{\mathbb{1}(j\in\mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right)^2\right] \\
&\le \mathbb{E}\left[\mathbb{E}\left[\left(f_1^{-2}(X_j)\left[f_1(X_j) - \frac{1}{n_1}\sum_{k\neq j, T_k=1}\frac{\mathbb{1}(k\in\mathcal{C}_j){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right]\frac{\mathbb{1}(j\in\mathcal{C}_i){s_n}^d}{\int g(\|y\|)\,\mathrm{d}y}\right)^2\middle|X_i, X_j\right]\right].
\end{aligned}
$$

Decomposing the expectation of the square into squared bias and variance, and proceeding as for R_2 and R_3 then implies $\mathbb{E}\left[\mathbb{E}\left[\left(\frac{nR_4}{n_1}\right)^2 \middle| \boldsymbol{T}, T_i = 1\right] T_i\right] \to 0.$ □

The proof follows arguments similar to those in Lin and Han (2025), who establish the asymptotic normality of ATE estimators that use kernel weights.

First, note that the ATE estimator can be written as

$$
\begin{aligned}
\widehat{\text{ATE}} &= \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\theta_i Y_i \\
&= \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\theta_i (Y_i - \mu_{T_i}(X_i) + \mu_{T_i}(X_i)) \\
&= \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\theta_i \varepsilon_i \\
&+ \frac{1}{n}\sum_{i,j,T_i=T_j}(2T_j - 1)\omega_{ijT_j}\mu_{T_j}(X_i) - \frac{1}{n}\sum_{i,j,T_i=1-T_j}(2T_j - 1)\omega_{ij(1-T_j)}\mu_{1-T_j}(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\theta_i \varepsilon_i \\
&+ \frac{1}{n}\sum_{i,j,T_i=T_j}(2T_i - 1)\omega_{jiT_i}\mu_{T_i}(X_j) - \frac{1}{n}\sum_{i,j,T_i=1-T_j}(2T_i - 1)\ \omega_{ji(1-T_i)}\ \mu_{1-T_i}(X_j) \\
&= \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\theta_i \varepsilon_i + \frac{1}{n}\sum_{i=1}^{n}\mu_1(X_i) - \mu_0(X_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\sum_j \omega_{jiT_i}\left(\mu_{T_i}(X_j) - \mu_{T_i}(X_i)\right) \\
&- \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\sum_j \omega_{ji(1-T_i)}\left(\mu_{1-T_i}(X_j) - \mu_{1-T_i}(X_i)\right) \\
&+ \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\mu_{T_i}(X_i)\left(\sum_j \omega_{jiT_i} - 1\right) - \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\mu_{1-T_i}(X_i)\left(\sum_j \omega_{ji(1-T_i)} - 1\right)
\end{aligned}
$$

The first term, $\frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)\theta_i\varepsilon_i + \frac{1}{n}\sum_{i=1}^{n}\mu_1(X_i) - \mu_0(X_i)$, is asymptotically normal while the other terms are $\sqrt{n}$-negligible.

Decompose

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\theta_i\varepsilon_i &= \frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\left(\frac{T_i}{p(X_i)}+\frac{1-T_i}{1-p(X_i)}\right)\varepsilon_i \\ &+\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\left(\theta_i-\frac{T_i}{p(X_i)}-\frac{1-T_i}{1-p(X_i)}\right)\varepsilon_i.\end{aligned}$$

Then,

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\left(\theta_i-\frac{T_i}{p(X_i)}-\frac{1-T_i}{1-p(X_i)}\right)\epsilon_i\,\middle|\,\boldsymbol{X},\boldsymbol{T}\right]=0,$$

$$\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\left(\theta_i-\frac{T_i}{p(X_i)}-\frac{1-T_i}{1-p(X_i)}\right)\epsilon_i\,\middle|\,\boldsymbol{X},\boldsymbol{T}\right]=\frac{1}{n^2}\sum_{i=1}^{n}\left(\theta_i-\frac{T_i}{p(X_i)}-\frac{1-T_i}{1-p(X_i)}\right)^2\sigma^2_{T_i}(X_i),$$

and

$$\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\left(\theta_i-\frac{T_i}{p(X_i)}-\frac{1-T_i}{1-p(X_i)}\right)\epsilon_i=o_P(n^{-1/2}).$$

Then,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\theta_i\varepsilon_i+\frac{1}{n}\sum_{i=1}^{n}\mu_1(X_i)-\mu_0(X_i)-\text{ATE} &= \frac{1}{n}\sum_{i=1}^{n}\left[\mu_1(X_i)-\mu_0(X_i)-\text{ATE}\right] \\ &+\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\left(\frac{T_i}{p(X_i)}+\frac{1-T_i}{1-p(X_i)}\right)\epsilon_i \\ &+o_P(n^{-1/2}).\end{aligned}$$

Hence, by the central limit theorem,

$$\begin{aligned}&\sqrt{n}\left(\mathbb{E}\left[\frac{\mathbb{V}[Y_i(1)|X_i]}{p(X_i)}+\frac{\mathbb{V}[Y_i(0)|X_i]}{1-p(X_i)}\right]+\mathbb{V}[\text{CATE}(X_i)]\right)^{-1/2} \\ &\times\left(\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\theta_i\varepsilon_i+\frac{1}{n}\sum_{i=1}^{n}\mu_1(X_i)-\mu_0(X_i)-\text{ATE}\right)\xrightarrow{d}\mathcal{N}(0,1)\end{aligned} \tag{16}$$

The remaining terms vanish at a lower rate. The following terms can be bounded as

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^{n}(2T_i-1)\sum_j \omega_{ij(1-T_i)}[\mu_{1-T_i}(X_i)-\mu_{1-T_i}(X_j)]\right| \\ &\leq \max_{t\in\{0,1\},\alpha\in\{1,\dots,d\}}\|\partial_\alpha \mu_t\|\frac{1}{n}\sum_{i=1}^{n}\sum_j \omega_{ijt}\|X_j-X_i\|\end{aligned} \tag{17}$$

which is $o_p(n^{-1/2})$ if $\sqrt{n}{s_n}^{-2}=o(1)$. Finally, the last two terms feature factors of the form $\sum_j \omega_{jit}-1$ and thus immediately drop out when the weights sum up to 1.[21]

4.5 Theorem 2.4

Proof. To ease notation, the argument κ is omitted in all instances of $\mathcal{C}_{it}(\kappa)$. I also make use of the following shorthands: $\Delta_{ij}^u \overset{\text{def}}{=} \|x_j^u - x_i^u\|$, $h_{ij} \overset{\text{def}}{=} h(X_j^o, X_i^o)$. Decompose the centered mean of the group as

$$\frac{1}{|\mathcal{C}_{it}|}\sum_{j\in\mathcal{C}_{it}} Y_j(T_j)-\mathbb{E}[Y(t)|x_i] = \frac{1}{|\mathcal{C}_{it}|}\sum_{j\in\mathcal{C}_{it}} Y_j(T_j)-\mathbb{E}[Y_j(t)|\mathcal{C}_{it}]+\mathbb{E}[Y_j(t)|\mathcal{C}_{it}]-\mathbb{E}[Y(t)|x_i]$$

for $t\in\{0,1\}$.

The first term depends on sample fluctuations and converges (in probability to 0 and in distribution once re-scaled) provided the number of observations in the sum grows to infinity. The second term is a bias term and disappears under regularity conditions and the truncation $h_{ij} > \kappa$ with $\kappa\to\infty$. Thus, the main steps prove that (i) the number of observations in the sum grows to infinity and (ii) the bias disappears, for some sequence $\kappa\to\infty$. Consider the probability of an observation

[21] If some regularization is used, those terms vanish as long as they converge to normalized weights fast enough, *e.g.*, in the case of a fixed offset in the denominators.

belonging to $\mathcal{C}_i$ first:

$$\begin{aligned}
\mathbb{P}[\mathcal{C}_i] &\geq C \int_{h_{ij}>\kappa} w_{ij} f(x_j)\ dx_j \\
&\geq C \int_{\Delta^u_{ij} \leq \frac{b_n}{2}, \kappa < h_{ij} < \kappa + \frac{b_n}{2}} w(\kappa + b_n) f(x_j)\ dx_j \\
&= C w(\kappa + b_n) \mathbb{P}\left[\Delta^u_{ij} \leq \frac{b_n}{2}, \kappa < h_{ij} < \kappa + \frac{b_n}{2}\right] \\
&= C w(\kappa + b_n) \mathbb{P}\left[\Delta^u_{ij} \leq \frac{b_n}{2}\right] \mathbb{P}\left[\kappa < h_{ij} < \kappa + \frac{b_n}{2} \middle| \Delta^u_{ij} \leq \frac{b_n}{2}\right] \\
&\geq C w(\kappa + b_n) {b_n}^{d_u + 1}
\end{aligned}$$

Following previous proofs, it suffices to let $\kappa \to \infty$ slowly enough to induce a rate of $\frac{\lambda_n}{n}$ for the above probability. Next, the bias term reads

$$\begin{aligned}
|\mathbb{E}[Y_j(T_j)|\mathcal{C}_{it}] - \mathbb{E}[Y_j(t)|x_i]| &= ||\mathbb{E}[\mathbb{E}[Y_j(T_j)|X_j]|\mathcal{C}_{it}]] - \mathbb{E}[Y_j(t)|x_i]|| \\
&= \left| \int_{\mathbb{R}^d} (\mathbb{E}[Y_j(t)|x_j] - \mathbb{E}[Y_j(t)|x_i]) f_{X_j|\mathcal{C}_{it}}(x_j)\ dx_j \right| \\
&\leq \left| \int_{B_\varepsilon(x_i)} (\mathbb{E}[Y_j(t)|x_j] - \mathbb{E}[Y_j(t)|x_i]) f_{X_j|\mathcal{C}_{it}}(x_j)\ dx_j \right| \\
&+ \left| \int_{B^c_\varepsilon(x_i)} (\mathbb{E}[Y_j(t)|x_j] - \mathbb{E}[Y_j(t)|x_i]) f_{X_j|\mathcal{C}_{it}}(x_j)\ dx_j \right| \\
&\leq C\varepsilon^\alpha \\
&+ \frac{1}{\mathbb{P}[\mathcal{C}_{it}]} \int_{B^c_\varepsilon(x_i)} (\mathbb{E}[Y_j(t)|x_j] - \mathbb{E}[Y_j(t)|x_i]) \mathbb{P}[\mathcal{C}_{it}|x_j] f_{X_j,T_j}(x_j, t)\ dx_j \\
&\leq C\varepsilon^\alpha \\
&+ \frac{C}{\mathbb{P}[\mathcal{C}_{it}]} \int_{B^c_\varepsilon(x_i), h_{ij}>\kappa} (\mathbb{E}[Y_j(t)|x_j] - \mathbb{E}[Y_j(t)|x_i])(1 - w_{ij}) f_{X_j,T_j}(x_j, t)\ dx_j \\
&\leq C\varepsilon^\alpha + \frac{Cn}{\lambda_n} w(\kappa + \varepsilon) \left(\mathbb{E}[|\mathbb{E}[Y_j(t)|X_j]||T_j = t] + |\mathbb{E}[Y_j(t)|x_i] \right)
\end{aligned}$$

so that, with $\varepsilon \downarrow 0$, the bias disappears as κ rises provided $w(\kappa + \varepsilon) = o(\frac{\lambda_n}{n})$.

□

Appendix B: Further results

4.6 Clustering

Since the covariate density is continuous (then bounded), the clustering coefficient is

$$
\begin{aligned}
C &\overset{\text{def}}{=} \mathbb{P}[W_{jk} = 1 | W_{ij} = W_{ik} = 1] \\
&= \mathbb{E}[\mathbb{P}[W_{jk} = 1 | W_{ij} = W_{ik} = 1, X_j, X_k] | W_{ij} = W_{ik} = 1] \\
&= \mathbb{E}[\mathbb{P}[W_{jk} = 1 | X_j, X_k] | W_{ij} = W_{ik} = 1] \\
&= \mathbb{E}[g(s_n \| x_j - x_k \|) | W_{ij} = W_{ik} = 1] \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(s_n \| x_j - x_k \|) f_{X_j, X_k | W_{ij}=1, W_{ik}=1}(x_j, x_k) \; dx_j dx_k \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(s_n \| x_j - x_k \|) \frac{\mathbb{P}[W_{ij} = 1 = W_{ik} | X_j = x_j, X_k = x_k] f(x_k) f(x_j)}{\mathbb{P}[W_{ij} = 1 = W_{ik}]} \; dx_j dx_k \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(s_n \| x_j - x_k \|) \frac{\mathbb{E}[\mathbb{P}[W_{ij} = 1 = W_{ik} | X_i = x_i, X_j = x_j, X_k = x_k] | X_j = x_j, X_k = x_k]}{\mathbb{E}[\mathbb{P}[W_{ij} = 1 = W_{ik} | X_i = x_i, X_j = x_j, X_k = x_k]]} \\
& f(x_k) f(x_j) \; dx_j dx_k \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{g(s_n \| x_j - x_k \|) \int_{\mathbb{R}^d} g(s_n \| x_i - x_j \|) g(s_n \| x_i - x_k \|) f(x_i) \; dx_i}{\int_{\mathbb{R}^{3d}} g(s_n \| x_i - x_k \|) g(s_n \| x_i - x_j \|) f(x_i) f(x_j) f(x_k) \; dx_i dx_j dx_k} \; f(x_j) f(x_k) \; dx_j dx_k \\
&= \frac{{s_n}^{-3d}}{{s_n}^{-3d}} \int_{\mathbb{R}^{3d}} \frac{g(\| \hat{x}_j - \hat{x}_k \|) g(\| \hat{x}_i - \hat{x}_j \|) g(\| \hat{x}_i - \hat{x}_k \|) f\left(\frac{\hat{x}_i}{\sqrt{s_n}}\right) f\left(\frac{\hat{x}_j}{\sqrt{s_n}} r\right) f\left(\frac{\hat{x}_k}{\sqrt{s_n}}\right)}{\int_{\mathbb{R}^{3d}} g(\| \hat{x}_i - \hat{x}_k \|) g(\| \hat{x}_i - \hat{x}_j \|) f\left(\frac{\hat{x}_i}{\sqrt{s_n}}\right) f\left(\frac{\hat{x}_j}{\sqrt{s_n}} r\right) f\left(\frac{\hat{x}_k}{\sqrt{s_n}}\right) \; d\hat{x}_i d\hat{x}_j d\hat{x}_k} \; d\hat{x}_i d\hat{x}_j d\hat{x}_k \\
&\to \frac{\int_{\mathbb{R}^{3d}} g(\| \hat{x}_j - \hat{x}_k \|) g(\| \hat{x}_i - \hat{x}_j \|) g(\| \hat{x}_i - \hat{x}_k \|) d\hat{x}_i d\hat{x}_j d\hat{x}_k}{\int_{\mathbb{R}^{3d}} g(\| \hat{x}_i - \hat{x}_j \|) g(\| \hat{x}_i - \hat{x}_k \|) \; d\hat{x}_i d\hat{x}_j d\hat{x}_k}
\end{aligned}
$$

where a hat represent a change of variable such as $\hat{x} = s_n x$.

4.7 Regression

Consider the *asymptotic homophily* framework with a linear outcome equation:

$$y_i = \text{CATE}(x_i) T_i + x_i' \beta + z_i' \gamma + \varepsilon_i \tag{18}$$

where z are additional controls, *i.e.*, variables that do not influence network formation, that include an intercept. The OLS estimator within $\mathcal{C}_i$ corresponds to

$$\hat{C}_i = \left(\frac{1}{|\mathcal{C}_i|}\sum_{j\in\mathcal{C}_i}\ddot{T}_j^2\right)^{-1}\frac{1}{|\mathcal{C}_i|}\sum_{j\in\mathcal{C}_i}\ddot{T}_j y_j \tag{19}$$

where $\ddot{T}_i$ are the residuals from a regression of T_i on z_i.

Expanding y_i according to (18) gives

$$\left(\frac{1}{|\mathcal{C}_i|}\sum_{j\in\mathcal{C}_i}\ddot{T}_i^2\right)^{-1}\frac{1}{|\mathcal{C}_i|}\sum_{j\in\mathcal{C}_i}\ddot{T}_j\,\text{CATE}(x_j) + \left(\frac{1}{|\mathcal{C}_i|}\sum_{j\in\mathcal{C}_i}\ddot{T}_j^2\right)^{-1}\frac{1}{|\mathcal{C}_i|}\sum_{j\in\mathcal{C}_i}\ddot{T}_j(x_j'\beta+\varepsilon_j) \tag{20}$$

Under asymptotic homophily, the difference between the first term and $\text{CATE}(x_i)$ is $o_p(1)$ while the second term approaches $x_i'\beta\mathbb{E}[\ddot{T}_i|x_i] + \mathbb{E}[\ddot{T}_i\varepsilon_i] = 0$. Furthermore, under regularity conditions and standard arguments, $\sqrt{n}(\widehat{\text{ATE}} - \text{ATE}) \xrightarrow{d} \mathcal{N}(0, V)$, provided second moments exist and $\sqrt{n}/{s_n}^2 \to 0$.

Appendix C: Simulation Results

Table 2: RMSE of ATE estimators (n=500)

y	β_3	M=1	M=2	M=3	M=4	c=2	c=3	c=4	OLS	Strat	IPW
	0	0.24	0.25	0.24	0.24	0.23	0.23	0.38	0.14	0.21	0.43
A	0.5	0.23	0.23	0.22	0.24	0.22	0.26	0.43	0.15	0.23	0.44
	1	0.24	0.23	0.23	0.24	0.22	0.26	0.43	0.21	0.29	0.36
	0	0.25	0.26	0.25	0.26	0.24	0.22	0.38	0.39	0.47	0.48
B	0.5	0.24	0.25	0.25	0.28	0.23	0.23	0.40	0.47	0.54	0.55
	1	0.26	0.26	0.26	0.28	0.24	0.25	0.41	0.48	0.57	0.52
	0	0.26	0.29	0.28	0.30	0.25	0.21	0.37	0.68	0.76	0.70
C	0.5	0.25	0.28	0.29	0.33	0.25	0.23	0.39	0.83	0.90	0.85
	1	0.28	0.30	0.30	0.34	0.26	0.24	0.41	0.83	0.91	0.84

Table 3: RMSE of ATE estimators (n=2000)

y	β_3	M=1	M=2	M=3	M=4	c=2	c=3	c=4	OLS	Strat	IPW
A	0	0.12	0.14	0.13	0.12	0.11	0.16	0.32	0.07	0.12	0.20
	0.5	0.13	0.13	0.12	0.12	0.11	0.19	0.36	0.08	0.13	0.22
	1	0.14	0.13	0.12	0.12	0.11	0.20	0.36	0.11	0.15	0.18
B	0	0.12	0.18	0.15	0.15	0.14	0.14	0.30	0.37	0.44	0.39
	0.5	0.12	0.16	0.15	0.15	0.12	0.16	0.33	0.45	0.51	0.47
	1	0.13	0.17	0.16	0.16	0.13	0.17	0.34	0.46	0.52	0.47
C	0	0.13	0.19	0.17	0.17	0.15	0.13	0.30	0.67	0.75	0.68
	0.5	0.12	0.20	0.18	0.19	0.15	0.15	0.33	0.81	0.88	0.83
	1	0.13	0.20	0.19	0.19	0.15	0.16	0.33	0.82	0.89	0.83

RMSE of the estimator using friends up to order M (columns 1-4), of the estimator using people with at least c friends in common (columns 5-7), of OLS, from stratification based on propensity scores, and of the inverse-propensity weighted estimator. y refers to the type of outcome equation (A: homogeneous effects with linear specification, B: heterogeneous effects, C: heterogeneous effects and quadratic specification).

Table 4: RMSE of ATE estimators (over-controlling; n=500)

y	β_3	M=1	M=2	M=3	M=4	c=2	c=3	c=4
A	0	0.36	0.38	0.37	0.38	0.35	0.19	0.28
	0.5	0.30	0.32	0.32	0.32	0.29	0.19	0.32
	1	0.30	0.31	0.32	0.33	0.29	0.20	0.34
B	0	0.38	0.38	0.38	0.39	0.36	0.19	0.28
	0.5	0.33	0.34	0.34	0.36	0.31	0.18	0.32
	1	0.35	0.35	0.34	0.37	0.33	0.20	0.31
C	0	0.38	0.38	0.37	0.40	0.36	0.19	0.28
	0.5	0.34	0.34	0.35	0.39	0.32	0.18	0.32
	1	0.35	0.36	0.36	0.39	0.34	0.21	0.34

Table 5: RMSE of ATE estimators (over-controlling; n=2000)

y	β_3	M=1	M=2	M=3	M=4	c=2	c=3	c=4
A	0	0.22	0.24	0.21	0.21	0.21	0.10	0.20
	0.5	0.17	0.19	0.18	0.17	0.17	0.11	0.25
	1	0.18	0.20	0.18	0.17	0.18	0.11	0.24
B	0	0.24	0.24	0.21	0.21	0.23	0.10	0.20
	0.5	0.20	0.21	0.19	0.19	0.19	0.10	0.24
	1	0.20	0.21	0.19	0.19	0.19	0.10	0.24
C	0	0.23	0.25	0.22	0.21	0.23	0.09	0.20
	0.5	0.22	0.23	0.21	0.21	0.21	0.10	0.24
	1	0.22	0.24	0.21	0.21	0.21	0.11	0.24

RMSE of the estimator using friends up to order M (columns 1-4), of the estimator using people with at least c friends in common (columns 5-7) in the over-controlling case. y refers to the type of outcome equation (A: homogeneous effects with linear specification, B: heterogeneous effects, C: heterogeneous effects and quadratic specification).