Identification and Estimation of Network Models with Nonparametric Unobserved Heterogeneity

Andrei ZELENEEV*

February 5, 2026
First version: November, 2019†

Abstract

Homophily based on observables is widespread in networks. Therefore, homophily based on unobservables (fixed effects) is also likely to be an important determinant of the interaction outcomes. Failing to properly account for latent homophily (and other complex forms of unobserved heterogeneity) can result in inconsistent estimators and misleading policy implications. To address this concern, we consider a network model with nonparametric unobserved heterogeneity, leaving the role of the fixed effects unspecified. We argue that the interaction outcomes can be used to identify agents with the same values of the fixed effects. The variation in the observed characteristics of such agents allows us to identify the effects of the covariates, while controlling for the fixed effects. Building on these ideas, we construct several estimators of the parameters of interest and characterize their large sample properties. Numerical experiments illustrate the usefulness of the suggested approaches and support the asymptotic theory.

Keywords: network data, homophily, fixed effects

*a.zeleneev@ucl.ac.uk. University College London, Department of Economics, and CeMMAP. I am grateful to my advisors Bo Honoré, Ulrich Müller and especially Kirill Evdokimov for their continuous guidance and support. I also thank Michal Kolesár, Mikkel Plagborg-Møller, Christopher Sims, Mark Watson, and the participants of numerous seminars and conferences for valuable comments and suggestions.

†The orignal JMP version is available here.

1 Introduction

Unobserved heterogeneity is pervasive in economics. The importance of accounting for unobserved heterogeneity is well recognized in microeconometrics, in general, as well as in the network context, in particular. For example, since Abowd et al. (1999), estimating a linear regression with additive (two-way) fixed effects has become a standard approach to analyzing interaction data. Originally employed to account for workers and firms fixed effects in the wage regression context, this technique has become a standard tool to control for two-sided unobserved heterogeneity and decompose it into agent specific effects.[1] Since the seminal work of Anderson and Van Wincoop (2003), the importance of controlling for exporters and importers fixed effects has also been well acknowledged in the context of the international trade network, including nonlinear settings of Santos Silva and Tenreyro (2006) and Helpman et al. (2008). Graham (2017) stresses the importance of accounting for agents' degree heterogeneity (captured by the additive fixed effects) in network formation models.

While the additive fixed effects framework is commonly employed to control for unobservables in networks, it is not flexible enough to capture more complicated forms of unobserved heterogeneity, which are likely to appear in many settings. This concern can be vividly illustrated in the context of estimation of homophily effects, one of the main focuses of the empirical network analysis.[2] Since homophily (assortative matching) based on observables is widespread in networks (e.g., McPherson et al., 2001), homophily based on unobservables (fixed effects) is also likely to be an important determinant of the interaction outcomes. Since observed and unobserved characteristics (i.e., covariates and fixed effects) are typically correlated, the presence of latent homophily significantly complicates identification of the homophily effects associated with the observables (e.g., Shalizi and Thomas, 2011). Failing to properly account for homophily based on unobservables (and other complex forms of unobserved heterogeneity, in general) is likely to result in inconsistent estimators and misleading policy implications.

To address the concern discussed above, we consider a dyadic network model with a flexible (nonparametric) form of unobserved heterogeneity, where the outcome of the

[1] For example, recent applications to employer-employee matched data feature Card et al. (2013); Helpman et al. (2017); Song et al. (2019) among others. The numerous applications of this approach also include the analysis of students-teachers (Hanushek et al., 2003; Rivkin et al., 2005; Rothstein, 2010), patients-hospitals (Finkelstein et al., 2016), firms-banks (Amiti and Weinstein, 2018), and residents-counties matched data (Chetty and Hendren, 2018).

[2] The term homophily typically refers to the tendency of individuals to assortatively match based on their characteristics. For example, individuals tend to form social connections based on gender, race, age, education level, and other socioeconomic characteristics. Similarly, countries that share a border, have the same legal system, language or currency, are more likely to have higher trade volumes.

interaction between agents i and j is given by

$$Y_{ij} = F(W_{ij}'\beta_0 + g(\xi_i, \xi_j)) + \varepsilon_{ij}. \tag{1.1}$$

Here, W_{ij} is a $p \times 1$ vector of pair-specific observed covariates, $\beta_0 \in \mathbb{R}^p$ is the parameter of interest, ξ_i and ξ_j are unobserved fixed effects, and ε_{ij} is an idiosyncratic error. The fixed effects are allowed to interact via the coupling function $g(\cdot, \cdot)$, which is treated as *unknown*. Importantly, we do not require $g(\cdot, \cdot)$ to have any particular structure and do not specify the dimension of ξ. Finally, $F(\cdot)$ is a known (up to location and scale normalizations) invertible link function. The presence of $F(\cdot)$ ensures that (1.1) is flexible enough to cover a broad range of the previously studied dyadic network models with unobserved heterogeneity, including important nonlinear specifications such as network formation or Poisson regression models.[3]

Being agnostic about the dimensions of the fixed effects and the nature of their interactions, (1.1) allows for a wide range of forms of unobserved heterogeneity, including homophily based on unobservables.

Example 1 (Nonparametric homophily based on unobservables)**.** Let $\xi = (\alpha, \nu)' \in \mathbb{R}^2$ and

$$g(\xi_i, \xi_j) = \alpha_i + \alpha_j - \psi(\nu_i, \nu_j),$$

where $\psi(\cdot, \cdot)$ is some function satisfying $\psi(\nu_i, \nu_j) = 0$ whenever $\nu_i = \nu_j$ and increasing in $|\nu_i - \nu_j|$ (e.g., $\psi(\nu_i, \nu_j) = c\,|\nu_i - \nu_j|^\zeta$ for some $c > 0$ and $\zeta \geqslant 1$). Here α represents the standard additive fixed effect, and $\psi(\cdot, \cdot)$ captures latent homophily based on ν: agents with similar values of ν tend to interact with higher intensity compared to agents distant in terms of ν. Again, since the dimension of ξ is not specified, (1.1) can also incorporate homophily based on several unobserved characteristics (multivariate ν) in a similar manner. ■

We study identification and estimation of (1.1) under the assumption that we observe a single network of a growing size.[4] First, we focus on a simpler version of (1.1)

$$Y_{ij} = W_{ij}'\beta_0 + g(\xi_i, \xi_j) + \varepsilon_{ij}. \tag{1.2}$$

We argue that the outcomes of the interactions can be used to identify agents with the same values of the unobserved fixed effects. Specifically, we introduce a certain pseudo-distance d_{ij}

[3]For example, with $F(\cdot)$ equal to the logistic CDF and $g(\xi_i, \xi_j) = \xi_i + \xi_j$, (1.1) corresponds to the network formation model of Graham (2017).

[4]The large single network asymptotics is standard for the literature focusing on identification and estimation of network models with unobserved heterogeneity. See, for example, Graham (2017); Dzemski (2018); Candelaria (2016); Jochmans (2018); Toth (2017); Gao (2020); Gao et al. (2023).

measuring similarity between agents i and j in terms of their latent characteristics. We argue that (i) $d_{ij} = 0$ if and only if $\xi_i = \xi_j$, and (ii) d_{ij} is identified and can be estimated from the data. Consequently, agents with the same values of ξ can be identified based on the pseudo-distance d_{ij}. Then, the variation in the observed characteristics of such agents allows us to identify the parameter of interest β_0 while controlling for the impact of the fixed effects. Importantly, this result is not driven by the linearity or particular functional form of (1.2): we show that a similar identification argument also applies when $W_{ij}'\beta_0$ is replaced by an unknown (nonparametric) function of observables.

Demonstrating that the introduced pseudo-distance d_{ij} is identified plays a central role in the argument outlined above. First, we show that, if the idiosyncratic errors ε_{ij} are homoskedastic, d_{ij} can be readily estimated from the data using simple pairwise-difference regressions. Second, we extend this argument to models with general heteroskedasticity by leveraging and advancing recent developments in the matrix estimation/completion literature. Specifically, we demonstrate that the error free outcomes $Y_{ij}^* := W_{ij}'\beta_0 + g(\xi_i, \xi_j)$ are identified and can be uniformly (across all pairs of agents) consistently estimated. This is a powerful result allowing us to treat Y_{ij}^* as effectively observed and thus greatly simplifying the analysis. In particular, working with Y_{ij}^* instead of Y_{ij} effectively reduces (1.2) to a model without the error term ε_{ij}, which can be interpreted as an extreme form of homoskedasticity. This, in turn, allows us to establish identification of β_0 by applying the same argument as in the homoskedastic model.

Building on these ideas, we construct an estimator of β_0 and characterize its rate of convergence. Following the identification argument, we first estimate the pseudo-distances $\hat{d}_{ij}$, which are used to find agents with similar latent characteristics. Then, we estimate β_0 by combining pairwise-difference regressions of $Y_{ik} - Y_{jk}$ on $W_{ik} - W_{jk}$ for all pairs of agents i and j sufficiently similar in terms of $\hat{d}_{ij}$. Consistency of this estimator crucially relies on the matched agents being different in terms of their observables, resulting in sufficient residual variation in $W_{ik} - W_{jk}$. We characterize the asymptotic behavior of the estimator's bias due to imperfect matching, and provide conditions under which it is consistent.

In the general heteroskedastic case, construction of $\hat{d}_{ij}$ also involves preliminary estimation of the error free outcomes Y_{ij}^*. To provide estimators of Y_{ij}^* and d_{ij} valid under heteroskedastic errors, we build on and extend the approach of Zhang et al. (2017) originally employed in the context of nonparametric graphon estimation. Moreover, to formally establish identification of the error free outcomes, we also propose a modified version of Zhang et al. (2017)'s estimator and demonstrate its consistency in the *max* (matrix) norm, meaning that, in a large network, our estimator recovers Y_{ij}^*'s for all pairs of agents with a high precision at once. To the best of our knowledge, this result is also new to the statistics literature on graphon estimation, which has previously focused on constructing estimators $\hat{Y}_{ij}^*$ and deriving their rates of convergence in terms of the mean square error/Frobenius norm (Chatterjee, 2015; Gao et al., 2015; Klopp et al., 2017; Zhang et al., 2017; Li et al., 2019).

Finally, we want to stress that identification of the error free outcomes is a powerful result, which applies to a general class of dyadic network models beyond (1.1) and can be used as a foundation for establishing new identification results. To the best of our knowledge, this result has not been previously recognized and leveraged in the econometrics literature. In particular, building on it, we demonstrate how the proposed identification and estimation strategies can be naturally extended to cover model (1.1), as well as its nonparametric version. We also argue that the pair-specific fixed effects $g_{ij} = g(\xi_i, \xi_j)$ are identified for all pairs of agents i and j and can be (uniformly) consistently estimated. Identification of g_{ij} is an important result in itself since in many applications the fixed effects are the central objects of interest. Moreover, this result is also of special significance when F is nonlinear because identification of g_{ij} allows us to identify important policy relevant quantities such as pair-specific and average partial effects.

This paper contributes to the literature on econometrics of networks and, more generally, two-way models. The distinctive feature of our model is allowing for flexible nonparametric unobserved heterogeneity: the fixed effects can interact via the unknown coupling function $g(\cdot,\cdot)$. Importantly, we do not require $g(\cdot,\cdot)$ to have any particular structure (other than satisfying a weak smoothness requirement) and do not specify the dimensionality of the fixed effects. This is in contrast to most of the existing approaches, which either explicitly specify the form of $g(\cdot,\cdot)$ or impose additional restrictive assumptions on its shape and smoothness.

Among explicitly specified forms of $g(\cdot,\cdot)$, the additive fixed effects structure $g(\xi_i, \xi_j) = \xi_i + \xi_j$ is by far the most popular way of incorporating unobserved heterogeneity in dyadic network models, e.g., see Graham (2017), Charbonneau (2017), Jochmans (2018), Dzemski (2018), Yan et al. (2019), Candelaria (2016), Gao (2020) and Toth (2017) among others. While this specification provides a practical way of controlling for degree heterogeneity, it does not account for more complicated forms of unobserved heterogeneity including latent homophily. Recent semiparametric extensions of Graham (2017) and related frameworks feature Gao et al. (2023) relaxing separability between $W_{ij}'\beta_0$, ξ_i, and ξ_j. However, Gao et al. (2023) still require ξ to be scalar and assume that the linking probability is increasing in ξ_i and ξ_j, thus ruling out homophily based on unobservables.

The linear factor specification $g(\xi_i, \xi_j) = \xi_i'\xi_j$ is widely used in both network and panel models as a generalization of the additive fixed effect framework.[5] While, as argued in Chen et al. (2021), this specification allows for certain forms of latent homophily, approximating a general function $g(\cdot,\cdot)$ by the linear factor model requires a growing number of factors (e.g., Fernández-Val et al., 2021). Such low-rank approximations, for example, are utilized by Freeman and Weidner (2023) and Beyhum and Mugnier (2024) who consider a panel variation of (1.2). To control the accuracy of the proposed low-rank approximations, both of these papers require $g(\cdot,\cdot)$ to have sufficiently many continuous derivatives and, more importantly, to ensure consistency of their estimator of β_0, they require regressors W to be "high-rank". Both of these requirements are restrictive in the network setting;[6] see Section 2.2.1 for a detailed comparison of the frameworks and approaches to identification. Allowing for both low-rank regressors W and possibly non-differentiable $g(\cdot,\cdot)$ is an important and unique combination of features of our setting differentiating this paper from the rest of the literature.[7]

Employing clustering methods to partition units into groups with similar unobserved characteristics is another method commonly employed to control for interactive (time-varying) unobserved heterogeneity. This approach was proposed and applied to likelihood panel models in the seminal work by Bonhomme et al. (2022) and then extended to semiparametric panel regressions by Beyhum and Mugnier (2024). While the specific clustering methods (involving the observed covariates as well) employed in these papers proved to be instrumental in panel settings, they would not allow one to identify and estimate β_0 in network models when the regressors take the typical form of $W_{ij} = w(X_i, X_j)$, where X_i and X_j are observed characteristics of agents i and j. In this case, the previously employed methods would cluster units into groups with similar values of both ξ and X preventing one from disentangling the effects of observables and unobservables and thus from identifying β_0 as well.[8] Importantly, unlike the previously employed clustering methods, our approach can be used to find units with similar values of ξ and yet different values of X allowing us to identify β_0. Thus, we also complement the methodology of Bonhomme et al. (2022) by providing a new grouping approach which can be instrumental in both network and panel models with low-rank regressors.

[5]Recent studies considering network models with unobserved effects having a linear factor structure include, among others, Chen et al. (2021); Ma et al. (2022); Zeleneev and Zhang (2025).

[6]Non-differentiable functions $g(\cdot,\cdot)$, such as the one provided in Example 1 for $\zeta = 1$, are a common feature of popular latent homophily models (e.g., Hoff et al., 2002; Handcock et al., 2007).

[7]Low-rank regressors are typically ruled out to ensure identification of β_0 even in linear factor models; see, for example, Bai (2009); Moon and Weidner (2015); Armstrong et al. (2022).

[8]Moreover, the clustering methods employed in Bonhomme et al. (2022) and Beyhum and Mugnier (2024) are based on individual-specific moments. Such clustering methods might fail to meaningfully group units in network settings, in which informative individual-specific moments might not exist at all; see Section 2.2.1 and specifically Footnote 12 for details.

When ξ is discrete, the considered model (1.1) belongs to the class of stochastic block models (SBM) with covariates (e.g., Mele et al., 2023; Ma et al., 2022). While our framework and estimation approach are general enough to cover this special setup, in this paper, we focus on the case when unobserved heterogeneity is continuous.[9] In a recent study, Kitamura and Laage (2024) study a nonseparable nonparametric variation of SBM with covariates. Since their approach crucially relies on the discreteness of ξ whereas ours exploits separability between the observables and unobservables as in (1.1), the frameworks considered in this paper and by Kitamura and Laage (2024) are non-nested and complementary.

Another strand of the literature emphasizes the importance of using network data to control for endogeneity in peer effects and other related models (e.g., Goldsmith-Pinkham and Imbens, 2013; Johnsson and Moon, 2021; Auerbach, 2022; Starck, 2025). In these papers, the agents' latent characteristics ξ affect both the individual outcomes of interest as well as the network formation process. To tackle this problem, Auerbach (2022) and Johnsson and Moon (2021) propose using certain network statistics to identify agents with the same values of ξ allowing the authors to control for the unobservables in the other (cross-sectional) regression of interest (e.g., in a peer effects model). The important difference between their and our work is that we use the *same* network data both to control for unobserved heterogeneity and to identify the parameters of interest, allowing us to disentangle the effects of observables and unobservables within a single network model.[10]

Finally, we emphasize that the considered model (1.1) does not incorporate interaction externalities. Specifically, we assume that conditional on the agents' observed and unobserved characteristics, the interaction outcomes are independent. This assumption is plausible when the interactions are primarily bilateral. For excellent recent reviews of econometrics of networks, with and without strategic interactions, we refer the reader to Graham and De Paula (2020); Graham (2020); De Paula (2020).

The rest of the paper is organized as follows. In Section 2, we formally introduce the framework and provide (heuristic) identification arguments for the semiparametric regression model (1.2). Section 3 turns these ideas into estimators of the parameters of interest. In Section 4, we establish consistency of the proposed estimators and derive their rates of convergence. In Section 5, we generalize the proposed identification argument to cover more general settings including the nonlinear model (1.1) as well as its nonparametric analogue.

[9] In fact, when ξ is discrete, agents can be correctly classified into groups with the same values of ξ based on the already introduced pseudo-distance $\hat{d}_{ij}$, which greatly simplifies the asymptotic analysis; we will discuss this in more detail shortly after we introduce Assumption 3 in Section 4.

[10] For example, similarly to the clustering methods of Bonhomme et al. (2022) and Beyhum and Mugnier (2024), a direct application of Auerbach (2022)'s approach to the observed network only allows one to identify agents with the same values of both ξ and X once again precluding identification of β_0.

Section 6 provides numerical and empirical illustrations. A supplementary appendix contains all proofs, and additional discussions and illustrations.

2 Identification of the Semiparametric Model

In this section, we introduce the framework and provide a conceptual discussion of identification of the semiparametric regression model. This discussion is supposed to illustrate the anatomy of the model and to highlight the main insights of our identification strategy, and is deliberately not formalized here. In Section 3, we will turn these ideas into practical estimators. Identification then is formally demonstrated in Section 4, where establish consistency of our estimators and provide their rates of convergence.

2.1 The model

We consider a network consisting of n agents. Each agent i is endowed with characteristics $Z_i = (X_i, \xi_i)$, where $X_i \in \mathcal{X}$ is observed by the econometrician, while $\xi_i \in \mathcal{E}$ is not. We consider the following semiparametric regression model, where the (scalar) outcome of the interaction between agents i and j is given by

$$Y_{ij} = w(X_i, X_j)'\beta_0 + g(\xi_i, \xi_j) + \varepsilon_{ij}, \quad i \neq j. \tag{2.1}$$

Here, $w : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^p$ is a known function, which transforms the observed characteristics of agents i and j into a pair-specific vector of covariates $W_{ij} := w(X_i, X_j)$, $\beta_0 \in \mathbb{R}^p$ is the parameter of interest, and ε_{ij} is an unobserved idiosyncratic error. Note that unlike $w(\cdot,\cdot)$, the coupling function $g : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ is *unknown*, and the dimension of the fixed effect $\xi_i \in \mathcal{E}$ is *not specified*. For simplicity of exposition, we will suppose that $\xi_i \in \mathbb{R}^{d_\xi}$ even though, in principle, the same insights apply when $\mathcal{E}$ is a general metric space.

For concreteness, we will also focus on an undirected model with $Y_{ij} = Y_{ji}$, so $w(\cdot,\cdot)$ and $g(\cdot,\cdot)$ are symmetric functions, and $\varepsilon_{ij} = \varepsilon_{ji}$. The methodology presented in this paper straightforwardly extends to directed networks and general two-way settings including panel models; see Section 5.3 for a more detailed discussion.

The following assumption formalizes the sampling process.

Assumption 1.

(i) $\{Z_i\}_{i=1}^n$ are i.i.d.;

(ii) conditional on $\{Z_i\}_{i=1}^n$, the idiosyncratic errors $\{\varepsilon_{ij}\}_{i<j}$ are independent draws from $P_{\varepsilon_{ij}|Z_i,Z_j}$ with $\mathbb{E}[\varepsilon_{ij}|Z_i,Z_j] = 0$, and $\varepsilon_{ij} = \varepsilon_{ji}$;

(iii) the econometrician observes $\{X_i\}_{i=1}^n$ and $\{Y_{ij}\}_{i\neq j}$ determined by (2.1).

Assumption 1 is standard for the networks literature.[11] The sampling process could be thought of as follows. First, the characteristics of agents $\{Z_i\}_{i=1}^n$ are independently drawn from some population distribution. Then, conditional on the drawn characteristics, the idiosyncratic errors $\{\varepsilon_{ij}\}_{i<j}$ are independently drawn from the conditional distributions, which potentially depend on the characteristics of the corresponding agents Z_i and Z_j.

Remark 2.1. For simplicity of exposition, we assume that we observe Y_{ij} for all pairs of agents i and j. In Section 5.2, we will discuss how to incorporate missing outcomes and sparse networks into the considered framework.

2.2 Identification of β_0: main insights

We study identification and estimation of β_0 under the large network asymptotics, which takes $n \to \infty$. The identification argument is based on the following observation. Suppose that we can identify two agents i and j with the same unobserved characteristics, i.e., with $\xi_i = \xi_j$. Then, for any third agent k, the difference between Y_{ik} and Y_{jk} is given by

$$Y_{ik} - Y_{jk} = \underbrace{(w(X_i, X_k) - w(X_j, X_k))'}_{(W_{ik}-W_{jk})'}\beta_0 + \varepsilon_{ik} - \varepsilon_{jk}. \tag{2.2}$$

The conditional mean independence of the regression errors now guarantees that β_0 can be identified from the regression of $Y_{ik} - Y_{jk}$ on $W_{ik} - W_{jk}$, provided that we have "enough" variation in $W_{ik} - W_{jk}$. Formally, we have

$$\beta_0 = \mathbb{E}\left[(W_{ik} - W_{jk})(W_{ik} - W_{jk})'|Z_i, Z_j\right]^{-1} \mathbb{E}\left[(W_{ik} - W_{jk})(Y_{ik} - Y_{jk})|Z_i, Z_j\right], \tag{2.3}$$

provided that $\mathbb{E}\left[(W_{ik} - W_{jk})(W_{ik} - W_{jk})'|Z_i, Z_j\right]$ is invertible. Since agents i and j are treated as fixed, the expectations are conditional on their characteristics Z_i and Z_j. At the same time, Z_k, the characteristics of agent k, and the idiosyncratic errors ε_{ik} and ε_{jk} are treated as random and integrated over. Note that the invertibility requirement insists on X_i and X_j, the observed characteristics of agents i and j, to be "sufficiently different". Indeed, if not only $\xi_i = \xi_j$ but also $X_i = X_j$, this condition is clearly violated since $W_{ik} - W_{jk} = 0$ for any agent k: in this case, β_0 cannot be identified from the regression (2.2).

Hence, the problem of identification of β_0 can be reduced to the problem of identification of agents i and j with the same values of the unobserved fixed effects ($\xi_i = \xi_j$) but with "sufficiently different" values of X_i and X_j.

[11] See, e.g., Graham (2017); Gao (2020); Gao et al. (2023).

Let Y^*_{ij} be the error free part of Y_{ij}, i.e.,

$$Y^*_{ij} \coloneqq \mathbb{E}\left[Y_{ij}|Z_i, Z_j\right] = w(X_i, X_j)'\beta_0 + g(\xi_i, \xi_j).$$

Consider the following (squared) pseudo-distance between agents i and j

$$\begin{aligned} d^2_{ij} &\coloneqq \min_{\beta\in\mathcal{B}} \mathbb{E}\left[(Y^*_{ik} - Y^*_{jk} - (W_{ik} - W_{jk})'\beta)^2 | Z_i, Z_j\right] \qquad (2.4) \\ &= \min_{\beta\in\mathcal{B}} \mathbb{E}\Bigg[(\underbrace{g(\xi_i, \xi_k) - g(\xi_j, \xi_k)}_{=0, \text{ when } \xi_i = \xi_j} - (W_{ik} - W_{jk})'(\beta - \beta_0))^2 | Z_i, Z_j\Bigg], \end{aligned}$$

where $\mathcal{B} \ni \beta_0$ is some parameter space. Here, the expectation is conditional on the characteristics of agents i and j and is taken over Z_k. Clearly, $d^2_{ij} = 0$ when $\xi_i = \xi_j$: in this case, the minimum is achieved at $\beta = \beta_0$. Moreover, under a suitable (rank) condition (which we will formally discuss in Section 4.1), $d^2_{ij} = 0$ also necessarily implies that $\xi_i = \xi_j$. Consequently, if d^2_{ij} were available, agents with the same values of ξ could be identified based on this pseudo-distance.

However, the expectation (2.4) cannot be directly identified, since the error free outcomes Y^*_{ij} are not observed. In the following Sections 2.3 and 2.4, we will argue that the pseudo-distances d^2_{ij} (or their close analogues) are identified for all pairs of agents i and j and, hence, can be used to identify agents with the same values of ξ (and different values of X).

Remark 2.2. Notice that none of the arguments provided in this section changes if $\xi_i = \xi_j$ is understood as equivalence of the associated functions $g(\xi_i, \cdot)$ and $g(\xi_j, \cdot)$ (in terms of the L^2 distance associated with the distribution P_ξ). Thus, for simplicity of exposition, we will implicitly assume that different values of ξ are associated with different functions $g(\xi, \cdot)$. This assumption is not restrictive since we do not normalize the distribution of ξ, e.g., we do not impose $\xi \sim U[0, 1]$. In particular, the absence of unobserved heterogeneity is allowed: in this case, we have $\xi = \xi_0$ for all agents for some fixed ξ_0.

2.2.1 Comparison with the existing approaches to identification

Before we proceed with identification of d^2_{ij}, we would like to compare our identification argument with the existing results for (interactive fixed effects) panel and network models and highlight some fundamental differences.

One of the most common ways to ensure identification of β_0 in panel regression models with interactive fixed effects is to assume that covariates W_{it} have sufficient independent variation over i and t. This assumption is also commonly referred to as the "high-rank" regressors condition. It plays a crucial role in establishing consistency and deriving asymptotic properties of various estimators of β_0 in models with unobserved heterogeneity following a linear factor form (e.g., Bai, 2009; Moon and Weidner, 2015; Armstrong et al., 2022) and in settings with more general nonparametric structures of unobserved heterogeneity similar to the one studied in this work (Bonhomme et al., 2022; Freeman and Weidner, 2023; Beyhum and Mugnier, 2024).

In the context of this paper, the "high-rank" regressors condition translates into assuming that $W_{ij} = w(X_i, X_j) + \eta_{ij}$, where η_{ij} exhibits sufficient independent variation across dyads, e.g., η_{ij}'s are (conditionally) independent. Identification of β_0 then can be ensured using the "exogenous" ("high-rank") variation in η_{ij} orthogonal to the unobserved (low-dimensional or even low-rank) component $g(\xi_i, \xi_j)$. While this strategy is commonly employed in panel models with high-rank regressors, it cannot be applied in network models with W_{ij} typically taking the (low-dimensional) form of $W_{ij} = w(X_i, X_j)$ *without* the additional exogenous "high-rank" component η_{ij}.

The absence of the η_{ij} component in the model substantially complicates both identification of β_0 as well as the asymptotic analysis of the subsequently constructed analogue estimators. To see this, notice that if $W_{ij} = w(X_i, X_j) + \eta_{ij}$, it would suffice to match agents with the same values of *both* X and ξ, i.e., with $X_i = X_j$ and $\xi_i = \xi_j$, because in this case β_0 can still be identified as in (2.3) thanks to the remaining variation in $W_{ik} - W_{jk} = \eta_{ik} - \eta_{jk}$. This matching strategy, for example, is employed by Beyhum and Mugnier (2024) who generalized the clustering approach of Bonhomme et al. (2022) to semiparametric panel regressions. However, when $W_{ij} = w(X_i, X_j)$ and the η_{ij}-component is absent, such matching approaches fail to identify β_0 due to the lack of the remaining variation in $W_{ik} - W_{jk} = 0$.

In this paper, we overcome this difficulty and establish identification of β_0 by demonstrating how to find agents with the same values of ξ but *different* values of X allowing us to use the variation in $W_{ik} - W_{ij}$ *without* relying on the "high-rank" exogenous variation in η_{ij} like the panel literature does. The latter is the most important and fundamental difference between our work and the recent panel papers by Freeman and Weidner (2023) and Beyhum and Mugnier (2024). Likewise, our identification approach is new and structurally different from the ones previously employed in the literature, and while we focus on network models in this paper, the proposed methodology can also be instrumental in establishing new identification results in panels with low-rank regressors.

Finally, we want to stress that the proposed approach to finding units with the same values of ξ and different values of X is new and structurally different from the previously employed approaches used to match agents with similar latent characteristics in panel and network models. In particular, our framework features two types of individual specific-characteristics, observed X_i and unobserved ξ_i, which are needed to be treated differently: we want to identify the "causal" effect of X_i while keeping ξ_i fixed. At the same time, the existing matching methods, including k-means clustering employed in Bonhomme et al. (2022) and Beyhum and Mugnier (2024), and the similarity based approach pioneered by Zhang et al. (2017) and Auerbach (2022), either drop X_i altogether or effectively match on $Z_i = (X_i, \xi_i)$, which, as explained above, does not allow one to identify β_0.[12]

2.3 Identification under conditional homoskedasticity

In this section, we consider the case when the regression errors are homoskedastic, i.e., when

$$\mathbb{E}\left[\varepsilon_{ij}^2 | Z_i, Z_j\right] = \sigma^2 \quad \text{a.s.} \tag{2.5}$$

For a pair of agents i and j, consider the following conditional expectation

$$q_{ij}^2 \coloneqq \min_{\beta \in \mathcal{B}} \mathbb{E}\left[(Y_{ik} - Y_{jk} - (W_{ik} - W_{jk})'\beta)^2 | Z_i, Z_j\right]. \tag{2.6}$$

Essentially, q_{ij}^2 is a feasible analogue of d_{ij}^2 with Y_{ik} and Y_{jk} replacing Y_{ik}^* and Y_{jk}^*. Importantly, unlike d_{ij}^2, q_{ij}^2 is immediately identified and can be estimated by

$$\hat{q}_{ij}^2 \coloneqq \min_{\beta \in \mathcal{B}} \frac{1}{n-2} \sum_{k \neq i,j} (Y_{ik} - Y_{jk} - (W_{ik} - W_{jk})'\beta)^2. \tag{2.7}$$

Notice that since $Y_{ik} = Y_{ik}^* + \varepsilon_{ik}$ and $Y_{jk} = Y_{jk}^* + \varepsilon_{jk}$,

$$\begin{aligned} q_{ij}^2 &= \min_{\beta \in \mathcal{B}} \mathbb{E}\left[(Y_{ik}^* - Y_{jk}^* - (W_{ik} - W_{jk})'\beta + \varepsilon_{ik} - \varepsilon_{jk})^2 | Z_i, Z_j\right] \\ &= \min_{\beta \in \mathcal{B}} \mathbb{E}\left[(Y_{ik}^* - Y_{jk}^* - (W_{ik} - W_{jk})'\beta)^2 | Z_i, Z_j\right] + \mathbb{E}\left[\varepsilon_{ik}^2 + \varepsilon_{jk}^2 | Z_i, Z_j\right] \\ &= d_{ij}^2 + \mathbb{E}\left[\varepsilon_{ik}^2 | Z_i\right] + \mathbb{E}\left[\varepsilon_{jk}^2 | Z_j\right], \end{aligned} \tag{2.8}$$

where the second and the third equalities follow from Assumption 1(ii). Hence, when the

[12] Also, notice that, even if one abstracts from having or explicitly controlling for observed X_i in the studied network model, clustering based on individual-specific moments h_i originally proposed by Bonhomme et al. (2022) and subsequently employed in Beyhum and Mugnier (2024) might fail to match agents with similar values of ξ. Specifically, their approach requires availability of (consistently estimable) $h_i = \varphi(\xi_i)$ informative about ξ_i but such individual-specific moments might not exist at all in the network setting. To see this, consider a simplified version of the studied model $Y_{ij} = g(\xi_i, \xi_j) + \varepsilon_{ij}$ without covariates, where (i) ξ_i is uniformly distributed on some sphere in $\mathbb{R}^{d_\xi}$, (ii) $g(\xi_i, \xi_j) = g(\|\xi_i - \xi_j\|)$ is effectively determined by $\|\xi_i - \xi_j\|$ exclusively and captures homophily based on ξ, (iii) ε_{ij}'s are iid and independent from all the ξ's. It is clear that, given the spherical symmetry of this model, any individual-specific moments h_i are completely uninformative about ξ_i, rendering the described clustering method inappropriate.

errors are homoskedastic and (2.5) holds, we have

$$q_{ij}^2 = d_{ij}^2 + 2\sigma^2. \tag{2.9}$$

Thus, for every pair of agents i and j, q_{ij}^2 differs from d_{ij}^2 by a constant term $2\sigma^2$.

Imagine that for a fixed agent i, we are looking for a match j with the same value of ξ. As discussed in Section 2.2, such an agent can be identified by minimizing d_{ij}^2 over potential matches. Then, (2.9) ensures that, in the homoskedastic setting, such an agent can also be identified by minimizing q_{ij}^2. Hence, agents with the same values of ξ (and different values of X) can be identified based on q_{ij}^2, which can be directly estimated.

Remark 2.3. The identification argument provided for the homoskedastic model can be naturally extended to allow for $\mathbb{E}[\varepsilon_{ij}^2|Z_i, Z_j] = \mathbb{E}[\varepsilon_{ij}^2|X_i, X_j]$. Indeed, if the skedastic function does not depend on the unobserved characteristics, conditioning on some fixed value $X_j = x$ makes the last term $\mathbb{E}[\varepsilon_{jk}^2|X_j = x, \xi_j] = \mathbb{E}[\varepsilon_{jk}^2|X_j = x]$ in (2.8) constant again. In this case, like in the homoskedastic model, q_{ij}^2 is minimized whenever d_{ij}^2 is, which allows us to identify agents with the same values of ξ.

2.4 Identification under general heteroskedasticity

Under general heteroskedasticity of the errors, the identification strategy based on q_{ij}^2 no longer guarantees finding agents with the same values of ξ. Consider the same process of finding an appropriate match j for a fixed agent i. As shown in (2.8), q_{ij}^2 can be represented as a sum of three components. The first term d_{ij}^2, which we will call the signal, identifies agents with the same values of ξ. The second term $\mathbb{E}\left[\varepsilon_{ik}^2|X_i, \xi_i\right]$ does not depend on j. However, under general heteroskedasticity, the third term $\mathbb{E}[\varepsilon_{jk}^2|X_j, \xi_j]$ depends on ξ_j and distorts the signal. Hence, the identification argument provided in Section 2.3 is no longer valid in this case.

In this section, we will address this issue and extend the arguments of Sections 2.2 and 2.3 to a model with general heteroskedasticity. Specifically, we will (heuristically) argue that the error free outcomes Y_{ij}^* are identified for all pairs of agents i and j. As a result, the pseudo-distance d_{ij}^2 introduced in (2.4) is also identified and can be directly employed to find agents with the same values of ξ (and different values of X).

2.4.1 Identification of Y_{ij}^*

With Y_{ij}^* and $Y_{ij} = Y_{ij}^* + \varepsilon_{ij}$ collected as entries of $n \times n$ matrices Y^* and Y (with diagonal elements of Y missing), the problem of identification and estimation of Y^* based on its noisy proxy Y can be interpreted as a particular variation of the classic matrix estimation/completion problem. Specifically, it turns out that the considered network model (2.1) is an example of the latent space model (see, for example, Chatterjee (2015) and the references therein). In the standard formulation of the latent space model, the entries of (symmetric) matrix Y have the form of

$$Y_{ij} = f(Z_i, Z_j) + \varepsilon_{ij}, \tag{2.10}$$

where f is some (unknown) symmetric function, $Z_1, \dots, Z_n$ are some latent variables associated with the corresponding rows and columns of Y, and the errors $\{\varepsilon_{ij}\}_{i<j}$ are assumed to be (conditionally) independent.[13][14] Notice, that the studied model fits this general formulation, even though, in our setting, we observe X_i (a subvector of Z_i).

It turns out that the particular structure of the latent space model (2.10) allows one to construct a consistent estimator of Y^* based on a single measurement Y. For example, Chatterjee (2015); Gao et al. (2015); Klopp et al. (2017); Zhang et al. (2017) construct such estimators and establish their consistency in terms of the mean square error (MSE).

In particular, we build on the estimation strategy of Zhang et al. (2017) to argue that the error free outcomes Y^*_{ij} are identified for all pairs of agents i and j. The proposed identification strategy consists of two main steps. First, we argue that we can identify agents with the same values of (both) X and ξ. Then, building on this result, we demonstrate how Y^*_{ij} can be constructively identified.

Step 1: Identification of agents with the same values of X and ξ

Consider a subpopulation of agents with a fixed value of $X = x$ exclusively. Let $g_x(\xi_i, \xi_j) \coloneqq w(x,x)'\beta_0 + g(\xi_i, \xi_j)$ and $P_{\xi|X}(\xi|x)$ denote the conditional distribution of ξ given $X = x$. In this subpopulation, consider the following (squared) pseudo-distance between agents i and j

$$\begin{aligned} d^2_\infty(i,j;x) &\coloneqq \sup_{\xi_k \in \operatorname{supp}(\xi|X=x)} \left|\mathbb{E}\left[(Y_{i\ell} - Y_{j\ell})Y_{k\ell} | \xi_i, \xi_j, \xi_k, X = x\right]\right| \\ &= \sup_{\xi_k \in \operatorname{supp}(\xi|X=x)} \left| \int (g_x(\xi_i, \xi_\ell) - g_x(\xi_j, \xi_\ell)) g_x(\xi_k, \xi_\ell) dP_{\xi|X}(\xi_\ell; x) \right|, \end{aligned}$$

[13] As noted, for example, in Bickel and Chen (2009) and Bickel et al. (2011), the latent space model is natural in exchangeable settings due to the Aldous-Hoover theorem (Aldous, 1981; Hoover, 1979). For a detailed discussion of this result and other representation theorems for exchangeable random arrays, see, for example, Kallenberg (2005) and Orbanz and Roy (2015).

[14] The problem of estimation of $Y^*_{ij} = f(Z_i, Z_j)$ from Y is also known as nonparametric regression without knowing the design (Gao et al., 2015) or blind regression (Li et al., 2019). If Y_{ij} is binary, Y can be interpreted as the adjacency matrix of a random graph. In this case, the function $f(\cdot,\cdot)$ is called a graphon, and this problem is commonly referred to as graphon estimation (see, for example, Gao et al., 2015; Klopp et al., 2017; Zhang et al., 2017).

where the second equality uses Assumption 1(ii).

The finite sample analogue of d_∞^2 was originally introduced in Zhang et al. (2017) in the context of nonparametric graphon estimation. It is also closely related to the so-called similarity distance inducing a weak topology on graphons (e.g., see, Lovász (2012) and the references therein).

First, notice that (under weak smoothness conditions) $d_\infty^2(i,j;x)$ is directly identified and, if a sample of n_x agents with $X = x$ is available, it can be estimated by

$$\hat{d}_\infty^2(i,j;x) \coloneqq \max_{k \neq i,j} \left| (n_x - 3) \sum_{\ell \neq i,j,k} (Y_{i\ell} - Y_{j\ell}) Y_{k\ell} \right|.$$

Second, note that $d_\infty^2(i,j;x) = 0$ implies that

$$\int \left(g_x(\xi_i, \xi_\ell) - g_x(\xi_j, \xi_\ell)\right) g_x(\xi_k, \xi_\ell) dP_{\xi|X}(\xi_\ell; x) = 0 \tag{2.11}$$

for almost all ξ_k. Evaluating (2.11) at $\xi_k = \xi_i$ and $\xi_k = \xi_j$ and subtracting the latter from the former, we conclude

$$\int \left(g_x(\xi_i, \xi_\ell) - g_x(\xi_j, \xi_\ell)\right)^2 dP_{\xi|X}(\xi_\ell; x) = \int \left(g(\xi_i, \xi_\ell) - g(\xi_j, \xi_\ell)\right)^2 dP_{\xi|X}(\xi_\ell; x) = 0.$$

Thus, $d_\infty^2(i,j;x) = 0$ implies that $g(\xi_i, \cdot)$ and $g(\xi_j, \cdot)$ are the same (in terms of the L^2 distance associated with the conditional distribution of $\xi | X = x$).[15] Finally, since the equivalence of $g(\xi_i, \cdot)$ and $g(\xi_j, \cdot)$ is understood as $\xi_i = \xi_j$, we can use $d_\infty^2(i,j;x)$ to identify agents with the same values of both X and ξ.

Remark 2.4. The idea of using different versions of d_∞^2 for finding agents with similar characteristics is not new. For example, Zhang et al. (2017) originally employed it for nonparametric graphon estimation, and Auerbach (2022) used it to control for unobservables in a partially linear model using network data.[16] Once again, we want to stress that this strategy alone only allows one to identify agents with the same values of *both* X and ξ. Such matches, however, cannot be used to disentangle the effects of observables and unobservables and to identify β_0, which is the primary focus of this paper.

Step 2: Identification of Y_{ij}^*

Now, being able to identify agents with the same values of X and ξ, we can also identify the

[15] Similar arguments are also provided in Lovász (2012) and Auerbach (2022).

[16] Other recent applications of this methodology also include construction of network resampling methods (Nowakowicz, 2024), estimation of grouped fixed effects models (Mugnier, 2025), and treatment effect estimation in panels (Athey and Imbens, 2025; Deaner et al., 2025).

error free outcome $Y_{ij}^* = w(X_i, X_j)'\beta_0 + g(\xi_i, \xi_j)$ for any pair of agents i and j. Specifically, for a fixed agent i, we can construct a collection of agents with $X = X_i$ and $\xi = \xi_i$, i.e., $\mathcal{N}_i \coloneqq \{i' : X_{i'} = X_i, \xi_{i'} = \xi_i\}$. Similarly, we construct $\mathcal{N}_j \coloneqq \{j' : X_{j'} = X_j, \xi_{j'} = \xi_j\}$. Then,

$$\begin{aligned}\frac{1}{n_i n_j} \sum_{i' \in \mathcal{N}_i} \sum_{j' \in \mathcal{N}_j} Y_{i'j'} &= \frac{1}{n_i n_j} \sum_{i' \in \mathcal{N}_i} \sum_{j' \in \mathcal{N}_j} \left(w(X_i, X_j) + g(\xi_i, \xi_j) + \varepsilon_{i'j'}\right) \\ &= Y_{ij}^* + \frac{1}{n_i n_j} \sum_{i' \in \mathcal{N}_i} \sum_{j' \in \mathcal{N}_j} \varepsilon_{i'j'} \quad \xrightarrow{p} \quad Y_{ij}^*, \qquad n_i, n_j \to \infty, \qquad (2.12)\end{aligned}$$

where n_i and n_j denote the number of elements in $\mathcal{N}_i$ and $\mathcal{N}_j$, respectively. Since in the population we can construct arbitrarily large $\mathcal{N}_i$ and $\mathcal{N}_j$, (2.12) implies that Y_{ij}^* is identified.

Remark 2.5. Although the identification argument provided above is heuristic, it captures the main insights and will be formalized later. Specifically, in Section 4.3, we will construct a particular estimator $\tilde{Y}_{ij}^*$ and establish its uniform consistency, i.e., we will demonstrate that $\max_{i,j} |\tilde{Y}_{ij}^* - Y_{ij}^*| = o_p(1)$. This formally proves that Y_{ij}^* is identified for all i and j.

Identifiability of Y_{ij}^* is a strong result, which, to the best of our knowledge, is new to the econometrics literature on identification of network and, more generally, two-way models. Importantly, it is not due to the specific parametric form or additive separability (in X and ξ) of the model (2.1). In fact, by essentially the same argument, the error free outcomes $Y_{ij}^* = f(X_i, \xi_i, X_j, \xi_j)$ are also identified in a fully non-separable nonparametric model

$$Y_{ij} = f(X_i, \xi_i, X_j, \xi_j) + \varepsilon_{ij}, \quad \mathbb{E}\left[\varepsilon_{ij} | X_i, \xi_i, X_j, \xi_j\right] = 0.$$

The established result implies that for studying identification of such models, the error free outcome Y_{ij}^* can be treated as directly observed. Removing the error part greatly simplifies the analysis and provides a powerful foundation for establishing further identification results in more complicated settings; see Section 5 for details and examples.

For example, in the particular context of the model (2.1), identifiability of Y_{ij}^* implies that the pseudo-distances d_{ij}^2 are also identified for all pairs of agents i and j. Hence, as discussed in Section 2.2, agents with the same values of ξ (and different values of X) and, subsequently, β_0 can be identified based on d_{ij}^2.

3 Estimation of the Semiparametric Model

In this section, we turn the ideas of Section 2 into an estimation procedure. First, we construct an estimator of β_0 assuming that some estimator of the pseudo-distances $\hat{d}_{ij}^2$ is already available for the researcher. Then, we discuss how to construct $\hat{d}^2_{ij}$ in the homoskedastic and general heteroskedastic settings.

3.1 Estimation of β_0

Suppose that we start with some (uniformly consistent) estimator of the pseudo-distances d^2_{ij} denoted by $\hat{d}^2_{ij}$. Using $\hat{d}^2_{ij}$, we construct the following kernel based estimator of β_0

$$\hat{\beta} := \left(\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h^2_n}\right) \sum_{k\neq i,j} \Delta W_{ijk}\Delta W'_{ijk}\right)^{-1} \left(\sum_{i<j} K\left(\frac{\hat{d}^2_{ij}}{h^2_n}\right) \sum_{k\neq i,j} \Delta W_{ijk}\Delta Y_{ijk}\right), \tag{3.1}$$

where $\Delta W_{ijk} := W_{ik} - W_{jk}$ and $\Delta Y_{ijk} := Y_{ik} - Y_{jk}$, $K : \mathbb{R}_+ \to \mathbb{R}$ is some kernel supported on $[0,1]$, and h_n is a bandwidth, which needs to satisfy $h_n \to 0$ (and some additional requirements) as $n \to \infty$. Hereafter, we will also use the notations $\sum_{i<j} := \sum_{i,j\in[n],i<j}$ and $\sum_{k\neq i,j} := \sum_{k\in[n],k\neq i,j}$, where $[n] = \{1,\dots,n\}$.

As discussed previously, β_0 can be estimated by the regression of ΔY_{ijk} on ΔW_{ijk} with fixed agents i and j satisfying $\xi_i = \xi_j$, and, consequently, $d^2_{ij} = 0$; see (2.2) and (2.3). However, in a finite sample, we are never guaranteed to find a pair of agents with exactly the same values of unobserved characteristics. The proposed estimator $\hat{\beta}$ addresses this issue: it combines all of the pairwise-difference regressions weighted by $K(\hat{d}^2_{ij}/h^2_n)$. Typically, the smaller $\hat{d}^2_{ij}$ is, the closer agents i and j appear to be in terms of ξ_i and ξ_j, and the higher weight is given to the corresponding pairwise-difference regression. Specifically, we show that, with probability approaching one, only the pairs that satisfy $\|\xi_i - \xi_j\| \leqslant \alpha h_n$ are given positive weights, where α is some positive constant. Since $h_n \to 0$, the quality of these matches increases and the bias introduced by the imperfect matching vanishes as the sample size grows. In Section 4.1, we formalize this discussion by providing the necessary regularity conditions and establishing the rate of convergence for $\hat{\beta}$.

3.2 Estimation of d^2_{ij}

The kernel based estimator (3.1) builds on the estimated pseudo-distances $\{\hat{d}^2_{ij}\}_{i\neq j}$. In this section, we construct particular estimators of d^2_{ij} for both homoskedastic and general heteroskedastic settings. Their asymptotic properties will be established in Section 4.2.

3.2.1 Estimation of d^2_{ij} under conditional homoskedasticity of ε_{ij}

We start with considering the homoskedastic setting. Recall that in this case, the pseudo-distance of interest d^2_{ij} is closely related to another quantity q^2_{ij} defined in (2.6). Specifically, according to (2.9), $q_{ij}^2 = d_{ij}^2 + 2\sigma^2$, where σ^2 stands for the conditional variance of ε_{ij}. Moreover, unlike d_{ij}^2, q_{ij}^2 can be directly estimated from the raw data as in (2.7).

Then, a natural way to estimate d_{ij}^2 is to subtract $2\hat{\sigma}^2$ from $\hat{q}_{ij}^2$, where $\hat{\sigma}^2$ is an estimator σ^2. One candidate estimator of σ^2 is given by

$$2\hat{\sigma}^2 = \min_{i,j\neq j} \hat{q}_{ij}^2. \tag{3.2}$$

Indeed, in large samples, we expect $\min_{i,j\neq i} d_{ij}^2$ to be small since we are likely to find a pair of agents similar in terms of ξ. Hence, in large samples, $\min_{i,j\neq i} q_{ij}^2 = \min_{i,j\neq i} d_{ij}^2 + 2\sigma^2$ is expected to be close to $2\sigma^2$. Then, d_{ij}^2 can be estimated by

$$\hat{d}_{ij}^2 = \hat{q}_{ij}^2 - 2\hat{\sigma}^2 = \hat{q}_{ij}^2 - \min_{i,j\neq i} \hat{q}_{ij}^2. \tag{3.3}$$

3.2.2 Estimation of d_{ij}^2 under general heteroskedasticity of ε_{ij}

As suggested by Section 2.4, under general heteroskedasticity of the errors, the first step of estimation of d_{ij}^2 is to construct an estimator of Y_{ij}^*. Once $\hat{Y}_{ij}^*$ are constructed for all pairs of agents, d_{ij}^2 can be estimated by

$$\hat{d}_{ij}^2 \coloneqq \min_{\beta\in\mathcal{B}} \frac{1}{n}\sum_{k=1}^{n} (\hat{Y}_{ik}^* - \hat{Y}_{jk}^* - (W_{ik} - W_{jk})'\beta)^2. \tag{3.4}$$

As pointed out in Section 2.4.1, the problem of estimation of Y_{ij}^* is closely related to the classic matrix (graphon) estimation/completion problem, with a number of candidate estimators available in the literature (see, e.g., Chatterjee, 2015). The researcher can pick an appropriate estimator of $\hat{Y}_{ij}^*$ depending on the context.

In this paper, we propose estimating Y_{ij}^* extending the approach of Zhang et al. (2017). For simplicity, first, we consider the case when X is discrete and takes finitely many values. Formal statistical guarantees for the proposed estimator and a discussion of the general case are provided in Section 4.2.2.

First, for all pairs of agents i and j, we estimate

$$\hat{d}_{\infty}^2(i,j) \coloneqq \max_{k\neq i,j} \left| (n-3)^{-1} \sum_{\ell\neq i,j,k} (Y_{i\ell} - Y_{j\ell})Y_{k\ell} \right|. \tag{3.5}$$

Then, for any agent i, we define its neighborhood $\hat{\mathcal{N}}_i(n_i)$ as a collection of n_i agents closest to agent i in terms of $\hat{d}^2_\infty$ among all agents with $X = X_i$, i.e.

$$\hat{\mathcal{N}}_i(n_i) := \{i' : X_{i'} = X_i, \text{Rank}(\hat{d}^2_\infty(i,i')|X = X_i) \leqslant n_i\}. \tag{3.6}$$

Also notice that by construction, $i \in \hat{\mathcal{N}}_i(n_i)$, so agent i is always included in its neighborhood. Essentially, for any agent i, its neighborhood $\hat{\mathcal{N}}_i(n_i)$ is a collection of agents with the same observed and similar unobserved characteristics. Note that since X is discrete and takes finitely many values, we require $X_{i'} = X_i$. Also, note that the number of agents included in the neighborhoods should grow (at a certain rate) as the sample size increases.

Once the neighborhoods are constructed, we estimate Y^*_{ij} by

$$\hat{Y}^*_{ij} = \frac{\sum_{i' \in \hat{\mathcal{N}}_i(n_i)} Y_{i'j}}{n_i}, \tag{3.7}$$

where, for the ease of notation, we let $Y_{i'j} = 0$ whenever $i' = j$. Note that $\hat{Y}^*_{ij}$ is also defined for $i = j$: despite Y_{ii} is not observed, we still can estimate the associated error free outcome $Y^*_{ii} := w(X_i, X_i) + g(\xi_i, \xi_i)$.

Remark 3.1. Notice that the proposed estimator (3.7) differs from the one discussed in Section 2.4.1. Specifically, (2.12) suggests using

$$\tilde{Y}^*_{ij} = \frac{1}{n_i n_j} \sum_{i' \in \hat{\mathcal{N}}_i(n_i)} \sum_{j' \in \hat{\mathcal{N}}_j(n_j)} Y_{i'j'}. \tag{3.8}$$

Next, we will demonstrate that the rate of convergence for $\hat{\beta}$ depends on the asymptotic properties of the first step estimator $\hat{d}^2_{ij}$. While $\tilde{Y}^*_{ij}$ is a natural and (uniformly) consistent estimator of Y^*_{ij}, i.e., we will later establish $\max_{i,j} |\tilde{Y}^*_{ij} - Y^*_{ij}| = o_p(1)$, it turns out that using $\hat{Y}^*_{ij}$ as in (3.7) delivers better rates of converges for $\hat{d}^2_{ij}$ and, consequently, for $\hat{\beta}$ too.

4 Large Sample Theory

In this section, we formally study the asymptotic properties of the estimators we provided in Section 3. The following set of basic regularity conditions will be used throughout the rest of the paper.

Assumption 2.

(i) $w : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^p$ is a symmetric bounded function, where $\text{supp}\,(X) \subseteq \mathcal{X}$;

(ii) $\text{supp}\,(\xi) \subseteq \mathcal{E}$, where $\mathcal{E}$ is a compact subset of $\mathbb{R}^{d_\xi}$;

(iii) $g : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ is a symmetric bounded function; moreover, for some $\overline{G} > 0$, we have

$$|g(\xi_1, \xi) - g(\xi_2, \xi)| \leqslant \overline{G}\, \|\xi_1 - \xi_2\| \quad \text{for all } \xi_1, \xi_2, \xi \in \mathcal{E};$$

(iv) for some $c > 0$, $\mathbb{E}\left[e^{\lambda \varepsilon_{ij}} | X_i, \xi_i, X_j, \xi_j\right] \leqslant e^{c\lambda^2}$ for all $\lambda \in \mathbb{R}$ a.s.

Conditions (i) and (ii) are standard. Condition (iii) requires $g(\cdot, \cdot)$ to be (uniformly) Lipschitz continuous. Condition (iv) requires the conditional distribution of the error $\varepsilon_{ij} | X_i, \xi_i, X_j, \xi_j$ to be sub-Gaussian, uniformly over (X_i, ξ_i, X_j, ξ_j). It allows us to invoke certain concentration inequalities and derive rates of uniform convergence.

4.1 Rate of convergence for $\hat{\beta}$

In this section, we provide necessary regularity conditions and establish the rate of convergence for the kernel based estimator $\hat{\beta}$ introduced in (3.1). We will do that assuming that our candidate estimator $\hat{d}^2_{ij}$ converges to d^2_{ij} (uniformly across all pairs) at a certain rate $R_n \to \infty$, i.e.,

$$\min_{i, j \neq i} |\hat{d}^2_{ij} - d^2_{ij}| = O_p(R_n^{-1}). \tag{4.1}$$

We will first derive the result treating (4.1) as a high-level assumption, and then verify that it holds and characterize R_n for the candidate estimators (3.3) and (3.4).

We establish consistency and derive a guaranteed rate of convergence of $\hat{\beta}$ under the following regularity conditions. For simplicity of exposition, we state these conditions for the case when ξ is scalar but the main result of this section still holds for multivariate ξ under minimal appropriate modifications of the assumptions below.

Assumption 3.

(i) $\xi \in \mathbb{R}$, and $\xi | X = x$ is continuously distributed for all $x \in \operatorname{supp}(X)$; its conditional density $f_{\xi|X}$ (with respect to the Lebesgue measure) satisfies $\sup_{x \in \operatorname{supp}(X)} \sup_{\xi \in \mathcal{E}} f_{\xi|X}(\xi|x) \leqslant \overline{f}_{\xi|X}$ for some constant $\overline{f}_{\xi|X} > 0$;

(ii) for all $x \in \operatorname{supp}(X)$, $f_{\xi|X}(\xi|x)$ is continuous at almost all ξ (with respect to the conditional distribution of $\xi|X = x$); moreover, there exist positive constants $\overline{\delta}$ and γ such that for all $\delta \in (0, \overline{\delta})$ and for all $x \in \operatorname{supp}(X)$,

$$\mathbb{P}\left(\xi_i \in \{\xi : f_{\xi|X}(\xi|x) \text{ is continuous on } B_\delta(\xi)\} | X_i = x\right) \geqslant 1 - \gamma\delta; \tag{4.2}$$

(iii) there exists $C_\xi > 0$ such that for all $x \in \text{supp}(X)$ and for any convex set $\mathcal{D} \in \mathcal{E}$ such that $f_{\xi|X}(\cdot; x)$ is continuous on $\mathcal{D}$, we have $\left|f_{\xi|X}(\xi_1|x) - f_{\xi|X}(\xi_2|x)\right| \leqslant C_\xi |\xi_1 - \xi_2|$.

Assumption 3 describes the properties of the conditional distribution of $\xi|X$. Note that we focus on the case when $\xi|X = x$ is continuously distributed for all $x \in \text{supp}(X)$ even though our framework straightforwardly allows for the (conditional) distribution of ξ to have point masses or to be discrete. In fact, the asymptotic analysis is substantially simpler in the latter case. Specifically, if ξ is discrete (and takes finitely many values), the agents can be consistently clustered into groups with the same values of ξ based on the same pseudo-distance $\hat{d}_{ij}^2$. In this case, $\hat{\beta}$ is asymptotically equivalent to the oracle pairwise-difference estimator using the knowledge of the true cluster membership, and, as a result, it is asymptotically normal and unbiased. Moreover, in this case, β_0 can also be estimated by the pooled linear regression, which includes additional interactions of the dummy variables for the estimated cluster membership.[17]

Conditions (ii) and (iii) are weak smoothness requirements. The second part of Condition (ii) bounds the probability mass of $\xi|X = x$, for which $f_{\xi|X}(\xi|X)$ is not potentially continuous on a ball $B_\delta(\xi)$. It allows us to control the probability mass of ξ close to the boundary of its support, where $f_{\xi|X}(\xi|X)$ is allowed to be discontinuous.

Example (Illustration of Assumption 3(ii))**.** Suppose $\xi|X = x$ is supported and continuously distributed on $[0,1]$ for all $x \in \text{supp}(X)$. Then $f_{\xi|X}(\xi|x)$ is continuous on $B_\delta(\xi)$ for all $\xi \in [\delta, 1-\delta]$. Then (4.2) is satisfied with $\gamma = 2\overline{f}_{\xi|X}$, where $\overline{f}_{\xi|X}$ is as in Assumption 3(i). ■

Assumption 4.

(i) there exist $\underline{\lambda} > 0$ and $\underline{\delta} > 0$ such that

$$\mathbb{P}\left((X_i, X_j) \in \left\{(x_1, x_2) : \lambda_{min}(\mathcal{C}(x_1, x_2)) > \underline{\lambda}, \int f_{\xi|X}(\xi|x_1) f_{\xi|X}(\xi|x_2) d\xi > \underline{\delta}\right\}\right) > 0,$$

where

$$\mathcal{C}(x_1, x_2) := \mathbb{E}\left[(w(x_1, X) - w(x_2, X))(w(x_1, X) - w(x_2, X))'\right]; \tag{4.3}$$

(ii) for each $\delta > 0$, there exists $C_\delta > 0$ such that

$$\inf_\beta \mathbb{E}\left[\left(g(\xi_i, \xi_k) - g(\xi_j, \xi_k) - (w(X_i, X_k) - w(X_j, X_k))'\beta\right)^2 |X_i, \xi_i, X_j, \xi_j\right] > C_\delta$$

a.s. for (X_i, ξ_i) and (X_j, ξ_j) satisfying $|\xi_i - \xi_j| \geqslant \delta$;

[17] Similar ideas are also explored in Mugnier (2025) in the context of grouped panel models.

(iii) $d_{ij}^2 \equiv d^2(X_i, \xi_i, X_i, \xi_j) = c(X_i, X_j, \xi_i)(\xi_j - \xi_i)^2 + r(X_i, \xi_i, X_j, \xi_j)$, where $|r(X_i, \xi_i, X_j, \xi_j)| \leqslant C\,|\xi_j - \xi_i|^3$ a.s. for some $C > 0$, and $0 < \underline{c} < c(X_i, X_j, \xi_i) < \overline{c}$ a.s.

Assumption 4 is a collection of identification conditions. Specifically, Condition (i) is the identification condition for β_0. It ensures that in a growing sample, it is possible to find a pair of agents i and j such that (i) X_i and X_j are "sufficiently different", so the minimal eigenvalue $\lambda_{min}(\mathcal{C}(X_i, X_j)) > \underline{\lambda} > 0$, (ii) and yet ξ_i and ξ_j are increasingly similar. The latter is guaranteed by $\int f_{\xi|X}(\xi|X_i) f_{\xi|X}(\xi|X_j) d\xi > \underline{\delta}$, which implies that the conditional supports of $\xi_i|X_i$ and $\xi_j|X_j$ have a non-trivial overlap. Condition (i) is crucial for establishing consistency of $\hat{\beta}$.

Condition (ii) ensures that d_{ij}^2 is bounded away from zero whenever $|\xi_i - \xi_j|$ is. Notice that it also guarantees that agents that are close in terms of d_{ij}^2, must also be similar in terms of ξ. Hence, Condition (ii) justifies using the pseudo-distance d_{ij}^2 for finding agents with similar values of ξ in finite samples. It also can be interpreted as a rank type condition: for fixed agents i and j with $\xi_i \neq \xi_j$, $g(\xi_i, \xi_k) - g(\xi_j, \xi_k)$ cannot be expressed as a linear combination of the components of $W_{ik} - W_{jk}$.

Condition (iii) is a local counterpart of Condition (ii). It says that, as a function of ξ_j, $d^2(X_i, X_j, \xi_i, \xi_j)$ can be locally quadratically approximated around $\xi_j = \xi_i$, and the approximation remainder can be uniformly bounded as $O(|\xi_j - \xi_i|^3)$. Condition (iii) also explains why we divide $\hat{d}_{ij}^2$ by h_n^2 for computing the kernel weights. Indeed, locally $d_{ij}^2 \propto (\xi_j - \xi_i)^2$, so the bandwidth h_n effectively controls how large $|\xi_j - \xi_i|$ can be for the pair of agents i and j to get a positive weight $K(\frac{\hat{d}_{ij}^2}{h_n^2})$.

Assumption 5.

(i) $K : \mathbb{R}_+ \to \mathbb{R}$ is supported on $[0, 1]$ and bounded by $\overline{K} < \infty$. K satisfies $\mu_K := \int K(u^2) du > 0$ and $|K(z) - K(z')| \leqslant \overline{K}'\,|z - z'|$ for all $z, z' \in \mathbb{R}_+$ for some $\overline{K}' > 0$;

(ii) $h_n \to 0$, $nh_n / \ln n \to \infty$ and $R_n h_n^2 \to \infty$ for some $R_n \to \infty$ satisfying (4.1).

Assumption 5 specifies the properties of the kernel K and the bandwidth h_n. Condition (i) imposes a number of fairly standard restrictions on K including Lipschitz continuity. Condition (ii) restricts the rates at which the bandwidth is allowed to shrink towards zero. The requirement $nh_n / \ln n \to \infty$ ensures that we have a growing number of potential matches as the sample size increases. Additionally, to get the desired results we need $R_n h_n^2 \to \infty$: the bandwidth cannot go to zero faster than $R_n^{-1/2}$. This requirement allows us to bound the effect of the sampling variability coming from the first step (estimation of $\{d_{ij}^2\}_{i \neq j}$) on the second step (estimation of β_0).

Assumption 6. There exists a bounded function $G : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ such that for all $\xi_1, \xi_2, \xi \in \mathcal{E}$

$$g(\xi_1, \xi) - g(\xi_2, \xi) = G(\xi_1, \xi)(\xi_1 - \xi_2) + r_g(\xi_1, \xi_2, \xi);$$

and there exists $C > 0$ such that for all $\delta_n \downarrow 0$

$$\limsup_{n\to\infty} \frac{\sup_{\xi} \sup_{\xi_1 : |\xi_1 - \xi| > \delta_n} \sup_{\xi_2 : |\xi_2 - \xi_1| \leqslant \delta_n} |r_g(\xi_1, \xi_2, \xi)|}{\delta_n^2} < C.$$

Assumption 6 is a weak smoothness requirement. It guarantees that as a function of ξ_2, the difference $g(\xi_1, \xi) - g(\xi_2, \xi)$ can be (locally) linearized around $\xi_2 = \xi_1$ provided that ξ_2 is close to ξ_1 relative to the distance between ξ_1 and ξ (guaranteed by the restrictions $|\xi_1 - \xi| > \delta_n$ and $|\xi_2 - \xi_1| \leqslant \delta_n$). The goal of introducing these restrictions is to allow for a *possibly non-differentiable* g, e.g., $g(\xi_i, \xi_j) = \kappa\,|\xi_i - \xi_j|$, since such models of latent homophily are common in the literature (e.g., Hoff et al., 2002; Handcock et al., 2007). We provide an illustration of Assumption 6 in Appendix B.

Theorem 1. *Suppose that (4.1) holds for some $R_n \to \infty$. Then, under Assumptions 1-6,*

$$\hat{\beta} - \beta_0 = O_p\left(h_n^2 + \frac{R_n^{-1}}{h_n} + \frac{R_n^{-1}}{h_n^2}\left(\frac{\ln n}{n}\right)^{1/2} + n^{-1}\right). \tag{4.4}$$

Theorem 1 provides a guaranteed rate of convergence of $\hat{\beta}$ and, thus, formally establishes identification of β_0. The derived rate consists of four terms. The h_n^2 term accounts for the bias due to the imperfect ξ-matching. The terms involving R_n are due to the fact that $\{d_{ij}^2\}_{i\neq j}$ are unknown and need to be estimated. Finally, n^{-1} is the standard sampling variability term in a regression with $O(n^2)$ observations.

In Section 4.2, we will derive R_n for the considered estimators of $\{d_{ij}^2\}_{i\neq j}$ and complete the characterization of the rate of convergence for $\hat{\beta}$ provided in (4.4). In particular, we will demonstrate that, under certain conditions, the candidate estimators (3.3) and (3.4) satisfy (4.1) with $R_n = (\frac{n}{\ln n})^{1/2}$, or with an even slower-growing R_n depending on d_ξ. For such values of R_n, (4.4) effectively simplifies as

$$\hat{\beta} - \beta_0 = O_p\left(h_n^2 + \frac{R_n^{-1}}{h_n}\right), \tag{4.5}$$

where we also used $\frac{R_n^{-1}}{h_n^2} = o(1)$ implied by Assumption 5(ii). In particular, under $h_n \propto R_n^{-1/3}$,

$\hat{\beta}$ is guaranteed to achieve the following rate of convergence

$$\hat{\beta} - \beta_0 = O_p\left(R_n^{-2/3}\right). \tag{4.6}$$

The rate of convergence established by Theorem 1 is not necessarily optimal and potentially can be improved, especially if additional smoothness conditions are imposed. The main goal of Theorem 1 is to demonstrate consistency of $\hat{\beta}$ and thus complement Section 2 by formally establishing identification of β_0 under minimal primitive conditions.

Finally, we want to highlight again two important features of the setting differentiating it from the previously studied frameworks and complicating the asymptotic analysis. First, as discussed in more detail in Section 2.2.1, we consider low-dimensional (or even low-rank) regressors $W_{ij} = w(X_i, X_j)$ typical for network models instead of relying on "high-rank" regressors like most of the related literature does. Second, we only impose minimal smoothness assumptions on $g(\cdot,\cdot)$ and do not require it to be differentiable whereas the alternative approaches building low-rank approximations of interactive unobserved heterogeneity rely on the existence of multiple continuous derivatives of $g(\cdot,\cdot)$ (e.g., Fernández-Val et al., 2021; Freeman and Weidner, 2023; Beyhum and Mugnier, 2024); see also Remark 4.1 below for a clarification of the role of Assumption 6.

As discussed earlier, these features are highly representative of network models. Moreover, their unique combination makes the asymptotic analysis highly nonstandard, potentially invalidating the previously proposed methods and established statistical guarantees. For this reason, in this paper, we do not attempt to construct an asymptotically normal estimator of β_0 or provide an inference method at the cost of imposing additional assumptions restrictive in network models. Instead, we focus on establishing identification of β_0 by providing a new estimator and showing its consistency in the general setting of the paper.

Remark 4.1. Assumption 6 is not essential for consistency of $\hat{\beta}$. It can be dropped at the cost of increasing the magnitude of the bias of $\hat{\beta}$ from h_n^2 to h_n. Even when Assumption 6 is imposed, bounding the bias of $\hat{\beta}$ is non-trivial. In particular, it requires careful accounting for the agents with ξ "close" to the boundary of its support because kernel smoothing methods suffer from larger biases on the boundary even when the estimated function is smooth.

Remark 4.2 (Extension to $d_\xi > 1$)**.** Importantly, the result of Theorem 1 remains the same for $d_\xi > 1$ provided that the condition $nh_n/\ln n \to \infty$ in Assumption 5(ii) is generalized as $nh_n^{d_\xi}/\ln n \to \infty$, and the other conditions are analogously restated in terms of multivariate ξ, if needed. Also notice that, the rates provided in (4.4) and (4.5) would still implicitly depend on d_ξ through R_n (and the restrictions imposed on h_n by Assumption 5(ii)).

4.2 Rates of uniform convergence for $\hat{d}_{ij}^2$

In this section, we complement the result of Theorem 1 by characterizing R_n, the rate of uniform convergence for $\hat{d}_{ij}^2$ defined in (4.1), for the candidate estimators (3.3) and (3.4) in the homoskedastic and then in the general heteroskedastic settings.

4.2.1 Homoskedastic model

First, we consider the homoskedastic case, i.e., we assume that the idiosyncratic errors satisfy (2.5). As discussed in Section 3.2, in this case, the suggested estimator is given by $\hat{d}_{ij}^2 = \hat{q}_{ij}^2 - 2\hat{\sigma}^2$, and $d_{ij}^2 = q_{ij}^2 - 2\sigma^2$ (see (3.3) and (2.9), respectively). Thus,

$$\max_{i,j\neq i} \left|\hat{d}_{ij}^2 - d_{ij}^2\right| = \max_{i,j\neq i} \left|(\hat{q}_{ij}^2 - q_{ij}^2) - (2\hat{\sigma}^2 - 2\sigma^2)\right| \leqslant \max_{i,j\neq i} \left|\hat{q}_{ij}^2 - q_{ij}^2\right| + \left|2\hat{\sigma}^2 - 2\sigma^2\right|. \tag{4.7}$$

This bound, together with the following lemma, allows us to characterize R_n.

Lemma 1. *Suppose that (2.5) holds and $\mathcal{B}$ is compact. Then, under Assumptions 1 and 2,*

$$\max_{i,j\neq i} \left|\hat{q}_{ij}^2 - q_{ij}^2\right| = O_p\left((\ln n/n)^{1/2}\right), \tag{4.8}$$

$$2\hat{\sigma}^2 - 2\sigma^2 = \overline{G}^2 \min_{i\neq j} \|\xi_i - \xi_j\|^2 + O_p\left((\ln n/n)^{1/2}\right), \tag{4.9}$$

where $\hat{q}_{ij}^2$, q_{ij}^2, $2\hat{\sigma}^2$ are given by (2.7), (2.6), (3.2), and $\overline{G}$ is defined in Assumption 2(iii).

The bounds (4.8) and (4.9) established by Lemma 1 together with (4.7) allow us to provide R_n for the estimator (3.3) in the homoskedastic setting. To make this characterization complete, we also need to provide a bound on $\|\xi_i - \xi_j\|^2$ in (4.9).

In particular, since $\mathcal{E}$ is bounded (Assumption 2(ii)), we can guarantee that

$$\min_{i\neq j} \|\xi_i - \xi_j\|^2 \leqslant C n^{-2/d_\xi} \tag{4.10}$$

for some $C > 0$. While this bound is conservative, it ensures that, for $d_\xi \leq 4$, the contribution of $\|\xi_i - \xi_j\|^2$ is negligible, resulting in the following corollary of Lemma 1.

Corollary 1. *Suppose that the hypotheses of Lemma 1 are satisfied. Then, for $d_\xi \leqslant 4$,*

$$\max_{i,j\neq i} \left|\hat{d}_{ij}^2 - d_{ij}^2\right| = O_p\left((\ln n/n)^{1/2}\right),$$

where $\hat{d}_{ij}^2$ and d_{ij}^2 are given by (3.3) and (2.4), respectively.

Corollary 1 ensures that when the errors are homoskedastic and $d_\xi \leqslant 4$, $\hat{d}_{ij}^2$ given by (3.3) satisfies (4.1) with $R_n = \left(\frac{n}{\ln n}\right)^{1/2}$. Hence, when $h_n \propto R_n^{-1/3} = \left(\frac{\ln n}{n}\right)^{-1/6}$, (4.6) implies that

$$\hat{\beta} - \beta_0 = O_p\left((\ln n/n)^{1/3}\right),$$

so the guaranteed rate of convergence for $\hat{\beta}$ is $\left(\frac{n}{\ln n}\right)^{1/3}$.

Remark 4.3. For $d_\xi > 5$, it is still possible to provide a simple yet *conservative* bound on R_n by combining (4.10) with the result of Lemma 1. In particular, in this case, we can guarantee that $2\hat{\sigma}^2 - 2\sigma^2 = O_p\left(n^{-2/d_\xi}\right)$, and, consequently, we also have

$$\max_{i,j\neq i}\left|\hat{d}_{ij}^2 - d_{ij}^2\right| = O_p\left(n^{-2/d_\xi}\right).$$

We stress that these bounds are loose. With a more detailed analysis of the asymptotic behavior of $\min_{i\neq j}\|\xi_i - \xi_j\|^2$ (which is outside the scope of this paper), these results can be substantially refined.

4.2.2 Model with general heteroskedasticity

In this section, we establish the rate of uniform convergence for $\hat{d}_{ij}^2$ under general heteroskedasticity of the errors. First, we suppose that X is discrete and derive R_n for the estimator given by (3.4)-(3.7). Then, we discuss how this estimator can be modified to accommodate continuously distributed X. Finally, we also provide a general result establishing R_n for $\hat{d}_{ij}^2$ based on a generic estimator of Y_{ij}^*, potentially other than (3.7).

Estimation of d_{ij}^2 when X is discrete

Now we formally derive the rate of uniform convergence for $\hat{d}_{ij}^2$ given by (3.4)-(3.7). This rate crucially depends on the asymptotic properties of the denoising estimator $\hat{Y}_{ij}^*$.

To formally state the asymptotic properties of $\hat{Y}_{ij}^*$, we introduce the following assumption.

Assumption 7.

(i) X is discrete and takes finitely many values $\{x_1, \ldots, x_R\}$;

(ii) there exist positive constants κ and $\overline{\delta}$ such that for all $x \in \text{supp}(X)$, for all $\xi' \in \text{supp}(\xi|X = x)$, $\mathbb{P}(\xi \in B_\delta(\xi')|X = x) \geqslant \kappa\delta^{d_\xi}$ for all positive $\delta \leqslant \overline{\delta}$.

As pointed out before, we suppose that X is discrete and takes finitely many values. Condition (ii) is a weak condition imposed on the conditional distribution of $\xi|X$. If $\xi|X$ is continuously distributed, it is satisfied when the conditional density $f_{\xi|X}(\xi|x)$ is (uniformly) bounded away from zero and its support is not "too irregular".

For any matrix $A \in \mathbb{R}^{n\times n}$, let $\|A\|_{2,\infty} := \max_i \sqrt{\sum_{j=1}^n A_{ij}^2}$. Also let $\hat{Y}^*$ and Y^* denote $n \times n$ matrices with entries given by $\hat{Y}_{ij}^*$ and Y_{ij}^*.

Theorem 2. *Suppose that for all i, $\underline{C}(n\ln n)^{1/2} \leqslant n_i \leqslant \overline{C}(n\ln n)^{1/2}$ for some positive constants $\underline{C}$ and $\overline{C}$. Then, under Assumptions 1, 2, 7, for $\hat{Y}_{ij}^*$ given by (3.7) we have:*

(i) $n^{-1}\|\hat{Y}^* - Y^*\|_{2,\infty}^2 = O_p\left((\ln n/n)^{\frac{1}{2d_\xi}}\right)$;

(ii) $\max_k \max_i |n^{-1}\sum_\ell Y_{k\ell}^*(\hat{Y}_{i\ell}^* - Y_{i\ell}^*)| = O_p\left((\ln n/n)^{\frac{1}{2d_\xi}}\right)$.

Theorem 2 establishes two important asymptotic properties of $\hat{Y}_{ij}^*$. In fact, both results play key roles in bounding R_n, the rate of uniform convergence for $\hat{d}_{ij}^2$.

Part (i) is analogous to the result of Zhang et al. (2017). While Zhang et al. (2017) only consider binary outcomes and do not allow for observed covariates, we extend their result by allows for (i) $d_\xi > 1$, (ii) possibly non-binary outcomes and unbounded idiosyncratic errors, (iii) observed (discrete) covariates X. Part (ii) is new. It allows us to substantially improve on R_n compared to what Part (i) can guarantee individually.[18]

Note that Theorem 2 requires n_i, the number of agents included into $\hat{\mathcal{N}}_i(n_i)$, to grow at $(n\ln n)^{1/2}$ rate. As shown in Zhang et al. (2017), this rate is optimal.[19] In applications, the authors also recommend taking $n_i = C(n\ln n)^{1/2}$ with $C \simeq 1$ and document robustness of their numerical results to the choice of C.

Building on Theorem 2, we now provide rate of uniform convergence for $\hat{d}_{ij}^2$.

Theorem 3. *Suppose that the hypotheses of Theorem 2 are satisfied and $\mathcal{B} = \mathbb{R}^p$. Then,*

$$\max_{i,j\neq i} |\hat{d}_{ij}^2 - d_{ij}^2| = O_p\left((\ln n/n)^{\frac{1}{2d_\xi}}\right),$$

where $\hat{d}_{ij}^2$ and d_{ij}^2 are given by (3.4)-(3.7) and (2.4), respectively.

Theorem 3 establishes the rate of uniform convergence for the proposed estimator of d_{ij}^2. Specifically, it ensures that, for the considered $\hat{d}_{ij}^2$, (4.1) holds with $R_n = \left(\frac{n}{\ln n}\right)^{\frac{1}{2d_\xi}}$. Combined with (4.6), this guarantees that under the appropriate choice of the bandwidth h_n, we have

$$\hat{\beta} - \beta_0 = O_p\left((\ln n/n)^{\frac{1}{3d_\xi}}\right).$$

Thus, when X is discrete, β_0 can be estimated (at least) at $\left(\frac{n}{\ln n}\right)^{\frac{1}{3d_\xi}}$ rate. If ξ is scalar, the guaranteed rate of convergence for $\hat{\beta}$ is the same as in the homoskedastic case.

[18] See Lemma 2 for the comparison of the rates.

[19] The optimal choice of n_i remains the same for $d_\xi > 1$.

Estimation of d_{ij}^2 when X is continuously distributed
The proposed estimator $\hat{Y}_{ij}^*$ given by (3.6)-(3.7) can be straightforwardly modified to allow for continuously distributed (components of) X. Since, in this case, the probability of finding two agents with exactly the same values of X is zero, we have to modify the construction of $\hat{\mathcal{N}}_i(n_i)$ previously provided in (3.6). One natural possibility is to consider

$$\hat{\mathcal{N}}_i(n_i;\delta_n) \coloneqq \{i' : \|X_{i'} - X_i\| \leqslant \delta_n, \mathrm{Rank}(\hat{d}_\infty^2(i,i') | \, \|X - X_i\| \leqslant \delta_n) \leqslant n_i\}. \tag{4.11}$$

The parameter δ_n controls the quality of matching based on X. To guarantee consistency of $\hat{Y}_{ij}^*$, we will need δ_n to converge to zero but slowly enough to ensure that we can still find a growing number of good matches increasingly similar in terms of both X and ξ. Once the constructed neighborhoods are appropriately modified, $\hat{Y}_{ij}^*$ can still be computed as in (3.7), and, with some work, the result of Theorem 2 can be generalized accordingly to allow for continuously distributed covariates.

Remark 4.4. Another possibility to allow for continuously distributed (components of) X is to simply treat it as unobserved, similarly to ξ. In this case, (X,ξ) becomes the effective latent variable and $\hat{\mathcal{N}}_i(n_i)$ can be constructed without conditioning on $X = X_i$. Then the result of Theorem 2 can be applied to the resulting estimator delivering analogous rates of convergence with $d_\xi + d_X$ taking of the place of d_ξ, where d_X denotes the dimension of X. While this construction of $\hat{Y}_{ij}^*$ might not necessarily be optimal, it allows us to formally cover the case when X is continuously distributed by providing analogous consistency results.

Estimation of d_{ij}^2 using general matrix denoising techniques
Theorem 3 establishes uniform consistency of $\hat{d}_{ij}^2$ leveraging the specific structure of $\hat{Y}_{ij}^*$ in (3.7) and requires X to be discrete. To complement Theorem 3 and formally establish identification of d_{ij}^2 and β_0 in the general setting, we will now provide a generic consistency result, which does not require $\hat{Y}_{ij}^*$ to have any particular structure and holds regardless of whether X is discrete or continuously distributed.

Lemma 2. *Suppose that $\hat{Y}^*$ satisfies $n^{-1}\|\hat{Y}^* - Y^*\|_{2,\infty}^2 = O_p(\mathcal{R}_n^{-1})$ for some $\mathcal{R}_n \to \infty$. Also suppose that $\mathcal{B}$ is compact. Then, under Assumptions 1 and 2, we have*

$$\max_{i,j\neq i} |\hat{d}_{ij}^2 - d_{ij}^2| = O_p\left((\ln n/n)^{1/2} + \mathcal{R}_n^{-1/2} \right),$$

where $\hat{d}_{ij}^2$ and d_{ij}^2 are given by (3.4) and (2.4), respectively.

Lemma 2 guarantees that $\hat{d}_{ij}^2$ in (3.4) is uniformly consistent for d_{ij}^2 provided that $\hat{Y}_{ij}^*$ is consistent for Y^*_{ij} in the $(2,\infty)$ norm, and thus it justifies using alternative estimators $\hat{Y}^*_{ij}$ available in the literature.

Together with the result of Theorem 4.4, Lemma 2 delivers consistency of $\hat{\beta}$ and thus formally establishes identification of β_0 in the general setting. In particular, as argued in Remark 4.4, if X is continuously distributed, we can still construct $\hat{Y}^*$ satisfying the requirement of the lemma with $\mathcal{R}_n = \left(\frac{n}{\ln n}\right)^{\frac{1}{2(d_\xi + d_X)}}$. Thus, Lemma 2 guarantees that β_0 can be consistently estimated when X is continuously distributed.

4.3 Uniformly consistent estimation of Y^*_{ij}

One of the contributions of this paper is establishing identification of the error free outcomes Y^*_{ij}'s. In Section 2.4.1, we heuristically argued that Y^*_{ij} is identified. In this section, we construct a uniformly consistent estimator of Y^*_{ij} and, hence, formally prove its identification.

As before, first, we suppose that X is discrete and takes finitely many values. The estimator we propose is an analogue of $\hat{Y}^*_{ij}$ given by (3.7), which we used before to construct $\hat{d}^2_{ij}$. It utilizes exactly the same neighborhoods as in (3.6) but, unlike $\hat{Y}^*_{ij}$, averages over all unique outcomes $Y_{i'j'}$ with $i' \in \hat{\mathcal{N}}_i(n_i)$ and $j' \in \hat{\mathcal{N}}_j(n_j)$.[20] For example, if $\hat{\mathcal{N}}_i(n_i)$ and $\hat{\mathcal{N}}_j(n_j)$ have no elements in common, then the proposed estimator takes a simple form as in (3.8). More generally, for any i and j, let

$$\hat{\mathcal{M}}_{ij} \coloneqq \{(i',j') : i' < j', (i' \in \hat{\mathcal{N}}_i(n_i), j' \in \hat{\mathcal{N}}_j(n_j)) \text{ or } (i' \in \hat{\mathcal{N}}_j(n_j), j' \in \hat{\mathcal{N}}_i(n_i))\}. \tag{4.12}$$

Essentially, $\hat{\mathcal{M}}_{ij}$ is a collection of unique unordered pairs of indices from the Cartesian product of $\hat{\mathcal{N}}_i(n_i)$ and $\hat{\mathcal{N}}_j(n_j)$. Then, Y^*_{ij} is estimated by

$$\tilde{Y}^*_{ij} = \frac{1}{m_{ij}} \sum_{(i',j') \in \hat{\mathcal{M}}_{ij}} Y_{i'j'}, \tag{4.13}$$

where m_{ij} denotes the number of elements in $\hat{\mathcal{M}}_{ij}$.

Theorem 4. *Suppose that the hypotheses of Theorem 2 hold. Suppose that for any $\delta > 0$, there exists $C_\delta > 0$ such that $\int (g(\xi_i,\xi) - g(\xi_j,\xi))^2 dP_\xi(\xi) > C_\delta$ a.s. for $|\xi_i - \xi_j| \geqslant \delta$. Then,*

$$\max_{i,j} |\tilde{Y}^*_{ij} - Y^*_{ij}| = o_p(1).$$

Theorem 4 demonstrates that $\tilde{Y}^*_{ij}$ is uniformly consistent for Y^*_{ij} and, consequently, it formally proves that Y^*_{ij} is identified. We also stress that the previously employed estimator

[20] Recall that in the studied undirected model, $Y_{ij} = Y_{ji}$, and Y_{ij} is not observed for $i = j$.

$\hat{Y}_{ij}^*$ is not necessarily uniformly consistent since it averages over only n_i outcomes $Y_{i'j}$. At the same time, $\tilde{Y}_{ij}^*$ averages over $m_{ij} = O(n_i n_j)$ outcomes, which allows us to establish the desired result.

Remark 4.5. Recall that, as discussed in Section 2.4.1, the similarity distance $\hat{d}_\infty^2(i,j)$ used to construct $\hat{Y}_{ij}^*$ and $\tilde{Y}_{ij}^*$ allows us to find agents i and j similar in terms of the L^2 distance between functions $g(\xi_i,\cdot)$ and $g(\xi_j,\cdot)$. The additional requirement imposed in Theorem 4 ensures that this similarity also translates into similarity between ξ_i and ξ_j, which helps us to establish uniform consistency of $\tilde{Y}_{ij}^*$.[21]

Remark 4.6. Analogous uniform consistency results can also be obtained when X is continuously distributed if $\tilde{Y}_{ij}^*$ is properly adjusted. As previously discussed, the possible adjustments include (i) constructing neighborhoods as in (4.11), or (ii) treating X as unobserved (see Remark 4.4). The result of Theorem 4 can be directly applied to the latter estimator to formally establish identification of Y_{ij}^* in this case.

Remark 4.7. To the best of our knowledge, identifiability of the error free outcomes Y_{ij}^* is a new result to the econometrics literature on identification of network and, more generally, two-way models. Moreover, Theorem 4 also contributes to the statistics literature on graphon and, more generally, the latent space model estimation. Specifically, most of the previous work focused on establishing consistency and deriving rates in terms of the mean squared error (MSE) for $\hat{Y}^*$ (e.g., Chatterjee, 2015; Gao et al., 2015; Klopp et al., 2017; Zhang et al., 2017; Li et al., 2019). Theorem 4 contributes to this literature by establishing $\|\tilde{Y}^* - Y^*\|_{\max} = o_p(1)$, i.e., demonstrating consistency of $\tilde{Y}^*$ in the max norm.

Another implication of Theorem 4 is that the pair-specific fixed effects $g(\xi_i,\xi_j)$ can also be consistently estimated and, hence, are identified for all pair of agents i and j. Consider

$$\hat{g}_{ij} = \tilde{Y}_{ij}^* - W_{ij}'\hat{\beta}, \tag{4.14}$$

where $\tilde{Y}_{ij}^*$ is given by (4.13). Since we have already demonstrated consistency of $\hat{\beta}$ and uniform consistency of $\tilde{Y}_{ij}^*$, $\hat{g}_{ij}$ is also uniformly consistent for $g_{ij} \coloneqq g(\xi_i,\xi_j)$.

Corollary 2. *Suppose that the hypotheses of Theorem 4 are satisfied. Also, suppose that* $\hat{\beta} - \beta_0 = o_p(1)$. *Then,* $\max_{i,j} |\hat{g}_{ij} - g_{ij}| = o_p(1)$, *where* $\hat{g}_{ij}$ *is given by* (4.14).

[21] This condition is a weaker version of Assumption 4(ii). Together with Assumption 2(iii), it allows us to guarantee that the matched agents i and j are also similar in terms of the L^∞ distance $\|g(\xi_i,\cdot) - g(\xi_j,\cdot)\|_\infty$. While it is also possible to characterize the rate of uniform convergence under additional conditions allowing one to translate the L^2 rate into the L^∞ rate for $\|g(\xi_i,\cdot) - g(\xi_j,\cdot)\|$ such as Assumption 4(iii), we do not pursue this direction here because the primary goal of Theorem 4 is establishing identification of Y_{ij}^*.

Establishing nonparametric identification of the pair-specific fixed effects g_{ij}'s is another contribution of the paper. This result is also of high empirical importance since in certain applications, the fixed effects are the primary object of interest.

5 Extensions

5.1 Identification of single index and nonparametric models

In this section, we extend the identification arguments of Section 2 to cover a wide range of network models, both semiparametric and nonparametric, beyond the model (2.1).

First, recall that, as discussed in Section 2.4.1, the error free outcomes are still identified in the most general analogue of (2.1) given by

$$Y_{ij} = f(X_i, \xi_i, X_j, \xi_j) + \varepsilon_{ij}, \quad \mathbb{E}\left[\varepsilon_{ij} | X_i, \xi_i, X_j, \xi_j\right] = 0. \tag{5.1}$$

For example, we can formally establish identification of $Y_{ij}^* = f(X_i, \xi_i, X_j, \xi_j)$ by constructing a version of $\tilde{Y}_{ij}^*$ introduced in (4.13), treating X as unobserved, and and showing its uniform consistency using the result of Theorem 4.

However, identification of Y_{ij}^*, *the value of* $f(X_i, \xi_i, X_j, \xi_j)$, for any pair of agents i and j is fundamentally different from identification of *function* f. Importantly, Y_{ij}^* is not a causal object and cannot be directly employed in counterfactual analysis. Moreover, since ξ is not observed, function f or any features of it relevant for counterfactual analysis cannot be identified unless some additional structure is imposed on f.

For example, in Sections 2-4, we demonstrated that β_0 is identified and can be consistently estimated when $f(X_i, \xi_i, X_j, \xi_j) = W_{ij}'\beta_0 + g(\xi_i, \xi_j)$, which allows us to recover the ceteris paribus effect of W_{ij} (or X_{ij}) on Y_{ij} in this model. Below, we extend these identification results to more general forms of f covering nonlinear and nonparametric models.

5.1.1 Identification of the semiparametric single index model

One empirically relevant generalization of (2.1) is allowing f to have a single index structure

$$f(X_i, \xi_i, X_j, \xi_j) = F(W_{ij}'\beta_0 + g(\xi_i, \xi_j)), \tag{5.2}$$

where $F(\cdot)$ is a known invertible link function. Notice that the presence of the link function $F(\cdot)$ ensures that (5.2) is flexible enough to cover a wide range of the previously studied nonlinear network models. For example, consider the following network formation model

$$Y_{ij} = \mathbb{1}\{W_{ij}'\beta_0 + g(\xi_i, \xi_j) - U_{ij} \geqslant 0\},$$

where Y_{ij} is a binary variable indicating whether agents i and j are connected or not, and U_{ij}'s are iid draws from some distribution, e.g., logistic or $N(0,1)$. This model is covered by (5.2) with $F(\cdot)$ standing for the CDF of U_{ij}. If $F(\cdot) = \exp(\cdot)$, then (5.2) generalizes the dyadic Poisson regression model commonly used to analyze trade networks.

The previously developed arguments can be immediately applied to establish identification of β_0 and to construct an analogue estimator. First, note that since $Y_{ij}^* = F(W_{ij}'\beta_0 + g(\xi_i, \xi_j))$ is identified and $F(\cdot)$ is invertible, we can also identify

$$\mathcal{Y}_{ij}^* := W_{ij}'\beta_0 + g(\xi_i, \xi_j) = F^{-1}(Y_{ij}^*).$$

Next, since $\mathcal{Y}_{ij}^*$ are effectively observed, we are back in the additively separable setting previously studied in Sections 2-4 with $\mathcal{Y}_{ij}^*$ replacing Y_{ij}^*, and all the relevant features of the model such as β_0 and the pair-specific fixed effects $g_{ij} = g(\xi_i, \xi_j)$ are identified for all i and j.

In particular, we can still estimate β_0 as in (3.1) with $\hat{\mathcal{Y}}_{ij}^*$ replacing Y_{ij}, using $\hat{d}_{ij}^2$ given by (3.4) with $\hat{\mathcal{Y}}_{ij}^*$ replacing $\hat{Y}_{ij}^*$. In this case, the preliminary step involves denoising the observed outcomes by constructing $\hat{Y}_{ij}^*$ as before, and then computing $\hat{\mathcal{Y}}_{ij}^* = F^{-1}(\hat{Y}_{ij}^*)$.

Finally, we want to highlight the importance of identification of the pair-specific fixed effects $g_{ij} = g(\xi_i, \xi_j)$ in the model (5.2). Since $F(\cdot)$ is potentially nonlinear, the knowledge of β_0 alone is not sufficient for identifying some policy relevant quantities such as partial effects. However, since the fixed effects g_{ij}'s are identified for all pairs of agents, we can also identify both pair-specific and average partial effects as well as other policy relevant counterfactuals.

5.1.2 Identification of the nonparametric model

In the previously considered models (2.1) and (5.2), the contribution of observables is parameterized by $\beta_0 \in \mathbb{R}^p$, which was the primary object of interest in our analysis. In this section, we consider a nonparametric version of (2.1) with f given by

$$f(X_i, \xi_i, X_j, \xi_j) = h(X_i, X_j) + g(\xi_i, \xi_j), \tag{5.3}$$

where both $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{E} \times \mathcal{E} \to \mathbb{R}$ are unknown symmetric functions.

In Sections 2-4, we considered a special case of (5.3) with $h(X_i, X_j) = w(X_i, X_j)'\beta_0$ and established identification of β_0. In this section, we will show that, in the studied setting, function $h(\cdot,\cdot)$ is nonparametrically identified. Importantly, nonparametric identification of

$h(\cdot,\cdot)$ implies that the previously obtained identification results were not driven by the parametric restrictions or linearity imposed on $h(\cdot,\cdot)$, justifying our focus on the semiparametric model (2.1) as a practical approximation of the general nonparametrically identified model (5.3).

We establish identification of $h(\cdot,\cdot)$ and g_{ij} in (5.3) under the following assumption.

Assumption 8. Suppose that (5.3) holds and

(i) $h: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a symmetric measurable function, and $h(x,x) = 0$ for all $x \in \mathcal{X}$;

(ii) For any $x, \tilde{x} \in \operatorname{supp}(X)$, there exists $\mathcal{E}_{x,\tilde{x}} \subseteq \mathcal{E}$ such that $\mathbb{P}(\xi \in \mathcal{E}_{x,\tilde{x}} | X = x) > 0$ and $\mathbb{P}(\xi \in \mathcal{E}_{x,\tilde{x}} | X = \tilde{x}) > 0$.

Discussion of Assumption 8(i). The requirement $h(x,x) = 0$ is a normalization. Indeed, since $g(\cdot,\cdot)$ and the dimension of ξ are not specified, it is without loss of generality to let $g(\xi_i, \xi_j) = \alpha_i + \alpha_j + \psi(\theta_i, \theta_j)$, where $\psi(\cdot,\cdot)$ is symmetric, and $\xi = (\alpha, \theta')'$. Consider

$$f(X_i, \xi_i, X_j, \xi_j) = h(X_i, X_j) + \alpha_i + \alpha_j + \psi(\theta_i, \theta_j), \tag{5.4}$$

where $h(\cdot,\cdot)$ is symmetric. Then, we can construct an observationally equivalent model with

$$\begin{aligned} \tilde{f}(X_i, \tilde{\xi}_i, X_j, \tilde{\xi}_j) &= \tilde{h}(X_i, X_j) + \tilde{\alpha}_i + \tilde{\alpha}_j + \psi(\theta_i, \theta_j), \\ \tilde{h}(X_i, X_j) &= h(X_i, X_j) - (h(X_i, X_i) + h(X_j, X_j))/2, \end{aligned} \tag{5.5}$$

where $\tilde{\xi}_i = (\tilde{\alpha}_i, \theta_i')'$ with $\tilde{\alpha}_i = \alpha_i + h(X_i, X_i)/2$. Note that $\tilde{h}(\cdot,\cdot)$ is symmetric, continuous, and satisfies $\tilde{h}(x,x) = 0$ for all $x \in \mathcal{X}$. Since $\tilde{f}(X_i, \tilde{\xi}_i, X_j, \tilde{\xi}_j) = f(X_i, \xi_i, X_j, \xi_j)$ for any pair of agents i and j, the normalized model (5.5) is equivalent to the original model (5.4). ■

Remark 5.1. While the normalization introduced in Assumption 8(i) is not the only possible one, it is natural for network models, especially when $h(X_i, X_j)$ captures homophily based on observables (e.g., similar normalizations are also imposed in Toth, 2017 and Gao, 2020).

First, we argue that $h(x, \tilde{x})$ is identified for any fixed $x, \tilde{x} \in \operatorname{supp}(X)$. Specifically, fix $X_i = x$ and $X_j = \tilde{x}$, and consider

$$\begin{aligned} d_{ij}^2(\mu; x, \tilde{x}) \coloneqq & \mathbb{E}\left[\left(Y_{ik}^* - Y_{jk}^* + \mu\right)^2 | (X_i, X_j, X_k) = (x, \tilde{x}, x), \xi_i, \xi_j\right] \\ & + \mathbb{E}\left[\left(Y_{ik}^* - Y_{jk}^* - \mu\right)^2 | (X_i, X_j, X_k) = (x, \tilde{x}, \tilde{x}), \xi_i, \xi_j\right], \\ d_{ij}^2(x, \tilde{x}) \coloneqq & \min_{\mu \in \mathbb{R}} d_{ij}^2(\mu; x, \tilde{x}), \qquad \mu_{ij}^*(x, \tilde{x}) \coloneqq \operatorname*{argmin}_{\mu \in \mathbb{R}} d_{ij}^2(\mu; x, \tilde{x}). \end{aligned} \tag{5.6}$$

Note that since Y^*_{ik} and Y^*_{jk} are identified in the general model (5.1), $\mathbf{d}^2_{ij}(x, \tilde{x})$ and $\mu^*_{ij}(x, \tilde{x})$ are also identified.

$\mathbf{d}^2_{ij}(x, \tilde{x})$ is a nonparametric analogue of the previously considered pseudo-distance d^2_{ij}. Since we are interested in identifying $\mathbf{h}(x, \tilde{x})$, we additionally condition on $(X_i, X_j) = (x, \tilde{x})$ and only consider $X_k = x$ and $X_k = \tilde{x}$. Notice that when $\xi_i = \xi_j$, then we have

$$\mathbf{d}^2_{ij}(x, \tilde{x}) = \min_{\mu \in \mathbb{R}} \left((-\mathbf{h}(x, \tilde{x}) + \mu)^2 + (\mathbf{h}(x, \tilde{x}) - \mu)^2 \right) = 0,$$

where the minimum is achieved at $\mu^*_{ij}(x, \tilde{x}) = \mathbf{h}(x, \tilde{x})$. The following lemma establishes that the converse is also true: if $\mathbf{d}_{ij}(x, \tilde{x})^2 = 0$, then we also necessarily have $\mu^*_{ij}(x, \tilde{x}) = \mathbf{h}(x, \tilde{x})$.

Lemma 3. *Suppose that Assumption 8 holds, and that the expectations in (5.6) exist for any agents i and j, and for any $x, \tilde{x} \in \mathcal{X}$. Then, for any $x, \tilde{x} \in$ supp(X), $x \neq \tilde{x}$, $\mathbf{d}^2_{ij}(x, \tilde{x}) = 0$ implies that $\mu^*_{ij}(x, \tilde{x}) = \mathbf{h}(x, \tilde{x})$.*

Lemma 3 establishes identification of $\mathbf{h}(x, \tilde{x})$ by showing that $\mathbf{h}(x, \tilde{x}) = \mu^*_{ij}(x, \tilde{x})$ for all agents i and j, with $X_i = x$ and $X_j = \tilde{x}$, satisfying $\mathbf{d}_{ij}(x, \tilde{x}) = 0$; notice that Assumption 8(ii) guarantees that such agents exist. Moreover, since the value of $\mu^*_{ij}(x, \tilde{x})$ should be the same for all such i and j (and this is true for any fixed $x, \tilde{x} \in \text{supp}(X)$), the nonparametric model (5.3) as well as the previously considered models including (2.1) are overidentified, and, in principle, can be falsified.

Next, notice that since $\mathbf{h}(X_i, X_j)$ is identified for any pair of agents i and j, we can also identify the pair-specific fixed effect as $g_{ij} = Y^*_{ij} - \mathbf{h}(X_i, X_j)$. As a result, we conclude that both $\mathbf{h}$ and the pair-specific fixed effects are nonparametrically identified in model (5.3).

Finally, consider a nonlinear single index version of (5.3)

$$f(X_i, \xi_i, X_j, \xi_j) = F(\mathbf{h}(X_i, X_j) + g(\xi_i, \xi_j)), \tag{5.7}$$

where $F(\cdot)$ is a known and invertible link function. This model is a nonparametric version of (5.2) considered in Section 5.1.1. By first constructing $\mathcal{Y}^*_{ij} = F^{-1}(Y^*_{ij}) = \mathbf{h}(X_i, X_j) + g(\xi_i, \xi_j)$ and then applying the results of this section, we conclude that $\mathbf{h}$, the fixed effects g_{ij}'s, as well as both pair-specific and average partial effects are also identified in (5.7).

5.2 Incorporating missing outcomes

From the beginning of Section 2, to simplify the exposition and facilitate the formal analysis, we assumed that $\{Y_{ij}\}_{i \neq j}$ are observed for all pairs of agents i and j. While this assumption is standard in the network formation context, where the absence of an interaction between agents i and j is still recorded as observing $Y_{ij} = 0$, in many other applications, interaction outcomes are available only for a limited number of pairs of agents (e.g., in the matched employer-employee setting). Hence, it is important to discuss (i) how to properly adjust the constructed estimators to account for missing outcomes, and (ii) under which conditions the proposed method remains valid in this case. For simplicity, we will stick with considering an undirected network as before. We will discuss directed networks and general two-way settings covering the important matched employer-employee example in Section 5.3 below.

Let D_{ij} be a binary variable such that $D_{ij} = 1$ if Y_{ij} is observed and $D_{ij} = 0$ otherwise, with D denoting the resulting adjacency matrix (by construction, $D_{ii} = 0$). Also, let $\mathcal{O}_{ij} := \{k : D_{ik} = D_{jk} = 1\}$ denote a set of agents k such that Y_{ik} and Y_{jk} are observed.

To fix ideas, consider the homoskedastic estimator first. We adjust (2.7) and (3.1) as

$$\hat{q}_{ij}^2 = \min_{\beta \in \mathcal{B}} \frac{1}{|\mathcal{O}_{ij}|} \sum_{k \in \mathcal{O}_{ij}} (Y_{ik} - Y_{jk} - (W_{ik} - W_{jk})'\beta)^2, \tag{5.8}$$

$$\hat{\beta} = \left(\sum_{i<j} K\left(\frac{\hat{d}_{ij}^2}{h_n^2}\right) \sum_{k \in \mathcal{O}_{ij}} \Delta W_{ijk} \Delta W_{ijk}' \right)^{-1} \left(\sum_{i<j} K\left(\frac{\hat{d}_{ij}^2}{h_n^2}\right) \sum_{k \in \mathcal{O}_{ij}} \Delta W_{ijk} \Delta Y_{ijk} \right), \tag{5.9}$$

where $\hat{d}_{ij}^2$ is still computed as in (3.3), and $|\mathcal{O}_{ij}|$ denotes the cardinality of $\mathcal{O}_{ij}$. Notice that to consistently estimate q_{ij}^2, we need $|\mathcal{O}_{ij}| \to \infty$. Hence, in practice one may want to limit their attention to pairs of agents for which $|\mathcal{O}_{ij}|$ is sufficiently large for $\hat{q}_{ij}^2$ to be informative.

Next, we want to discuss under which conditions the proposed method, with appropriate modifications as described above, remains valid when some interaction outcomes are missing. First, the selection mechanism needs to be exogenous conditional on the observed and unobserved characteristics of agents $\{(X_i, \xi_i)\}_{i=1}^n$, meaning that the D should be (conditionally) independent of the errors $\{\varepsilon_{ij}\}$.[22] This assumption is standard in the network regression literature including the seminal AKM model of Abowd et al. (1999) and subsequent work, assuming exogeneity of the employer-employee network conditional on the firms' and workers' *additive* fixed effect (and their observed characteristics). While this assumption is widely used in the AKM literature, it is also often criticized for severely restricting the network formation process. We want to stress that, since we allow for a much more general form of unobserved heterogeneity than the additive AKM model, the network exogeneity assumption is substantially less restrictive in our setting. Specifically, since we do not specify the dimensionality of the fixed effects and their role in the model (e.g., we allow for complementarity between the firm and worker fixed effects), our

[22] This assumption is satisfied if agents form connections based on $\{(X_i, \xi_i)\}_{i=1}^n$ but not on the idiosyncratic errors. For example, in a structural model, it can be rationalized if ε_{ij}'s are drawn after the network is formed.

framework allows for a much broader class of selection mechanism compatible with the exogeneity assumption. Thus, rather than seeing network endogeneity as a potential threat to the validity of our approach, we consider it to be a new tool addressing this concern.

Second, for $\hat{\beta}$ in (5.9) to be consistent, we need to have a growing number of pairs i and j, for which $|\mathcal{O}_{ij}| \to \infty$ as $n \to \infty$. Specifically, the requirement $|\mathcal{O}_{ij}| \to \infty$ ensures that we can consistently estimate $\hat{d}_{ij}^2$ for a growing group of agents. At the same time, a growing pool of potential matches allows us to find increasingly similar agents controlling the bias of $\hat{\beta}$ and ensuring its consistency. Notice that this requirement still allows the network to be sparse, and that its adequacy can be evaluated in a given application. Moreover, in Section 5.3, we will argue that this requirement becomes even less restrictive in general two-way settings and is plausible for many matched employer-employee data sets.

In the general heteroskedastic case, the first estimation step is to construct $\hat{Y}_{ij}^*$. In fact, recent developments in the matrix completion literature allow one to consistently estimate Y^* (in terms of the MSE), even when the observed matrix Y is sparse (e.g., Chatterjee, 2015; Klopp et al., 2017; Li et al., 2019). Once $\hat{Y}_{ij}^*$ are constructed (for example, using one of the already developed matrix completion techniques), the rest of the estimation procedure remains the same. We provide an appropriate modification of the previously used estimator of $\hat{Y}_{ij}^*$ and discuss estimation of β_0 in more detail in Appendix C.

5.3 Extension to directed networks and two-way models

Finally, the proposed estimation procedure can also be generalized to cover directed networks and, more generally, two-way models. Specifically, consider a general interaction model

$$Y_{ik} = W_{ik}'\beta_0 + g(\xi_i, \eta_k) + \varepsilon_{ik},$$

where $i \in \mathcal{I}$ and $k \in \mathcal{K}$ index senders and receivers, and ξ_i and η_k denote the sender and receiver fixed effects. As in Section 5.2, we can also allow for missing interactions, with Y_{ik} observed whenever $D_{ik} = 1$.

As before, our identification and estimation strategies are based on finding agents with similar values of the fixed effects. However, the considered interaction model consists of two types of agents, senders and receivers (e.g., firms and workers). As a result, we have the flexibility to decide whether we want to match senders or receivers depending on the context. For example, consider the sender-to-sender approach. To fix ideas, we also focus on the homoskedastic setting (the general heteroskedastic estimator can be constructed in a similar fashion). In this case, the sender-to-sender estimator of β_0 has exactly the same form as in (5.8)-(5.9) previously provided in Section 5.2

As discussed in the previous section, for the sender-to-sender estimator to be consistent, we need a growing pool of senders that we can potentially match satisfying the same requirement $|\mathcal{O}_{ij}| \to \infty$, so we can consistently determine if senders i and j are similar in terms of ξ or not. Also notice that in this case receivers are allowed to participate only in a few interactions. For example, this suggests that in the matched employer-employee setting, the firm-to-firm approach can be appropriate even when the workers' mobility is limited, so long as we have a sufficient number of pairs of firms with a sufficient number of workers moving from one of them to another.

6 Numerical Evidence

6.1 Numerical experiment in a homoskedastic model

In this section, we illustrate the finite sample properties of the proposed estimators. Specifically, we consider the following homoskedastic variation of (2.1):

$$Y_{ij} = (X_i - X_j)^2\beta_0 - (\xi_i - \xi_j)^2 + \varepsilon_{ij}, \quad \begin{pmatrix} X_i \\ \xi_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{6.1}$$

where $\{\varepsilon_{ij}\}_{i<j}$ are independent draws from $N(0,1)$. The true value of the parameter of interest is $\beta_0 = -1$, so the considered model features homophily based on both X and ξ.

We study the performance of the following estimators. The first estimator $\hat{\beta}_{\text{FE}}$ is produced by the standard linear regression with additive fixed effects. The second estimator $\hat{\beta}$ is the kernel based estimator (3.1) with $\hat{d}^2_{ij}$ computed as in (3.3) and using the Epanechnikov kernel.[23] We choose $h_n^2 = 0.9 \min\{\hat{\sigma}_{\hat{d}^2}, \text{IQR}_{\hat{d}^2}/1.349\}\binom{n}{2}^{-1/5}$ following the standard (kernel density estimation) rule of thumb. Here $\hat{\sigma}_{\hat{d}^2}$ and $\text{IQR}_{\hat{d}^2}$ stand for the standard deviation and the interquartile range of the estimated pseudo-distances $\{\hat{d}^2_{ij}\}_{i<j}$, and $\binom{n}{2}$ corresponds to the number of the estimated pseudo-distances.[24] Finally, we also compute the 1 nearest neighbor pairwise-difference estimator $\hat{\beta}_{\text{NN1}}$, which, instead of using kernel weights as in (3.1), matches every unit i with exactly one unit j closest to it in terms of $\hat{d}^2_{ij}$.

We simulate the model (6.1) for $n \in \{30, 50, 100\}$ and $\rho \in \{0, 0.3, 0.5, 0.7\}$. The simulated finite sample properties of the considered estimators are reported in Table 1 below. The number of replications is 10,000. The naive estimator $\hat{\beta}_{\text{FE}}$ is biased whenever the observed and unobserved characteristics of agents are correlated. The magnitude of this

[23]The reported results are robust to the kernel choice.

[24]Note that since the kernel weights in (3.1) are $K(\frac{\hat{d}^2_{ij}}{h_n^2})$, the "effective" kernel density estimation bandwidth applied to $\{\hat{d}^2_{ij}\}_{i<j}$ is h_n^2, not h_n.

bias increases rapidly as ρ grows. The proposed estimators $\hat{\beta}$ and $\hat{\beta}_{\text{NN1}}$ effectively remove the bias even in networks of a moderate size with $n = 30$. Notice that the magnitudes of the bias for $\hat{\beta}$ and $\hat{\beta}_{\text{NN1}}$ are approximately the same but the kernel based estimator $\hat{\beta}$ is consistently less dispersed. This might suggest that in the studied setting, the 1 nearest neighbor estimator $\hat{\beta}_{\text{NN1}}$ tends to undersmooth. Finally, notice that the proposed estimators $\hat{\beta}$ and $\hat{\beta}_{\text{NN1}}$ dominate the naive estimator $\hat{\beta}_{\text{FE}}$ not only in terms of the bias but also in terms of the standard deviation/IQR even when $\rho = 0$, i.e., when $\hat{\beta}_{\text{FE}}$ is consistent. Indeed, when the fixed effects contribution to the variability in Y_{ij} is large, controlling for the unobservables (as both $\hat{\beta}$ and $\hat{\beta}_{\text{NN1}}$ do by differencing them out) can substantially improve precision even at the cost of significantly reducing the effective sample size.

Table 1: Simulation results for model (6.1)

	Bias			Med Bias			Std. Dev.			IQR/1.349		
ρ	$\hat{\beta}_{\text{FE}}$	$\hat{\beta}$	$\hat{\beta}_{\text{NN1}}$	$\hat{\beta}_{\text{FE}}$	$\hat{\beta}$	$\hat{\beta}_{\text{NN1}}$	$\hat{\beta}_{\text{FE}}$	$\hat{\beta}$	$\hat{\beta}_{\text{NN1}}$	$\hat{\beta}_{\text{FE}}$	$\hat{\beta}$	$\hat{\beta}_{\text{NN1}}$
$n = 30$												
0.0	0.000	-0.001	0.010	0.010	0.000	0.006	0.063	0.028	0.050	0.044	0.025	0.036
0.3	-0.090	-0.006	0.005	-0.061	-0.005	0.004	0.124	0.033	0.060	0.106	0.029	0.044
0.5	-0.251	-0.018	-0.009	-0.225	-0.016	-0.006	0.173	0.045	0.077	0.165	0.038	0.059
0.7	-0.491	-0.052	-0.058	-0.473	-0.048	-0.048	0.196	0.080	0.109	0.191	0.063	0.092
$n = 50$												
0.0	-0.000	-0.000	0.005	0.006	-0.000	0.003	0.035	0.015	0.029	0.026	0.014	0.021
0.3	-0.090	-0.004	0.004	-0.072	-0.004	0.002	0.090	0.018	0.036	0.085	0.016	0.027
0.5	-0.251	-0.013	-0.005	-0.235	-0.012	-0.004	0.130	0.024	0.046	0.128	0.022	0.036
0.7	-0.491	-0.036	-0.038	-0.481	-0.034	-0.032	0.148	0.042	0.068	0.147	0.037	0.057
$n = 100$												
0.0	-0.000	-0.000	0.003	0.003	-0.000	0.002	0.017	0.007	0.014	0.012	0.007	0.010
0.3	-0.091	-0.003	0.002	-0.082	-0.002	0.002	0.061	0.008	0.018	0.058	0.008	0.014
0.5	-0.251	-0.008	-0.002	-0.243	-0.007	-0.002	0.089	0.011	0.024	0.087	0.010	0.019
0.7	-0.491	-0.021	-0.020	-0.486	-0.021	-0.017	0.102	0.019	0.036	0.100	0.018	0.030

This table reports the simulated bias, median bias, standard deviation, and interquartile range (divided by 1.349) for the additive fixed effects estimator $\hat{\beta}_{\text{FE}}$, the kernel estimator $\hat{\beta}$, and the 1 nearest neighbor estimator $\hat{\beta}_{\text{NN1}}$. The results are presented for different values n (network size) and ρ (correlation between X_i and ξ_i). The number of replications is 10,000.

6.2 Empirical illustration: homophily in online social networks

In this section, we illustrate the usefulness of our method in the context of estimating homophily in online social networks using the Facebook 100 data set (Traud et al., 2012). This dataset contains Facebook friendship network as well as nodal covariates (gender, major, graduation year, etc.) collected at 100 colleges and universities in the US in 2005.

Specifically, we consider the following logistic network formation model

$$Y_{ij} = \mathbb{1}\{\mathbb{1}\{X_i = X_j\}\beta_0 + g(\xi_i, \xi_j) - U_{ij} \geqslant 0\}, \tag{6.2}$$

where Y_{ij} is a binary variable indicating whether students i and j are friend or not, X_i is a gender dummy, and $\{U_{ij}\}_{i<j}$ are iid draws from the logistic distribution. In this model, β_0 captures gender homophily. Notice that (6.2) can be represented in the regression form as

$$Y_{ij} = \Lambda(\mathbb{1}\{X_i = X_j\}\beta_0 + g(\xi_i, \xi_j)) + \varepsilon_{ij},$$

where $\Lambda(\cdot)$ stands for the logistic CDF. Note that in the studied network formation model the conditional mean is *nonlinear* and the errors ε_{ij}'s are *heteroskedastic*.

We estimate β_0 using the following three estimators. The first is $\hat{\beta}_{\text{MLE}}$, the naive MLE estimator that ignores unobserved heterogeneity. The second is $\hat{\beta}_{\text{TL}}$, the tetra-logit estimator introduced in Graham (2017) allowing for additive fixed effects, i.e., assuming that $g(\xi_i, \xi_j) = \xi_i + \xi_j$. Finally, we also construct the kernel estimator $\hat{\beta}$ following the procedure described in Section 5.1.1. Specifically, we use $n_i \approx 0.5(n \ln n)^{1/2}$ for all i for constructing Y^*_{ij}, and we choose K and h_n^2 as described in Section 6.1.

We report results for Princeton University, the class of 2004 (the results are qualitatively similar across the universities and cohorts). This network consists of 541 students with the mean degree of 37.6, i.e., on average the students in this sample are connected with about 7% of the whole class, so the studied network is characteristically sparse.

The results are reported in Table 2 below. Both the MLE and tetra-logit homophily estimates are much higher compared to the one produced by $\hat{\beta}$. This is in line with one of the well known perils of estimating homophily effects: naive methods are likely to overestimate homophily associated with observables when agents also exhibit homophily based on their latent characteristics. For example, if students meet their friends in classrooms and gender is predictive of their choices of classes, naive methods might misattribute this effect to gender homophily.

However, since we do not have a confidence interval available for $\hat{\beta}$, it is not immediately clear if the difference in the produced estimates is systematic or simply attributed to the large sampling uncertainty. In order to address this concern and study the performance of our method in an empirically relevant setting, we perform the following numerical experiment designed to mimic the studied application. Specifically, we consider a version of (6.2)

$$Y_{ij} = \mathbb{1}\{\mathbb{1}\{X_i = X_j\}\beta_0 + \alpha_0 - \kappa_0 |\xi_i - \xi_j| - U_{ij} \geqslant 0\},$$
$$X_i \sim Bernoulli(0.5), \quad \xi_i = \pi_0 X_i + V_i, \quad V_i \sim U([-1, 1]),$$

where (X_i, ξ_i, V_i) are iid. We choose $(\alpha_0, \beta_0, \kappa_0, \pi_0) = (-1.5, 0.1, 2, 0.2)$ and $n = 550$ to match the main features of the data including the average node degree, its standard deviation, as well as the estimates produced by the methods. Note that this model features homophily based on both X and ξ, and ignoring the latter would result in overestimation of β_0.

The simulation results are also reported in Table 2 below. First, as expected, $\hat{\beta}_{\text{MLE}}$ and $\hat{\beta}_{\text{TL}}$ indeed severely overestimate β_0, and their biases are substantially larger than their standard deviations. At the same time our estimator $\hat{\beta}$ does not suffer from this bias and correctly estimates the true effect. Moreover, we also find that, in the studied setting, its standard deviation is comparable to the standard deviations of the other estimators.

In summation, this numerical experiment demonstrates that our method can perform well in an empirically relevant setting featuring nonlinearity of the regression function, heteroskedasticity of the errors, and sparsity representative of real world social networks. Its results also support our empirical finding documenting that the standard approaches overestimate the gender homophily effects and suggest that the difference in the estimates is systematic and not likely to be (entirely) attributed to the sampling variability.

Table 2: Actual data and empirically calibrated simulation results

	Actual Data		Simulation Results			
	Estimate	95% CI	Bias	Med. Bias	Std. Dev.	IQR/1.349
$\hat{\beta}_{\text{MLE}}$	0.1792	[0.1387, 0.2196]	0.0814	0.0802	0.0297	0.0307
$\hat{\beta}_{\text{TL}}$	0.1862	[0.1419, 0.2306]	0.0849	0.0830	0.0301	0.0310
$\hat{\beta}$	0.1106		-0.0060	-0.0060	0.0321	0.0339

This table reports the estimates (and whenever available confidence intervals) of the homophily parameter β_0 in model (6.2) for the naive MLE estimator $\hat{\beta}_{\text{MLE}}$, the tetra-logit estimator $\hat{\beta}_{\text{TL}}$, and the kernel estimator $\hat{\beta}$ described in Section 5.1.1. It also reports the simulated bias, median bias, standard deviation, and interquartile range (divided by 1.349) for the same estimators in the considered empirically calibrated numerical experiment. The number of replications is 10,000.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598.

Amiti, M. and Weinstein, D. E. (2018). How much do idiosyncratic bank shocks affect investment? Evidence from matched bank-firm loan data. *Journal of Political Economy*, 126(2):525–587.

Anderson, J. E. and Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, 93(1):170–192.

Armstrong, T. B., Weidner, M., and Zeleneev, A. (2022). Robust estimation and inference in panels with interactive fixed effects. *arXiv preprint arXiv:2210.06639*.

Athey, S. and Imbens, G. (2025). Identification of average treatment effects in nonparametric panel models. *arXiv preprint arXiv:2503.19873*.

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

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Beyhum, J. and Mugnier, M. (2024). Inference after discretizing time-varying unobserved heterogeneity. *arXiv preprint arXiv:2412.07352*.

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073.

Bickel, P. J., Chen, A., and Levina, E. (2011). The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5):2280–2301.

Bonhomme, S., Lamadon, T., and Manresa, E. (2022). Discretizing unobserved heterogeneity. *Econometrica*, 90(2):625–643.

Candelaria, L. E. (2016). A Semiparametric Network Formation Model with Multiple Linear Fixed Effects. Working paper, Duke University.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, 128(3):967–1015.

Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *The Econometrics Journal*, 20(3):S1–S13.

Chatterjee, S. (2015). Matrix estimation by Universal Singular Value Thresholding. *The Annals of Statistics*, 43(1):177–214.

Chen, M., Fernández-Val, I., and Weidner, M. (2021). Nonlinear factor models for network and panel data. *Journal of Econometrics*, 220(2):296–324.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.

De Paula, Á. (2020). Econometric models of network formation. *Annual Review of Economics*, 12(1):775–799.

Deaner, B., Hsiang, C.-W., and Zeleneev, A. (2025). Inferring treatment effects in large panels by uncovering latent similarities. *arXiv preprint arXiv:2503.20769*.

Dzemski, A. (2018). An Empirical Model of Dyadic Link Formation in a Network with Unobserved Heterogeneity. *The Review of Economics and Statistics*, pages 1–14.

Fernández-Val, I., Freeman, H., and Weidner, M. (2021). Low-rank approximations of nonseparable panel models. *The Econometrics Journal*, 24(2):C40–C77.

Finkelstein, A., Gentzkow, M., and Williams, H. (2016). Sources of geographic variation in health care: Evidence from patient migration. *Quarterly Journal of Economics*, 131(4):1681–1726.

Freeman, H. and Weidner, M. (2023). Linear panel regressions with two-way unobserved heterogeneity. *Journal of Econometrics*, 237(1):105498.

Gao, C., Lu, Y. U., and Zhou, H. H. (2015). Rate-optimal graphon estimation. *The Annals of Statistics*, 43(6):2624–2652.

Gao, W. Y. (2020). Nonparametric identification in index models of link formation. *Journal of Econometrics*, 215(2):399–413.

Gao, W. Y., Li, M., and Xu, S. (2023). Logical differencing in dyadic network formation models with nontransferable utilities. *Journal of Econometrics*, 235(1):302–324.

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

Graham, B. and De Paula, Á. (2020). *The econometric analysis of network data.* Academic Press.

Graham, B. S. (2017). An Econometric Model of Network Formation With Degree Heterogeneity. *Econometrica*, 85(4):1033–1063.

Graham, B. S. (2020). Chapter 2 - network data. In Durlauf, S. N., Hansen, L. P., Heckman, J. J., and Matzkin, R. L., editors, *Handbook of Econometrics, Volume 7A*, volume 7 of *Handbook of Econometrics*, pages 111–218. Elsevier.

Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354.

Hanushek, E. A., Kain, J. F., Markman, J. M., and Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18(5):527–544.

Helpman, E., Itskhoki, O., Muendler, M. A., and Redding, S. J. (2017). Trade and inequality: From theory to estimation. *Review of Economic Studies*, 84(1):357–405.

Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *Quarterly Journal of Economics*, 123(2):441–487.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. Preprint, Institute for Advanced Study, Princeton, NJ.

Jochmans, K. (2018). Semiparametric Analysis of Network Formation. *Journal of Business and Economic Statistics*, 36(4):705–713.

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

Kallenberg, O. (2005). *Probabilistic symmetries and invariance principles.* Springer, New York.

Kitamura, Y. and Laage, L. (2024). Estimating stochastic block models in the presence of covariates. *arXiv preprint arXiv:2402.16322*.

Klopp, O., Tsybakov, A. B., and Verzelen, N. (2017). Oracle inequalities for network models and sparse graphon estimation. *The Annals of Statistics*, 45(1):316–354.

Li, Y., Shah, D., Song, D., and Yu, C. L. (2019). Nearest neighbors for matrix estimation interpreted as blind regression for latent variable model. *IEEE Transactions on Information Theory*, 66(3):1760–1784.

Lovász, L. (2012). *Large Networks and Graph Limits*, volume 60. American Mathematical Soc.

Ma, S., Su, L., and Zhang, Y. (2022). Detecting latent communities in network formation models. *Journal of Machine Learning Research*, 23(310):1–61.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444.

Mele, A., Hao, L., Cape, J., and Priebe, C. E. (2023). Spectral estimation of large stochastic blockmodels with discrete nodal covariates. *Journal of Business & Economic Statistics*, 41(4):1364–1376.

Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579.

Mugnier, M. (2025). A simple and computationally trivial estimator for grouped fixed effects models. *Journal of Econometrics*, 250:106011.

Nowakowicz, K. (2024). Nonparametric network bootstrap.

Orbanz, P. and Roy, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE transactions on pattern analysis and machine intelligence*, 37(2):437–461.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458.

Rothstein, J. (2010). Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *The Quarterly Journal of Economics*, 125(1):175–214.

Santos Silva, J. M. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4):641–658.

Shalizi, C. R. and Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research*, 40(2):211–239.

Song, J., Price, D. J., Guvenen, F., Bloom, N., and Von Wachter, T. (2019). Firming up inequality. *Quarterly Journal of Economics*, 134(1):1–50.

Starck, V. (2025). Improving control over unobservables with network data. *arXiv preprint arXiv:2511.00612*.

Toth, P. (2017). Semiparametric Estimation in Networks with Homophily and Degree Heterogeneity. Working paper, University of Nevada.

Traud, A. L., Mucha, P. J., and Porter, M. A. (2012). Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165–4180.

Yan, T., Jiang, B., Fienberg, S. E., and Leng, C. (2019). Statistical Inference in a Directed Network Model With Covariates. *Journal of the American Statistical Association*, 114(526):857–868.

Zeleneev, A. and Zhang, W. (2025). Tractable estimation of nonlinear panels with interactive fixed effects. *arXiv preprint arXiv:2511.15427*.

Zhang, Y., Levina, E., and Zhu, J. (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika*, 104(4):771–783.

Online Supplementary Appendix to "Identification and Estimation of Network Models with Nonparametric Unobserved Heterogeneity"

Andrei Zeleneev

February 5, 2026

A Proofs

A.1 Proofs of the results of Section 4.1

A.1.1 Auxiliary lemmas

Lemma A.1. *Let $c_{ij} \coloneqq c(X_i, X_j, \xi_i)$, where $c(X_i, X_j, \xi_i)$ is defined in Assumption 4(iii). Then, under hypotheses of Theorem 1, there exists $\alpha > 0$ such that, for sufficiently large n, (i) $|\xi_j - \xi_i| > \alpha h_n$ implies that $K\left(d^2(X_i, X_j, \xi_i, \xi_j)/h_n^2\right) = 0$ and $K\left(c_{ij}(\xi_j - \xi_i)^2/h_n^2\right) = 0$ with probability one; (ii) $\sum_{i<j} K(\hat{d}_{ij}^2/h_n^2)^2 \mathbb{1}\{|\xi_j - \xi_i| > \alpha h_n\} = 0$ with probability approaching one; (iii) $|r_K(X_i, X_j, \xi_i, \xi_j; h_n)| \leqslant C h_n \mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\}$ a.s. for some $C > 0$, where $r_K(X_i, X_j, \xi_i, \xi_j; h_n) \coloneqq K(d^2(X_i, X_j, \xi_i, \xi_j)/h_n^2) - K(c_{ij}(\xi_j - \xi_i)^2/h_n^2)$.*

Proof of Lemma A.1. Note that Assumption 4(iii) guarantees that there exists some $\delta_0 > 0$ and $c^* \in (0, \underline{c})$ such that $d^2(X_i, X_j, \xi_i, \xi_j) \geqslant c^*(\xi_j - \xi_i)^2$ a.s. for $|\xi_j - \xi_i| \leqslant \delta_0$. This implies that there exists $\alpha = 1/\sqrt{c^*}$ such that $d^2(X_i, X_j, \xi_i, \xi_j) > h_n^2$ whenever $\alpha h_n < |\xi_j - \xi_i| \leqslant \delta_0$ (notice that $\alpha h_n < \delta_0$ for large enough n). At the same, by Assumption 4(ii), time, for large enough n, $d^2(X_i, X_j, \xi_i, \xi_j) > h_n^2$ a.s. for $|\xi_i - \xi_j| \geqslant \delta_0$. Hence, we conclude that $|\xi_j - \xi_i| > \alpha h_n$ implies $d^2(X_i, X_j, \xi_i, \xi_j) > h_n^2$ and, by Assumption 5(i), $K\left(d^2(X_i, X_j, \xi_i, \xi_j)/h_n^2\right) = 0$. Similarly, since $c^* < \underline{c}$, $|\xi_j - \xi_i| > \alpha h_n$ also implies $c_{ij}(\xi_j - \xi_j)^2 > h_n^2$ and $K\left(c_{ij}(\xi_j - \xi_i)^2/h_n^2\right) = 0$ a.s. This completes the proof of part (i).

Clearly, we can also choose α such that $d^2(X_i, X_j, \xi_i, \xi_j) > c h_n^2$ for some $c > 1$ whenever $|\xi_j - \xi_i| > \alpha h_n$. Consequently, for all pairs of agents i and j satisfying $|\xi_j - \xi_i| > \alpha h_n$, we have $d_{ij}^2/h_n^2 > c > 1$. Using (4.1) and Assumption 5(ii), we also conclude that, for all pairs of agents i and j satisfying that $|\xi_j - \xi_i| > \alpha h_n$, with probability approaching one we also have $\hat{d}_{ij}^2/h_n^2 > 1$. Combining this with Assumption 5(i) completes the proof of part (ii).

Finally, notice that using the result of part (i), we obtain

$$r_K(X_i, X_j, \xi_i, \xi_j; h_n) = \left(K\left(\frac{d^2(X_i, X_j, \xi_i, \xi_j)}{h_n^2}\right) - K\left(\frac{c_{ij}(\xi_j - \xi_i)^2}{h_n^2}\right)\right) \mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\}.$$

Combining this with Assumptions 4(iii) and 5(i) delivers the required bound. Q.E.D.

Lemma A.2. *Suppose the hypotheses of Theorem 1 are satisfied. Then, for any* $\alpha > 0$*, we have: (i)* $h_n^{-1}\mathbb{E}\left[\mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\}|\xi_i\right] \leqslant C_\alpha$ *a.s. and, hence,* $h_n^{-1}\mathbb{E}\left[\mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\}\right] \leqslant C_\alpha$ *for some* $C_\alpha > 0$*; (ii)* $\binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} \mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\} = O_p(1)$*.*

Proof of Lemma A.2. Let $q_{n,ij} := h_n^{-1}\mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\}$. Using Assumption 3(i) $\mathbb{E}\left[q_{n,ij}|\xi_i\right] = h_n^{-1}\int \mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\} f_\xi(\xi_j) d\xi_j \leqslant 2\alpha \overline{f}_\xi = C_\alpha$, which proves the first part. Similarly, $\mathbb{E}\left[q_{n,ij}^2\right] \leqslant \frac{2\alpha \overline{f}_\xi}{h_n} = O(h_n^{-1}) = o(n)$, where the last equality uses Assumption 5(ii). Invoking Lemma A.3 of Ahn and Powell (1993), $\binom{n}{2}^{-1} \sum_{i<j} q_{n,ij} = \mathbb{E}\left[q_{n,ij}\right] + o_p(1) = O_p(1)$, where the last equality also uses $\mathbb{E}\left[q_{n,ij}\right] \leqslant C_\alpha$. Q.E.D.

A.1.2 Proof of Theorem 1

First, notice that $\hat{\beta} - \beta_0 = \hat{A}_n^{-1}(\hat{B}_n + \hat{C}_n)$, where

$$\hat{A}_n := \binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} K\left(\frac{\hat{d}_{ij}^2}{h_n^2}\right) \frac{1}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(W_{ik} - W_{jk})',$$

$$\hat{B}_n := \binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} K\left(\frac{\hat{d}_{ij}^2}{h_n^2}\right) \frac{1}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(g(\xi_i, \xi_k) - g(\xi_j, \xi_k)),$$

$$\hat{C}_n := \binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} K\left(\frac{\hat{d}_{ij}^2}{h_n^2}\right) \frac{1}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(\varepsilon_{ik} - \varepsilon_{jk}).$$

We proof Theorem 1 by inspecting the asymptotic behavior of $\hat{A}_n$, $\hat{B}_n$, and $\hat{C}_n$: below we show (i) $\hat{A}_n = A + o_p(1)$ for some symmetric A satisfying $\lambda_{min}(A) > C > 0$; (ii) $\hat{B}_n = O_p(h_n^2 + \frac{R_n^{-1}}{h_n} + n^{-1})$; (iii) $\hat{C}_n = O_p(\frac{R_n^{-1}}{h_n^2}\left(\frac{\ln n}{n}\right)^{1/2} + n^{-1})$, together delivering the desired result.

For simplicity of exposition and consistency with the assumptions provided in Section 4.1, we prove the result for $d_\xi = 1$. For $d_\xi > 1$ and with the assumptions appropriately modified as discussed in Remark 4.2, the proof of the theorem remains essentially the same with the exceptions that (i) $\hat{A}_n$, $\hat{B}_n$, and $\hat{C}_n$ should be normalized by $h_n^{-d_\xi}$ instead of h_n^{-1}, and (ii) all the quantities in the statements of Lemma A.2 should be normalized by $h_n^{-d_\xi}$ instead of h_n^{-1}.

A.1.3 Demonstrating consistency of $\hat{A}_n$

In this section, we argue that $\hat{A}_n = A + o_p(1)$, where A is a symmetric invertible matrix.

In particular, first, we argue that $\hat{A}_n - A_n = o_p(1)$, where

$$A_n := \binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} K\left(\frac{d_{ij}^2}{h_n^2}\right) \frac{1}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(W_{ik} - W_{jk})'.$$

Then

$$\hat{A}_n - A_n = \binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} \left(K\left(\frac{\hat{d}_{ij}^2}{h_n^2}\right) - K\left(\frac{d_{ij}^2}{h_n^2}\right)\right) \frac{1}{n-2} \sum_{k\neq i,j} (W_{ik} - W_{jk})(W_{ik} - W_{jk})'.$$

First, notice that $\left\| \frac{1}{n-2} \sum_{k\neq i,j} (W_{ik} - W_{jk})(W_{ik} - W_{jk})' \right\|$ is uniformly bounded thanks to Assumption 2(i). Then, using Assumption 5(i) and the result of Lemma A.1(ii), we have.

$$\begin{aligned}\binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} \left| K\left(\frac{\hat{d}_{ij}^2}{h_n^2}\right) - K\left(\frac{d_{ij}^2}{h_n^2}\right)\right| &\leqslant \overline{K}' \frac{\max_{i\neq j} \left|\hat{d}_{ij}^2 - d_{ij}^2\right|}{h_n^2} \binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} \mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\} \\ &= O_p\left(R_n^{-1}/h_n^2\right) = o_p(1),\end{aligned}$$

where the first equality uses (4.1) and Lemma A.1(ii), and the second equality is due to Assumption 5(ii). This completes the proof of $\hat{A}_n - A_n = o_p(1)$.

Next, notice that $A_n = \frac{1}{n(n-1)(n-2)} \sum_{i\neq j\neq k} \zeta_n(Z_i, Z_j, Z_k)$, where

$$\zeta_n(Z_i, Z_j, Z_k) := h_n^{-1} K\left(\frac{d^2(Z_i, Z_j)}{h_n^2}\right) (w(X_j, X_k) - w(X_i, X_k))(w(X_j, X_k) - w(X_i, X_k))'.$$

Note that ζ_n is symmetric in its first two arguments and, consequently, it can be symmetrized with

$$p_n(Z_i, Z_j, Z_k) := \frac{1}{3}\left(\zeta_n(Z_i, Z_j, Z_k) + \zeta_n(Z_k, Z_i, Z_j) + \zeta_n(Z_j, Z_k, Z_i)\right).$$

Then A_n is a third order U-statistic with kernel p_n. First, we want to show that $\mathbb{E}\left[\|p_n(Z_i, Z_j, Z_k)\|^2\right] = o(n)$. Indeed, again since W is bounded, using Assumption 5(i) and the result of Lemma A.1(i),

$$\mathbb{E}\left[\|\zeta_n(Z_i, Z_j, Z_k)\|^2\right] \leqslant C h_n^{-2} \mathbb{E}\left[\mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\}\right] = o(h_n^{-1}) = o(n)$$

where the first equality uses Lemma A.2, and the second uses Assumption 5(ii). Similarly, $\mathbb{E}\left[\|p_n(Z_i, Z_j, Z_k)\|^2\right] = O(h_n^{-1}) = o(n)$, and by Lemma A.3 of Ahn and Powell (1993), we obtain $A_n = \mathbb{E}\left[p_n(Z_i, Z_j, Z_k)\right] + o_p(1) = \mathbb{E}\left[\zeta_n(Z_i, Z_j, Z_k)\right] + o_p(1)$. Since we have previously established that $\hat{A}_n = A_n + o_p(1)$, we also get $\hat{A}_n = \mathbb{E}\left[\zeta_n(Z_i, Z_j, Z_k)\right] + o_p(1)$.

The rest of the proof deals with computing $\mathbb{E}\left[\zeta_n(Z_i, Z_j, Z_k)\right]$. First, note

$$\mathbb{E}\left[\zeta_n(Z_i, Z_j, Z_k)|X_i, X_j\right] = \mathbb{E}\left[h_n^{-1} K\left(\frac{d^2(X_i, X_j, \xi_i, \xi_j)}{h_n^2}\right)|X_i, X_j\right] \mathcal{C}(X_i, X_j),$$

where $\mathcal{C}(X_i, X_j)$ is as defined in (4.3). Using the result of Lemma A.1(iii) and Assumption 3(i), we obtain

$$I(X_i, X_j, \xi_i; h_n) \coloneqq \mathbb{E}\left[h_n^{-1} K\left(\frac{d^2(X_i, X_j, \xi_i, \xi_j)}{h_n^2}\right) |X_i, X_j, \xi_i\right] = I_1(X_i, X_j, \xi_i; h_n) + O(h_n),$$
$$I_1(X_i, X_j, \xi_i; h_n) \coloneqq \int h_n^{-1} K\left(\frac{c_{ij}(\xi_j - \xi_i)^2}{h_n^2}\right) f_{\xi|X}(\xi_j; X_j) d\xi_j = \frac{1}{\sqrt{c_{ij}}} \int K(u^2) f_{\xi|X}\left(\xi_i + \frac{h_n u}{\sqrt{c_{ij}}}; X_j\right) du,$$

where the last follows from the change of the variable $\xi_j = \xi_i + h_n u/\sqrt{c_{ij}}$. Next, note that for all values of ξ_i such that $f_{\xi|X}(\xi_i|X_j)$ is continuous at ξ_i, by the dominated convergence theorem (DCT), we have $I_1(X_i, X_j, \xi_i; h_n) \to \frac{\mu_K f_{\xi|X}(\xi_i|X_j)}{\sqrt{c(X_i, X_j, \xi_i)}}$ and, consequently, also $I(X_i, X_j, \xi_i; h_n) \to \frac{\mu_K f_{\xi|X}(\xi_i|X_j)}{\sqrt{c(X_i, X_j, \xi_i)}}$, where $c_{ij} = c(X_i, X_j, \xi_i)$, and μ_K is as defined in Assumption 5(i). By Assumption 3(ii), this applies for almost all ξ_i. Moreover, by Lemmas A.1(i) and A.2(i) and Assumption 5(i), $I(X_i, X_j, \xi_i; h_n)$ is uniformly bounded since we have

$$I(X_i, X_j, \xi_i; h_n) \leqslant \overline{K} h_n^{-1} \mathbb{E}\left[\mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha\} h_n |\xi_i\right] \leqslant C.$$

Hence, by the DCT, for all X_i and X_j, we have,

$$I(X_i, X_j; h_n) \coloneqq \mathbb{E}\left[I(X_i, X_j, \xi_i; h_n)|X_i, X_j\right] \to \underbrace{\int \frac{\mu_K}{\sqrt{c(X_i, X_j, \xi_i)}} f_{\xi|X}(\xi|X_i) f_{\xi|X}(\xi|X_j) d\xi}_{\coloneqq \lambda(X_i, X_j)},$$

Finally, since $\mathbb{E}\left[\zeta_n(Z_i, Z_j, Z_k)|X_i, X_j\right] = I(X_i, X_j; h_n)\mathcal{C}(X_i, X_j)$ is uniformly bounded, the DCT guarantees $\mathbb{E}\left[\zeta_n(Z_i, Z_j, Z_k)\right] \to A \coloneqq \mathbb{E}\left[\lambda(X_i, X_j)\mathcal{C}(X_i, X_j)\right]$. Hence, we conclude $\hat{A}_n = \mathbb{E}\left[\zeta_n(Z_i, Z_j, Z_k)\right] + o_p(1) = A + o_p(1)$, where $\lambda_{min}(A) > C$ for some $C > 0$ by Assumption 4(i), which completes the proof.

A.1.4 Bounding $\hat{B}_n$

First, we introduce the following notations and prove an auxiliary lemma stated below.

$$q_n(Z_i, Z_j, Z_k) \coloneqq h_n^{-1} K\left(\frac{d^2(Z_i, Z_j)}{h_n^2}\right)(W(X_i, X_k) - W(X_j, X_k))(g(\xi_i, \xi_k) - g(\xi_j, \xi_k)),$$
$$p_n(Z_i, Z_j, Z_k) \coloneqq \frac{1}{3}(q_n(Z_i, Z_j, Z_k) + q_n(Z_k, Z_i, Z_j) + q_n(Z_j, Z_k, Z_i)), \quad \text{(A.1)}$$

Lemma A.3. *Suppose that the hypotheses of Theorem 1 are satisfied. Then, (i)* $b_n \coloneqq \mathbb{E}\left[q_n(Z_i, Z_j, Z_k)\right] = O(h_n^2)$*, and (ii)* $\zeta_{1,n} \coloneqq \mathbb{E}\left[\|\mathbb{E}\left[p_n(Z_i, Z_j, Z_k)|Z_i\right] - b_n\|^2\right] = O(h_n^3)$*.*

Proof of Lemma A.3. Before starting the proof, we introduce the following notations

$s_n^{(1)}(Z_i) \coloneqq \mathbb{E}\left[q_n(Z_i, Z_j, Z_k)|Z_i\right] - b_n$, $s_n^{(2)}(Z_i) \coloneqq \mathbb{E}\left[q_n(Z_k, Z_i, Z_j)|Z_i\right] - b_n$, and $s_n^{(3)}(Z_i) \coloneqq \mathbb{E}\left[q_n(Z_j, Z_k, Z_i)|Z_i\right] - b_n$, so

$$\begin{aligned}\mathbb{E}\left[p_n(Z_i, Z_j, Z_k)|Z_i\right] - b_n &= \frac{1}{3}\left(s_n^{(1)}(Z_i) + s_n^{(2)}(Z_i) + s_n^{(3)}(Z_i)\right), \\ \zeta_{1,n}^2 &= \frac{1}{9}\mathbb{E}\left[\left\|s_n^{(1)}(Z_i) + s_n^{(2)}(Z_i) + s_n^{(3)}(Z_i)\right\|^2\right]. \end{aligned} \tag{A.2}$$

First, let us compute $E_q(X_i, \xi_i, X_j, Z_k; h_n) \coloneqq \mathbb{E}\left[q_n(Z_i, Z_j, Z_k)|X_i, \xi_i, X_j, Z_k\right]$. Note that

$$\begin{aligned} E_q(X_i, \xi_i, X_j, Z_k; h_n) =& \underbrace{\mathbb{E}\left[h_n^{-1}K\left(\frac{c_{ij}(\xi_j - \xi_i)^2}{h_n^2}\right)(W_{ik} - W_{jk})(g(\xi_i, \xi_k) - g(\xi_j, \xi_k))|X_i, \xi_i, X_j, Z_k\right]}_{\coloneqq E_{q,2}(X_i, \xi_i, X_j, Z_k; h_n)} \\ &+ \underbrace{\mathbb{E}\left[h_n^{-1}r_K(Z_i, Z_j; h_n)(W_{ik} - W_{jk})(g(\xi_i, \xi_k) - g(\xi_j, \xi_k))|X_i, X_j, \xi_i\right]}_{\coloneqq r_{E_q}(X_i, \xi_i, X_j, Z_k; h_n)}, \end{aligned}$$

where r_K is defined in Lemma A.1. Combining Lemma A.1(iii), boundedness of W, Assumption 2(iii), and Lemma A.2(i), we conclude that there exists $C > 0$ such that $\left|r_{E_q}(X_i, \xi_i, X_j, Z_k; h_n)\right| \leqslant Ch_n^2$ a.s.

Next, we compute $E_{q,2}(X_i, \xi_i, X_j, Z_k; h_n)$. We start with considering ξ_i and ξ_k such that $|\xi_i - \xi_k| > \alpha h_n$. Using Lemma A.1(i), Assumption 6, and boundedness of W, we obtain

$$\left\|K\left(\frac{c_{ij}(\xi_j - \xi_i)^2}{h_n^2}\right)(W_{ik} - W_{jk})(\underbrace{g(\xi_i, \xi_k) - g(\xi_j, \xi_k) - G(\xi_i, \xi_k)(\xi_i - \xi_j)}_{r_g(\xi_i, \xi_j, \xi_k)})\right\| \leq Ch_n^2 \mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\}$$

for some $C > 0$ for all (Z_i, Z_j, Z_k) satisfying $|\xi_i - \xi_k| > \alpha h_n$. Then, using Lemma A.2(i),

$$\begin{aligned} E_{q,2}(X_i, \xi_i, X_j, Z_k; h_n) =& \underbrace{\mathbb{E}\left[h_n^{-1}K\left(\frac{c_{ij}(\xi_j - \xi_i)^2}{h_n^2}\right)(W_{jk} - W_{ik})G(\xi_i, \xi_k)(\xi_j - \xi_i)|X_i, \xi_i, X_j, Z_k\right]}_{\coloneqq E_{q,l}(X_i, \xi_i, X_j, Z_k; h_n)} \\ &+ r_{E_{q,2}}(X_i, \xi_i, X_j, Z_k; h_n), \end{aligned}$$

where $\left|r_{E_{q,2}}(X_i, \xi_i, X_j, Z_k; h_n)\right| \leqslant Ch_n^2$ a.s. for some $C > 0$ for $|\xi_i - \xi_k| > \alpha h_n$. Next, consider

$$\begin{aligned} E_{q,l}(X_i, \xi_i, X_j, Z_k; h_n) &= (W_{jk} - W_{ik})G(\xi_i, \xi_k)\int h_n^{-1}K\left(\frac{c_{ij}(\xi_j - \xi_i)^2}{h_n^2}\right)(\xi_j - \xi_i)f_{\xi|X}(\xi_j; X_j)d\xi_j \\ &= \frac{h_n(W_{jk} - W_{ik})G(\xi_i, \xi_k)}{c_{ij}}\int K(u^2)uf_{\xi|X}(\xi_i + h_nu/\sqrt{c_{ij}}; X_j)du, \end{aligned}$$

where the last equality follows from the change of the variable $\xi_j = \xi_i + h_n u/\sqrt{c_{ij}}$. Note

$$\left|\int K(u^2)uf_{\xi|X}(\xi_i + h_n u/\sqrt{c_{ij}}; X_j)du\right|$$
$$\leqslant \int \left|K(u^2)u\left(f_{\xi|X}(\xi_i + h_n u/\sqrt{c_{ij}}; X_j) - f_{\xi|X}(\xi_i; X_j)\right)\right| du \leqslant \frac{\overline{K}C_\xi h_n}{\sqrt{c_{ij}}} \tag{A.3}$$

where the first inequality used $\int uK(u^2)du = 0$ and the triangle inequality, and the second inequality holds for all ξ_i such that $f_{\xi|X}(\cdot; X_j)$ is continuous on $B_{h_n u/\sqrt{c_{ij}}}(\xi_i)$ thanks to Assumptions 3(iii) and 5(i). (A.3) together with boundedness of W, G and c_{ij}^{-1} (Assumptions 2(i), 6, and 4(iii), respectively) implies that there exists a uniform constant C such that $|E_{q,l}(X_i, \xi_i, X_j, Z_k; h_n)| \leqslant Ch_n^2$ almost surely for all Z_i, Z_j, and Z_k such that $|\xi_i - \xi_k| > \alpha h_n$ and $f_{\xi|X}(\cdot; X_j)$ is continuous on $B_{h_n u/\sqrt{c_{ij}}}(\xi_i)$. Combining this with the previously obtained bounds on r_{E_q} and $r_{E_{q,2}}$, we conclude that there exists a uniform constant $C > 0$ such that (for sufficiently large n)

$$|E_q(X_i, \xi_i, X_j, Z_k; h_n)| \leqslant Ch_n^2 \tag{A.4}$$

almost surely for all Z_i, Z_j, and Z_k such that $|\xi_i - \xi_k| > \alpha h_n$ and $f_{\xi|X}(\cdot; X_j)$ is continuous on $B_{h_n u/\sqrt{c_{ij}}}(\xi_i)$. Moreover, for all (Z_i, Z_j, Z_k) without additional qualifiers, we also have

$$|E_q(X_i, \xi_i, X_j, Z_k; h_n)| \leqslant Ch_n, \tag{A.5}$$

for some $C > 0$. To inspect (A.5), note $\|q_n(Z_i, Z_j, Z_k)\| \leqslant C\mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\}$ a.s. for some $C > 0$ thanks to Lemma A.1(i) and Assumptions 2(i) and (iii), and apply Lemma A.2(i).

Equipped with the bounds (A.4) and (A.5), now we want to bound $b_n = \mathbb{E}[q_n(Z_i, Z_j, Z_k)]$. We start with considering $\mathbb{E}[q_n(Z_i, Z_j, Z_k)|Z_i]$. Since $c_{ij} > \underline{c} > 0$ (Assumption 4(iii)), Assumption 3(ii) guarantees that there exists $\gamma_1 > 0$ such that the probability mass of ξ_i such that $f_{\xi|X}(\cdot; X_j)$ is continuous on $B_{h_n u/\sqrt{c_{ij}}}(\xi_i)$ is at least $1 - \gamma_1 h_n$ irrespectively of the values of X_i and X_j. Also, by Assumption 3(i), the probability mass of ξ_k such that $|\xi_k - \xi_i| > \alpha h_n$ is at least $1 - \gamma_2 h_n$ irrespectively of the value of ξ_i for some $\gamma_2 > 0$. For such values of ξ_i and ξ_k, the bound (A.4) applies. Moreover, the bound (A.5) applies with probability one. Hence, integrating $E_q(X_i, \xi_i, X_j, Z_k; h_n)$ over (X_j, Z_k) ensures that there exists $C > 0$ such that $\|\mathbb{E}[q_n(Z_i, Z_j, Z_k)|Z_i]\| \leqslant Ch_n^2$ for all ξ_i such that $f_{\xi|X}(\cdot; X_j)$ is continuous on $B_{h_n u/\sqrt{c_{ij}}}(\xi_i)$, which happens with probability $1 - \gamma_1 h_n$ at least. Moreover, (A.5) immediately implies that $\|\mathbb{E}[q_n(Z_i, Z_j, Z_k)|Z_i]\| \leqslant Ch_n$ with probability one. Combining these bounds and integrating over Z_i gives

$$\|\mathbb{E}\left[q_n(Z_i, Z_j, Z_j)\right]\| \leqslant \mathbb{E}\left[\|\mathbb{E}\left[q_n(Z_i, Z_j, Z_j)|Z_i\right]\|\right] \leqslant (1-\gamma_1 h_n) \times Ch_n^2 + \gamma_1 h_n \times Ch_n = O(h_n^2),$$

which completes the proof of the first statement of the lemma.

To prove the second statement of the lemma, thanks to (A.2), it is sufficient to verify that $\mathbb{E}[\|s_n^{(\ell)}(Z_i)\|^2] = O(h_n^3)$ for $\ell \in \{1, 2, 3\}$. We start with $\ell \in \{1, 2\}$. By the same bounds on $\|\mathbb{E}\left[q_n(Z_i, Z_j, Z_j)|Z_i\right]\|$ as established above and $b_n = O(h_n^2)$, the following holds: (i) for some $C > 0$ and $\gamma_1 > 0$, we have $\|s_n^{(1)}(Z_i)\|^2 \leqslant Ch_n^4$ with probability at least $1 - \gamma_1 h_n$ and $\|s_n^{(1)}(Z_i)\|^2 \leqslant Ch_n^2$ with probability one. Hence, $\mathbb{E}[\|s_n^{(1)}(Z_i)\|^2] = \mathbb{E}[\|s_n^{(2)}(Z_i)\|^2] \leqslant Ch_n^3$ for some $C > 0$. Finally, notice we can write $s_n^{(3)}(Z_k) = \mathbb{E}\left[q_n(Z_i, Z_j, Z_k)|Z_k\right] - b_n$. Again, for a fixed ξ_k, the bound (A.4) applies for all ξ_i satisfying (i) $|\xi_k - \xi_i| > \alpha h_n$ and (ii) $f_{\xi|X}(\cdot; X_j)$ is continuous on $B_{h_n u/\sqrt{c_{ij}}}(\xi_i)$. By the same reasoning as above, the probability mass of such ξ_i is at least $1-\gamma_3 h_n$ for some $\gamma_3 > 0$ irrespectively of (X_i, X_j, Z_k). At the same time, the bound (A.5) applies with probability one. Hence, integrating $E_q(X_i, \xi_i, X_j, Z_k; h_n)$ over (X_i, ξ_i, X_j) gives $\|\mathbb{E}\left[q_n(Z_i, Z_j, Z_k)|Z_k\right]\| \leqslant Ch_n^2$ a.s. for some $C > 0$. This, paired with $b_n = O(h_n^2)$, implies that $\mathbb{E}[\|s_n^{(3)}(Z_k)\|^2] = \mathbb{E}[\|s_n^{(3)}(Z_i)\|^2] \leqslant Ch_n^4$, which completes the proof. Q.E.D.

Bounding $\hat{B}_n$ Equipped with Lemma A.3, we are now ready to bound $\hat{B}_n$. The first step of the proof is to bound B_n defined as

$$\begin{aligned} B_n &\coloneqq \binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} K\left(\frac{d_{ij}^2}{h_n^2}\right) \frac{1}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(g(\xi_i, \xi_k) - g(\xi_j, \xi_k)) \\ &= \binom{n}{3}^{-1} \sum_{i<j<k} p_n(Z_i, Z_j, Z_k) = b_n + \underbrace{\binom{n}{3}^{-1} \sum_{i<j<k} (p_n(Z_i, Z_j, Z_k) - b_n)}_{\coloneqq U_n}, \end{aligned}$$

so B_n is a third order U-statistic with the symmetrized kernel $p_n(Z_i, Z_j, Z_k)$ given by (A.1), and $b_n \coloneqq \mathbb{E}\left[q_n(Z_i, Z_j, Z_k)\right] = \mathbb{E}\left[p_n(Z_i, Z_j, Z_k)\right]$ is defined and bounded in Lemma A.3. We proceed with bounding using a Bernstein type inequality for U-statistic developed in Arcones (1995). Specifically, Theorem 2 in Arcones (1995) guarantees that $U_n = O_p\left(\max\left\{\frac{\zeta_{1,n}}{n^{1/2}}, \frac{1}{n}\right\}\right)$, where $\zeta_{1,n}$ is defined in Lemma A.3, which demonstrates $\zeta_{1,n}^2 = O(h_n^3)$. Using Assumption 5(ii), we also obtain $\frac{\zeta_{1,n}}{n^{1/2}} = \frac{O(h_n^{3/2})}{n^{1/2}} = o(h_n^2)$. Finally, since Lemma A.3(i) guarantees $b_n = O(h_n^2)$, we conclude $B_n = b_n + U_n = O_p(h_n^2 + n^{-1})$.

The second step is to bound $\hat{B}_n - B_n$. Combining the result of the result of Lemma

A.1(ii) with Assumption 5(i), we have, with probability approaching one,

$$\left\|\hat{B}_n - B_n\right\| \leqslant \overline{K}' \frac{\max_{i\neq j}\left|\hat{d}_{ij}^2 - d_{ij}^2\right|}{h_n^2} \times \binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} \underbrace{\mathbb{1}\{|\xi_j - \xi_i| \leqslant \alpha h_n\} \frac{\sum_{k\neq i,j}\|(W_{ik} - W_{jk})(g(\xi_i,\xi_k) - g(\xi_j,\xi_k))\|}{n-2}}_{\leqslant C h_n \mathbb{1}\{|\xi_j - \xi_i|\leqslant \alpha h_n\}}.$$

Notice that using Assumption 2(iii), boundedness of W, and Lemma A.2(ii), we can bound the second line of the display equation above by $O_p(h_n)$. Together with (4.1), this allows us to obtain $\hat{B}_n - B_n = O_p\left(\frac{R_n^{-1}}{h_n}\right)$. Together with the previously obtained bound on B_n, this delivers the required result.

A.1.5 Bounding $\hat{C}_n$

$$\hat{C}_n = \underbrace{\binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} K\left(\frac{d_{ij}^2}{h_n^2}\right) \frac{1}{n-2} \sum_{k\neq i,j} (W_{ik} - W_{jk})(\varepsilon_{ik} - \varepsilon_{jk})}_{:=C_n} + \underbrace{\binom{n}{2}^{-1} h_n^{-1} \sum_{i<j} \left(K\left(\frac{\hat{d}_{ij}^2}{h_n^2}\right) - K\left(\frac{d_{ij}^2}{h_n^2}\right)\right) \frac{1}{n-2} \sum_{k\neq i,j}(W_{ik} - W_{jk})(\varepsilon_{ik} - \varepsilon_{jk})}_{:=\Delta\hat{C}_n}. \quad \text{(A.6)}$$

The first step is to argue that $C_n = O_p\left(n^{-1}\right)$. Let $K_{n,ij} \coloneqq h_n^{-1} K\left(\frac{d_{ij}^2}{h_n^2}\right)$. Note

$$\sum_{i<j} K_{n,ij} \sum_{k\neq i,j}(W_{ik} - W_{jk})(\varepsilon_{ik} - \varepsilon_{jk}) = \sum_{i\neq j} K_{n,ij} \sum_{k\neq i,j}(W_{ik} - W_{jk})\varepsilon_{ik}$$
$$= \sum_{i<k}\left(\sum_{j\neq i,k}\left(K_{n,ij}(W_{ik} - W_{jk}) + K_{n,kj}(W_{ik} - W_{ji})\right)\right)\varepsilon_{ik}.$$

Hence, we can represent C_n as

$$C_n = \binom{n}{2}^{-1} \sum_{i<k} \hat{\omega}_{n,ik}\varepsilon_{ik}, \quad \hat{\omega}_{n,ik} \coloneqq \underbrace{\frac{\sum_{j\neq i,k} K_{n,ij}(W_{ik} - W_{jk})}{n-2}}_{:=\hat{\kappa}_{n,ik}} + \underbrace{\frac{\sum_{j\neq i,k} K_{n,kj}(W_{ik} - W_{ji})}{n-2}}_{:=\hat{\eta}_{n,ik}}.$$

Next, we want to argue that $\max_{i\neq k}\|\hat{\omega}_{n,ik}\| < C_\omega$ with probability approaching one for some $C_\omega > 0$. To this end, we first show that $\max_{i\neq k}\|\hat{\omega}_{n,ik} - \omega_{n,ik}\| = o_p(1)$ and then we verify that $\max_{i\neq k}\|\omega_{n,ik}\| < C$ for some $C>0$, where

$$\omega_{n,ik} \coloneqq \underbrace{\mathbb{E}\left[K_{n,ij}(W_{ik}-W_{jk})|Z_i,Z_k\right]}_{:=\kappa_{n,ik}} + \underbrace{\mathbb{E}\left[K_{n,kj}(W_{ik}-W_{ji})|Z_i,Z_k\right]}_{:=\eta_{n,ik}}.$$

Uniform consistency of $\hat{\omega}_{n,ik}$ would follow from uniform consistency of $\hat{\kappa}_{n,ik}$ and $\eta_{n,ik}$ to $\kappa_{n,ik}$ and $\eta_{n,ik}$, respectively. Next, we will establish this result for $\hat{\kappa}_{n,ik}$ noting that the analogous result for $\hat{\eta}_{n,ik}$ follows by the same argument. Note that, conditional on Z_i and Z_k, $\{K_{n,ij}(W_{ik}-W_{jk})\}_{j\neq i,k}$ is a collection of bounded (given the sample size) independent variables with $\|K_{n,ij}(W_{ik}-W_{jn})\| \leqslant Ch_n^{-1}$ and $\mathbb{E}\left[\|K_{n,ij}(W_{ik}-W_{jn})\|^2|Z_i,Z_k\right] \leqslant Ch_n^{-1}$ for all Z_i and Z_k for some $C>0$, where we used boundedness of W and K combined with Lemmas A.1(i) and A.2(i). Hence, applying Bernstein inequality A.1, we conclude that there exist positive constants a, b, and C such that for all Z_i, Z_k and $\epsilon>0$ we have

$$\mathbb{P}\left(\|\hat{\kappa}_{n,ik}-\kappa_{n,ik}\| \geqslant \epsilon|Z_i,Z_k\right) \leqslant C\exp\left(-\frac{(n-2)h_n\epsilon^2}{a+b\epsilon}\right),$$
$$\mathbb{P}\left(\max_{i\neq k}\|\hat{\kappa}_{n,ik}-\kappa_{n,ik}\| \geqslant \epsilon\right) \leqslant \binom{n}{2}C\exp\left(-\frac{(n-2)h_n\epsilon^2}{a+b\epsilon}\right) \to 0,$$

where the second inequality follows from the union bound, and the convergence follows from Assumption 5(ii). Hence, $\hat{\kappa}_{n,ik}$, and, consequently, $\hat{\omega}_{n,ik}$ are uniformly consistent for $\kappa_{n,ik}$, and $\omega_{n,ik}$, respectively. Finally, using the boundedness of W and K and Lemmas A.1(i) and A.2(i) again, we conclude that $\max_{i\neq k}\|\omega_{n,ik}\| \leq C$ for some $C>0$, which delivers the desired result for $\hat{\omega}_{n,ik}$.

We are now ready to bound C_n using that $\max_{i\neq k}\|\hat{\omega}_{n,ik}\| < C_\omega$ with probability approaching one. Notice that, conditional on $\{Z_i\}_{i=1}^n$, $\{\hat{\omega}_{n,ik}\varepsilon_{ik}\}_{i<k}$ is a collection of independent vectors with zero mean, which satisfy the requirements of Theorem A.2. Therefore, Theorem A.2 guarantees that there exist some positive constants C, a, b such that for all $\{Z_i\}_{i=1}^n$ satisfying $\max_{i\neq k}\|\hat{\omega}_{n,ik}\| < C_\omega$, for all $\epsilon>0$,

$$\mathbb{P}\left(\|C_n\| > \epsilon|\{X_i,\xi_i\}_{i=1}^n\right) \leqslant C\exp\left(-\frac{\binom{n}{2}\epsilon^2}{a+b\epsilon}\right).$$

Since the requirement $\max_{i\neq k}\|\hat{\omega}_{n,ik}\| < C_\omega$ is satisfied with probability approaching one, we conclude that $C_n = O_p(n^{-1})$.

The final part of the proof is to bound $\Delta\hat{C}_n$ defined in (A.6). Applying Corollary A.1 conditional on $\{Z_i\}_{i=1}^n$ and using boundedness of W, we conclude there exist positive constants C_2, a_2 and b_2 such that for all $\{Z_i\}_{i=1}^n$ and for all $\epsilon > 0$

$$\mathbb{P}\left(\left\|\frac{1}{n-2}\sum_{k\neq i,j}(W_{ik}-W_{jk})(\varepsilon_{ik}-\varepsilon_{jk})\right\| > \epsilon|\{Z_i\}_{i=1}^n\right) \leqslant C_2\exp\left(-\frac{(n-2)\epsilon^2}{a_2+b_2\epsilon}\right),$$
$$\mathbb{P}\left(\max_{i\neq j}\left\|\frac{1}{n-2}\sum_{k\neq i,j}(W_{ik}-W_{jk})(\varepsilon_{ik}-\varepsilon_{jk})\right\| > \epsilon\right) \leqslant \binom{n}{2}C_2\exp\left(-\frac{(n-2)\epsilon^2}{a_2+b_2\epsilon}\right),$$

where the second inequality uses the union bound and the uniformity of those constants over $\{Z_i\}_{i=1}^n$. Hence, we have $\max_{i\neq j}\|\frac{1}{n-2}\sum_{k\neq i,j}(W_{ik}-W_{jk})(\varepsilon_{ik}-\varepsilon_{jk})\| = O_p\left(\left(\frac{\ln n}{n}\right)^{1/2}\right)$ implying

$$\left\|\Delta\hat{C}_n\right\| \leqslant \binom{n}{2}^{-1} h_n^{-1}\sum_{i<j}\left|K\left(\frac{\hat{d}_{ij}^2}{h_n^2}\right)-K\left(\frac{d_{ij}^2}{h_n^2}\right)\right| \times O_p\left(\left(\frac{\ln n}{n}\right)^{1/2}\right).$$

Combining this with (4.1), we obtain $\|\Delta\hat{C}_n\| = O_p\left(\frac{R_n^{-1}}{h_n^2}\left(\frac{\ln n}{n}\right)^{1/2}\right)$, which, together with $C_n = O_p(n^{-1})$, delivers the desired result for $\hat{C}_n = C_n + \Delta\hat{C}_n$.

A.2 Proof of the results of Section 4.2.1

First, we introduce the following notations and prove an auxiliary lemma stated below.

$$\begin{aligned}
\hat{q}_{ij}^2(\beta) &:= \frac{1}{n-2}\sum_{k\neq i,j}\left(Y_{ik}-Y_{jk}-(W_{ik}-W_{jk})'\beta\right)^2,\\
d_{ij,n-2}^2(\beta) &:= \frac{1}{n-2}\sum_{k\neq i,j}(Y_{ik}^*-Y_{jk}^*-(W_{ik}-W_{jk})'\beta)^2,\\
d_{ij,n}^2(\beta) &:= \frac{1}{n}\sum_{k}(Y_{ik}^*-Y_{jk}^*-(W_{ik}-W_{jk})'\beta)^2, \qquad \text{(A.7)}\\
q_{ij}^2(\beta) &:= \mathbb{E}\left[(Y_{ik}-Y_{jk}-(W_{ik}-W_{jk})'\beta)^2\right], \quad d_{ij}^2(\beta) = \mathbb{E}\left[(Y_{ik}^*-Y_{jk}^*-(W_{ik}-W_{jk})'\beta)^2\right].
\end{aligned}$$

For brevity, we will also use $g_{ij} := g(\xi_i, \xi_i)$ for all i and j.

Lemma A.4. *Suppose that $\mathcal{B}$ is compact. Then, under Assumptions 1 and 2,*

(i) $\max_{i\neq j}\max_{\beta\in\mathcal{B}}\left|d_{ij,n-2}^2(\beta)-d_{ij}^2(\beta)\right| = O_p\left((\ln n/n)^{1/2}\right)$,

(ii) $\max_{i\neq j}\max_{\beta\in\mathcal{B}}\left|d_{ij,n}^2(\beta)-d_{ij}^2(\beta)\right| = O_p\left((\ln n/n)^{1/2}\right)$.

Proof of Lemma A.4. Using $Y^*_{ik} - Y^*_{jk} = (W_{ik} - W_{jk})'\beta_0 + g_{ik} - g_{jk}$, we obtain

$$\begin{aligned} d^2_{ij,n-2}(\beta) =& (\beta_0 - \beta)' \frac{1}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(W_{ik} - W_{jk})'(\beta_0 - \beta) \\ &+ \frac{2}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(g_{ik} - g_{jk})(\beta_0 - \beta) + \frac{1}{n-2} \sum_{k \neq i,j} (g_{ik} - g_{jk})^2, \\ d^2_{ij}(\beta) =& (\beta_0 - \beta)' \mathbb{E}\left[(W_{ik} - W_{jk})(W_{ik} - W_{jk})' | Z_i, Z_j\right](\beta_0 - \beta) \\ &+ 2\mathbb{E}\left[(W_{ik} - W_{jk})(g_{ik} - g_{jk}) | Z_i, Z_j\right](\beta_0 - \beta) + \mathbb{E}\left[(g_{ik} - g_{jk})^2 | Z_i, Z_j\right]. \end{aligned}$$

Since $\mathcal{B}$ is bounded, for some positive constants c_1 and c_2, we have

$$\max_{i \neq j} \max_{\beta \in \mathcal{B}} \left| d^2_{ij,n-2}(\beta) - d^2_{ij}(\beta) \right| \leqslant c_1 S_1 + c_2 S_2 + S_3, \tag{A.8}$$

$$S_1 \coloneqq \max_{i \neq j} \left\| \frac{1}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(W_{ik} - W_{jk})' - \mathcal{C}(X_i, X_j) \right\|,$$

$$S_2 \coloneqq \max_{i \neq j} \left\| \frac{1}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(g_{ik} - g_{jk})' - \mathbb{E}\left[(W_{ik} - W_{jk})(g_{ik} - g_{jk}) | Z_i, Z_j\right] \right\|,$$

$$S_3 \coloneqq \max_{i \neq j} \left| \frac{1}{n-2} \sum_{k \neq i,j} (g_{ik} - g_{jk})^2 - \mathbb{E}\left[(g_{ik} - g_{jk})^2 | Z_i, Z_j\right] \right|.$$

where we used $\mathcal{C}(X_i, X_j) = \mathbb{E}\left[(W_{ik} - W_{jk})(W_{ik} - W_{jk})' | Z_i, Z_j\right]$. Hence, to prove the first statement of the lemma, it is sufficient to show that $S_\ell = O_p\left((\ln n/n)^{1/2}\right)$ for $\ell \in \{1, 2, 3\}$.

We start with $\ell = 1$. Conditional on Z_i and Z_j, $\{(W_{ik} - W_{jk})(W_{ik} - W_{jk})'\}_{k \neq i,j}$ is a collection of iid matrices, which are uniformly bounded over Z_i and Z_j. Hence, by Bernstein inequality A.1, there exist positive constants a, b, and C such that for all $\epsilon > 0$ and (almost) all Z_i and Z_j, we have

$$\mathbb{P}\left(\left\| \frac{1}{n-2} \sum_{k \neq i,j} (W_{ik} - W_{jk})(W_{ik} - W_{jk})' - \mathcal{C}(X_i, X_j) \right\| > \epsilon | Z_i, Z_j \right) \leqslant C \exp\left(-\frac{(n-2)\epsilon^2}{a + b\epsilon} \right).$$

Since a, b, and C are uniform over Z_i and Z_j, applying the union bound, we have

$$\mathbb{P}(S_1 > \epsilon) \leqslant \binom{n}{2} C \exp\left(-\frac{(n-2)\epsilon^2}{a + b\epsilon} \right),$$

which implies $S_1 = O_p\left((\ln n/n)^{1/2}\right)$. The same argument can be applied to inspect that $S_\ell = O_p\left((\ln n/n)^{1/2}\right)$ for $\ell \in \{2, 3\}$, which together with (A.8) delivers the desired result for $d^2_{ij,n-2}$. Essentially the same argument applies to demonstrate the result for $d^2_{ij,n}$. Q.E.D.

Proof of Lemma 1. Recall that under (2.5), $q_{ij}^2(\beta) = d_{ij}^2(\beta) + 2\sigma^2$. First, we show that

$$\max_{i\neq j}\max_{\beta\in\mathcal{B}}\left|\hat{q}_{ij}^2(\beta) - q_{ij}^2(\beta)\right| = \max_{i\neq j}\max_{\beta\in\mathcal{B}}\left|\hat{q}_{ij}^2(\beta) - d_{ij}^2(\beta) - 2\sigma^2\right| = O_p\left((\ln n/n)^{1/2}\right). \quad \text{(A.9)}$$

Next, $\hat{q}_{ij}^2(\beta)$ can be decomposed as

$$\begin{aligned}\hat{q}_{ij}^2(\beta) = & \underbrace{\frac{1}{n-2}\sum_{k\neq i,j}\left(Y_{ik}^* - Y_{jk}^* - (W_{ik} - W_{jk})'\beta\right)^2}_{d_{ij,n-2}^2(\beta)} + \underbrace{\frac{1}{n-2}\sum_{k\neq i,j}(\varepsilon_{ik} - \varepsilon_{jk})^2}_{:=2\hat{\sigma}_{ij}^2} \\ & + \underbrace{\frac{2}{n-2}\sum_{k\neq i,j}\left((W_{ik} - W_{jk})'(\beta_0 - \beta) + g_{ik} - g_{jk}\right)(\epsilon_{ik} - \epsilon_{jk})}_{:=\hat{\Delta}_{ij}(\beta)}.\end{aligned}$$

Thanks to the result of Lemma A.4(i), to verify (A.9), it is sufficient to show that (i) $\max_{i\neq j}\max_{\beta}|\hat{\Delta}_{ij}(\beta)| = O_p\left((\ln n/n)^{1/2}\right)$, and (ii) $\max_{i\neq j}|2\hat{\sigma}_{ij}^2 - 2\sigma^2| = O_p\left((\ln n/n)^{1/2}\right)$. To inspect (i), recall that $\max_{i\neq j}\|\frac{1}{n-2}\sum_{k\neq i,j}(W_{ik} - W_{jk})(\varepsilon_{ik} - \varepsilon_{jk})\| = O_p\left((\ln n/n)^{1/2}\right)$ has already been demonstrated in Section A.1.5. By a nearly identical argument, we also have $\max_{i\neq j}\left|\frac{1}{n-2}\sum_{k\neq i,j}(g_{ik} - g_{jk})(\epsilon_{ik} - \epsilon_{jk})\right| = O_p\left((\ln n/n)^{1/2}\right)$. Together with boundedness of $\mathcal{B}$, these to convergence results imply that (i) holds. Next, to inspect (ii), note that, conditional on $\{Z_i, Z_j\}$, $\{\varepsilon_{ik} - \varepsilon_{jk}\}_{k\neq i,j}$ is a collection of iid sub-Gaussian random variables. Then Corollary A.1 guarantees that there exist some positive constants a, b, and C such that for all $\epsilon > 0$ and (almost) all Z_i and Z_j

$$\mathbb{P}\left(\left|\frac{1}{n-2}\sum_{k\neq i,j}\left((\varepsilon_{ik} - \varepsilon_{jk})^2 - 2\sigma^2\right)|Z_i, Z_j\right| > \epsilon\right) \leqslant C\exp\left(-\frac{(n-2)\epsilon^2}{a + b\epsilon}\right). \quad \text{(A.10)}$$

Importantly, Assumption 2(iv) guarantees that a, b, and C are uniform over Z_i and Z_j. Hence, (ii) following from turning (A.10) into an analogous unconditional statement and combining it with the union bound. Inspecting (i) and (ii), completes the proof that (A.9) holds. Finally, (A.9), in turn, guarantees that $\max_{i\neq j}\left|\hat{q}_{ij}^2 - q_{ij}^2\right| = O_p\left((\ln n/n)^{1/2}\right)$, which completes the proof of the first part of the lemma.

To prove the second part, note $2\hat{\sigma}^2 = \min_{i\neq j}\hat{q}_{ij}^2 = \min_{i\neq j}q_{ij}^2 + O_p\left((\ln n/n)^{1/2}\right)$, and $2\sigma^2 \leqslant \min_{i\neq j}q_{ij}^2 \leqslant 2\sigma^2 + \min_{i\neq j}\mathbb{E}\left[(g(\xi_i,\xi_k) - g(\xi_j,\xi_k))^2|\xi_i,\xi_j\right]$. Note that Assumption 2(iii) guarantees that $\min_{i\neq j}\mathbb{E}\left[(g(\xi_i,\xi_k) - g(\xi_j,\xi_k))^2|\xi_i,\xi_j\right] \leqslant \overline{G}^2\min_{i\neq j}\|\xi_i - \xi_j\|^2$, which delivers the result. Q.E.D.

A.3 Proofs of the results of Section 4.2.2

A.3.1 Proof of Theorem 2

First, we state prove two auxiliary lemmas.

Lemma A.5. *Suppose that the hypotheses of Theorem 2 are satisfied. Then, for any* $C > 0$*, there exists* $a_C > 0$ *such that we have* $\min_i \frac{|\mathcal{B}_i(\delta_{n,C})|}{n-1} \geqslant C(n^{-1}\ln n)^{1/2}$ *with probability approaching one, where* $\mathcal{B}_i(\delta) := \{i' \neq i : X_{i'} = X_i, \xi_{i'} \in B_\delta(\xi_i)\}$*, and* $\delta_{n,C} := a_C(n^{-1}\ln n)^{\frac{1}{2d_\xi}}$*.*

Proof of Lemma A.5. Let $p_Z(x,\xi;\delta) := \mathbb{P}(X_j = x)\mathbb{P}(\xi_j \in B_\delta(\xi)|X_j = x)$, $\mathcal{Q}_i(\delta) := \frac{|\mathcal{B}_i(\delta)|}{n-1}$, and $p_r := \mathbb{P}(X = x_r)$. Then, by Bernstein inequality A.1, we have for all $\epsilon \in (0,1)$

$$\mathbb{P}(|\mathcal{Q}_i(\delta) - p_Z(X_i,\xi_i;\delta)| \geqslant \epsilon | X_i, \xi_i) \leqslant 2\exp\left(-\frac{\frac{1}{2}(n-1)\epsilon^2}{1+\frac{1}{3}\epsilon}\right) \leqslant 2\exp\left(-\frac{1}{4}n\epsilon^2\right).$$

Hence, we also have $\mathbb{P}(|\mathcal{Q}_i(\delta) - p_Z(X_i,\xi_i;\delta)| \geqslant \epsilon) \leqslant 2\exp\left(-\frac{1}{4}n\epsilon^2\right)$ and, using the union bound, $\mathbb{P}(\max_i |\mathcal{Q}_i(\delta) - p_Z(X_i,\xi_i;\delta)| \geqslant \epsilon) \leqslant 2n\exp\left(-\frac{1}{4}n\epsilon^2\right)$. Hence, for any $\tilde{C} > 0$ (and large enough n), $\mathbb{P}(\max_i |\mathcal{Q}_i(\delta) - p_Z(X_i,\xi_i;\delta)| \geqslant \tilde{C}(n^{-1}\ln n)^{1/2}) \leqslant 2n^{1-\tilde{C}/4}$. Using Assumption 7(ii), for $\delta \leqslant \overline{\delta}$, we have $p_Z(X_i,\xi_i;\delta) \geqslant \underline{\kappa}\delta^{d_\xi}$ a.s. where $\underline{\kappa} = \kappa \min_r p_r$. Then, we have $\min p_Z(X_i,\xi_i;a(n^{-1}\ln n)^{\frac{1}{2d_\xi}}) \geqslant \underline{\kappa}a^{d_\xi}(n^{-1}\ln n)^{1/2}$ for $\delta = a(n^{-1}\ln n)^{\frac{1}{2d_\xi}}$ for a fixed $a > 0$ and for large enough n. Hence, with probability at least $1 - 2n^{1-\tilde{C}/4}$, we have

$$\min_i \mathcal{Q}_i(a(n^{-1}\ln n)^{\frac{1}{2d_\xi}}) \geqslant (\underline{\kappa}a^{d_\xi} - \tilde{C})(n^{-1}\ln n)^{1/2}. \tag{A.11}$$

Take some fixed $\tilde{C} > 4$. In this case, (A.11) holds with probability approaching one, and take any fixed $a_C > \left(\underline{\kappa}^{-1}(C + \tilde{C})\right)^{1/d_\xi}$. For such a_C, $(\underline{\kappa}a^{d_\xi} - \tilde{C})(n^{-1}\ln n)^{1/2} > C(n^{-1}\ln n)^{1/2}$. As a result, we conclude that with probability approaching one, we have $\min_i \mathcal{Q}_i(a(n^{-1}\ln n)^{\frac{1}{2d_\xi}}) > C(n^{-1}\ln n)^{1/2}$, which completes the proof. Q.E.D.

Lemma A.6. *Suppose that the hypotheses of Theorem 2 are satisfied. Then we have:*

(i) $\max_i \max_{i' \in \hat{\mathcal{N}}_i(n_i)} n^{-1}\sum_\ell (Y^*_{i\ell} - Y^*_{i'\ell})^2 = O_p\left((n^{-1}\ln n)^{\frac{1}{2d_\xi}}\right)$;

(ii) $\max_i \max_{i' \in \hat{\mathcal{N}}_i(n_i)} \max_k |n^{-1}\sum_\ell Y^*_{k\ell}(Y^*_{i'\ell} - Y^*_{i\ell})| = O_p\left((n^{-1}\ln n)^{\frac{1}{2d_\xi}}\right)$;

(iii) $\max_i \max_{i' \in \hat{\mathcal{N}}_i(n_i)} \int (g(\xi_i,\xi) - g(\xi_{i'},\xi))^2 dP_\xi(\xi) = O_p\left((n^{-1}\ln n)^{\frac{1}{2d_\xi}}\right)$.

Proof of Lemma A.6. Note that both statements of the lemma trivially hold when $i = i'$. For this reason, below $\max_{i' \in \hat{\mathcal{N}}_i(n_i)}$ should be read as $\max_{i' \in \hat{\mathcal{N}}_i(n_i): i' \neq i}$.

Proof of Part (i). First, take some $C > \overline{C}$, where $\overline{C}$ is defined in the text of Theorem 2. Lemma A.5 guarantees, that there exist some a_C and $\delta_{n,C} = a_C(n^{-1}\ln n)^{\frac{1}{2d_\xi}}$ such that with probability approaching one, we have $\min_i |\mathcal{B}_i(\delta_{n,C})| \geqslant C(n \ln n)^{1/2}$. Since we have $\max_i n_i \leqslant \overline{C}(n \ln n)^{1/2} < C(n \ln n)^{1/2}$, we conclude that $\min_i |\mathcal{B}_i(\delta_{n,C})| > n_i$ holds with probability approaching one. For any agents i and $i' \in \hat{\mathcal{N}}_i(n_i)$, let k and k' be two agents (other than i and i') such that $k \in \mathcal{B}_i(\delta_{n,C})$ and $k' \in \mathcal{B}_{i'}(\delta_{n,C})$ (by Lemma A.5, this happens simultaneously for all i and i' with probability approaching one for large enough n). By boundedness of Y^* and Assumption 2(iii), for $\mathcal{G} \coloneqq \|Y^*\|_\infty \overline{G}$, we have

$$
\begin{aligned}
&|n^{-1}\sum_\ell Y_{i\ell}^{*2} - n^{-1}\sum_\ell Y_{i\ell}^* Y_{k\ell}^*| \leqslant \mathcal{G}\delta_{n,C}, \quad |n^{-1}\sum_\ell Y_{i'\ell}^* Y_{i\ell}^* - n^{-1}\sum_\ell Y_{i'\ell}^* Y_{k\ell}^*| \leqslant \mathcal{G}\delta_{n,C}, \\
&|n^{-1}\sum_\ell Y_{i'\ell}^{*2} - n^{-1}\sum_\ell Y_{i'\ell}^* Y_{k'\ell}^*| \leqslant \mathcal{G}\delta_{n,C}, \quad |n^{-1}\sum_\ell Y_{i\ell}^* Y_{i'\ell}^* - n^{-1}\sum_\ell Y_{i\ell}^* Y_{k'\ell}^*| \leqslant \mathcal{G}\delta_{n,C} \qquad \text{(A.12)}
\end{aligned}
$$

with probability approaching one for all i, $i' \in \hat{\mathcal{N}}_i(n_i)$. Then

$$
\begin{aligned}
n^{-1}\sum_\ell (Y_{i\ell}^* - Y_{i'\ell}^*)^2 &\leqslant |n^{-1}\sum_\ell Y_{i\ell}^{*2} - n^{-1}\sum_\ell Y_{i'\ell}^* Y_{i\ell}^*| + |n^{-1}\sum_\ell Y_{i'\ell}^{*2} - n^{-1}\sum_\ell Y_{i\ell}^* Y_{i'\ell}^*| \\
&\leqslant |n^{-1}\sum_\ell (Y_{i\ell}^* - Y_{i'\ell}^*)Y_{k\ell}^*| + |n^{-1}\sum_\ell (Y_{i'\ell}^* - Y_{i\ell}^*)Y_{k'\ell}^*| + 4\mathcal{G}\delta_{n,C}.
\end{aligned}
$$

Note that since Y^* is bounded, there exists $\Delta > 0$ such that for all i, i', and k,

$$
|n^{-1}\sum_\ell (Y_{i\ell}^* - Y_{i'\ell}^*)Y_{k\ell}^*| \leqslant |(n-3)^{-1}\sum_{\ell \neq i,i',k} (Y_{i\ell}^* - Y_{i'\ell}^*)Y_{k\ell}^*| + \frac{\Delta}{n}. \qquad \text{(A.13)}
$$

Moreover, as usual, combining Corollary A.1 with the union bound, we have

$$
r_n \coloneqq \max_{i \neq i' \neq k} \left| \frac{\sum_{\ell \neq i,i',k}(Y_{i\ell} - Y_{i'\ell})Y_{k\ell} - \sum_{\ell \neq i,i',k}(Y_{i\ell}^* - Y_{i'\ell}^*)Y_{k\ell}^*}{n-3} \right| = O_p\left(\left(\frac{\ln n}{n}\right)^{1/2}\right). \qquad \text{(A.14)}
$$

So, we have

$$
\begin{aligned}
&\max_i \max_{i' \in \hat{\mathcal{N}}_i(n_i)} n^{-1}\sum_\ell (Y_{i\ell}^* - Y_{i'\ell}^*)^2 \leqslant \max_i \max_{i' \in \hat{\mathcal{N}}_i(n_i)} \max_{k \neq i,i'} |(n-3)^{-1}\sum_{\ell \neq i,i',k}(Y_{i\ell} - Y_{i'\ell})Y_{k\ell}| \\
&+ \max_i \max_{i' \in \hat{\mathcal{N}}_i(n_i)} \max_{k' \neq i,i'} |(n-3)^{-1}\sum_{\ell \neq i,i',k}(Y_{i'\ell} - Y_{i\ell})Y_{k'\ell}| + 4\mathcal{G}\delta_{n,C} + \frac{2\Delta}{n} + 2r_n \\
&= 2\max_i \max_{i' \in \hat{\mathcal{N}}_i(n_i)} \hat{d}_\infty^2(i,i') + O_p((n^{-1}\ln n)^{\frac{1}{2d_\xi}}), \qquad \text{(A.15)}
\end{aligned}
$$

where the last line uses the definitions of $\hat{d}^2_\infty(i,i')$ and $\delta_{n,C} = a_C(n^{-1}\ln n)^{\frac{1}{2d_\xi}}$.

Next, let $\hat{d}^2_{\infty,*}(i,i') := \max_{k\neq i,i'} |(n-3)^{-1}\sum_{\ell\neq i,i',k}(Y^*_{i\ell} - Y^*_{i'\ell})Y^*_{k\ell}|$. For any i and i', (A.14) guarantees that $\max_i \max_{i'\neq i} |\hat{d}^2_\infty(i,i') - \hat{d}^2_{\infty,*}(i,i')| = O_p\left((n^{-1}\ln n)^{1/2}\right)$. Analogously to (A.12), for all i and $i' \in \mathcal{B}_i(\delta_{n,C})$, we have $\hat{d}^2_{\infty,*}(i,i') \leqslant \mathcal{G}\delta_{n,C}$ and hence also

$$\hat{d}^2_\infty(i,i') \leqslant \mathcal{G}\delta_{n,C} + O_{p,n}\left((n^{-1}\ln n)^{1/2}\right).$$

Then, by definition of $\mathcal{N}_i$ and the fact that for all i we have $n_i < |\mathcal{B}_i(\delta_{n,C})|$ with probability approaching one, we conclude

$$\max_i \max_{i'\in\hat{\mathcal{N}}_i(n_i)} \hat{d}^2_\infty(i,i') \leqslant \mathcal{G}\delta_{n,C} + O_{p,n}\left((n^{-1}\ln n)^{1/2}\right) = O_{p,n}\left((n^{-1}\ln n)^{\frac{1}{2d_\xi}}\right), \tag{A.16}$$

which together with (A.15) completes the proof of Part (i).

Proof of Part (ii). We prove the result by bounding the quantity of interest for $k \neq i,i'$ and then for $k = i$ (or $k = i'$). First, we consider $k \neq i,i'$. Then, using (A.13) and (A.14),

$$\max_i \max_{i'\in\hat{\mathcal{N}}_i(n_i)} \max_{k\neq i,i'} |n^{-1}\sum_\ell Y^*_{k\ell}(Y^*_{i'\ell} - Y^*_{i\ell})| = \max_i \max_{i'\in\hat{\mathcal{N}}_i(n_i)} \max_{k\neq i,i'} |(n-3)^{-1}\sum_{\ell\neq i,i',k} Y_{k\ell}(Y_{i'\ell} - Y_{i\ell})|$$
$$+O_p\left((\ln n/n)^{1/2}\right) = \max_i \max_{i'\in\hat{\mathcal{N}}_i(n_i)} \hat{d}^2_\infty(i,i') + O_p\left((\ln n/n)^{1/2}\right) = O_{p,n}\left((n^{-1}\ln n)^{\frac{1}{2d_\xi}}\right), \tag{A.17}$$

where the last equality uses (A.16).

To complete the proof, consider $k = i$ (the case $k = i'$ is analogous). Then, as argued at the beginning of the proof of Part (i), for any pair agents i and i', there exist some $\tilde{k} \in \mathcal{B}_i(\delta_{n,C})$ (other than i and i') with probability approaching one. Then, by boundedness of Y^* and Assumption 2(iii),

$$\max_i \max_{i'\in\hat{\mathcal{N}}_i(n_i)} |n^{-1}\sum_\ell Y^*_{i\ell}(Y^*_{i'\ell} - Y^*_{i\ell})| = \max_i \max_{i'\in\hat{\mathcal{N}}_i(n_i)} |n^{-1}\sum_\ell Y^*_{\tilde{k}\ell}(Y^*_{i'\ell} - Y^*_{i\ell})| + O(\delta_{n,C})$$
$$= O_{p,n}\left((n^{-1}\ln n)^{\frac{1}{2d_\xi}}\right),$$

where the second equality follows from (A.17). This completes the proof of Part (ii).

Proof of Part (iii). Recall that for $i' \in \hat{\mathcal{N}}_i(n_i)$, we have $Y^*_{i\ell} - Y^*_{i'\ell} = g(\xi_i,\xi_\ell) - g(\xi_{i'},\xi_\ell)$. Hence, the result of Part (i) reads as

$$\max_i \max_{i'\in\hat{\mathcal{N}}_i(n_i)} n^{-1}\sum_\ell (g(\xi_i,\xi_\ell) - g(\xi_{i'},\xi_\ell))^2 = O_p\left((n^{-1}\ln n)^{\frac{1}{2d_\xi}}\right). \tag{A.18}$$

Applying Bernstein inequality A.1 conditional on ξ_i and $\xi_{i'}$ and the union bound, we obtain

$$\max_{i,i'}\left|n^{-1}\sum_{\ell}(g(\xi_i,\xi_\ell)-g(\xi_{i'},\xi_\ell))^2-\int(g(\xi_i,\xi)-g(\xi_{i'},\xi))^2dP_\xi(\xi)\right|=O_p\left((\ln n/n)^{1/2}\right). \tag{A.19}$$

Finally, (A.18) and (A.19) together deliver the result. Q.E.D.

Proof of Theorem 2. Proof of Part (i). First, notice that

$$n^{-1}\sum_j\left(\hat{Y}^*_{ij}-Y^*_{ij}\right)^2=n^{-1}\sum_j\left(\frac{\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}(Y_{i'j}-Y^*_{ij})}{n_i}\right)^2\leq 2(\mathcal{S}_i+\mathcal{J}_i), \tag{A.20}$$

$$\mathcal{S}_i:=\frac{1}{n}\sum_j\left(\frac{\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}(Y_{i'j}-Y^*_{i'j})}{n_i}\right)^2,\quad \mathcal{J}_i:=\frac{1}{n}\sum_j\left(\frac{\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}(Y^*_{i'j}-Y^*_{ij})}{n_i}\right)^2.$$

We start with bounding $\mathcal{S}_i$. Note that, for all i' and j, $Y_{i'j}-Y^*_{i'j}=\varepsilon_{i'j}$, with the convention $\varepsilon_{i'j}=-Y^*_{i'i'}$ when $i'=j$. Then $\mathcal{S}_i$ can be decomposed as follows

$$\begin{aligned}\mathcal{S}_i&=\frac{1}{nn_i^2}\sum_j\left(\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}\varepsilon^2_{i'j}+\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}\sum_{i''\in\hat{\mathcal{N}}_i(n_i),i''\neq i'}\varepsilon_{i'j}\varepsilon_{i''j}\right)\\&=\underbrace{\frac{1}{n_i^2}\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}\frac{1}{n}\sum_j\varepsilon^2_{i'j}}_{:=\mathcal{S}_{1i}}+\underbrace{\frac{1}{n_i^2}\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}\sum_{i''\in\hat{\mathcal{N}}_i(n_i),i''\neq i'}\frac{1}{n}\sum_j\varepsilon_{i'j}\varepsilon_{i''j}}_{:=\mathcal{S}_{2i}}.\end{aligned}$$

First, $\frac{1}{n}\sum_j\varepsilon^2_{i'j}=\frac{1}{n}\sum_{j\neq i'}\varepsilon^2_{i'j}-\frac{Y^*_{i'i'}}{n}$. Applying Corollary A.1 (conditional on $Z_{i'}$) and the union bound gives $\max_{i'}|\frac{1}{n-1}\sum_{j\neq i'}\varepsilon^2_{i'j}-\sigma^2_{i'}|=O_p\left((\ln n/n)^{1/2}\right)$, where $\sigma^2_i:=\mathbb{E}\left[\varepsilon^2_{ij}|Z_i\right]\leqslant C$ a.s. by Assumption 2(iv). Finally, since Y^* is bounded, we conclude

$$\max_{i'}\frac{1}{n}\sum_j\varepsilon^2_{i'j}\leqslant C+O_p\left((\ln n/n)^{1/2}\right),\quad \max_i\mathcal{S}_{1i}\leqslant\frac{C}{n_i}+o_p(n_i^{-1}). \tag{A.21}$$

To bound $\mathcal{S}_{2i}$, note $\frac{1}{n}\sum_j\varepsilon_{i'j}\varepsilon_{i''j}=\frac{1}{n}\sum_{j\neq i',i''}\varepsilon_{i'j}\varepsilon_{i''j}-\frac{1}{n}(Y^*_{i'i'}\varepsilon_{i''i'}+\varepsilon_{i'i''}Y^*_{i''i''})$. First, as usual, applying Bernstein inequality A.2 (conditional on $Z_{i'}$ and $Z_{i''}$) and the union bound gives $\max_{i'\neq i''}|\frac{1}{n-2}\sum_{j\neq i',i''}\varepsilon_{i'j}\varepsilon_{i''j}|=O_p\left((\ln n/n)^{1/2}\right)$. Similarly, since $\varepsilon_{i'i''}$'s are (uniformly) sub-Gaussian, we have $\max_{i'\neq i''}\left|\frac{1}{n}\varepsilon_{i'i''}\right|=O_p\left(n^{-1}(\ln n)^{1/2}\right)$. Since Y^* is bounded, we conclude

$$\max_{i'\neq i''}|\frac{1}{n}\sum_j\varepsilon_{i'j}\varepsilon_{i''j}|=O_p\left((\ln n/n)^{1/2}\right),\quad \max_i|\mathcal{S}_{2i}|=O_p\left((\ln n/n)^{1/2}\right). \tag{A.22}$$

Combining (A.21) and (A.22) gives $\max_i \mathcal{S}_i = O_p\left((\ln n/n)^{1/2}\right)$.

Finally, note $\mathcal{J}_i \leqslant \frac{1}{n}\sum_j \frac{1}{n_i}\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}(Y^*_{i'j} - Y^*_{ij})^2 = \frac{1}{n_i}\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}\frac{1}{n}\sum_j(Y^*_{i'j} - Y^*_{ij})^2$. Applying Lemma A.6(i), we obtain $\max_i \mathcal{J}_i = O_p\left((n^{-1}\ln n)^{\frac{1}{2d_\xi}}\right)$. This, together with the bound on $\mathcal{S}_i$ and (A.20), delivers the desired result.

Proof of Part (ii). First, notice that

$$n^{-1}\sum_\ell Y^*_{k\ell}(\hat{Y}^*_{i\ell} - Y^*_{i\ell}) = n_i^{-1}\sum_{i'\in\hat{\mathcal{N}}_i(n_i)}\left(n^{-1}\sum_\ell Y^*_{k\ell}(Y^*_{i'\ell} - Y^*_{i\ell}) + n^{-1}\sum_\ell Y^*_{k\ell}\varepsilon_{i'\ell}\right),$$

$$\max_k\max_i|n^{-1}\sum_\ell Y^*_{k\ell}(\hat{Y}^*_{i\ell} - Y^*_{i\ell})| \leqslant \underbrace{\max_i\max_{i'\in\hat{\mathcal{N}}_i(n_i)}\max_k|n^{-1}\sum_\ell Y^*_{k\ell}(Y^*_{i'\ell} - Y^*_{i\ell})|}_{=O_p\left(\left(\frac{\ln n}{n}\right)^{\frac{1}{2d_\xi}}\right)} + \max_k\max_{i'}|n^{-1}\sum_\ell Y^*_{k\ell}\varepsilon_{i'\ell}|,$$

where the bound on the first term on the right-hand side follows from Lemma A.6(ii). To complete the proof, we need to to bound the second term, which can be represented as

$$n^{-1}\sum_\ell Y^*_{k\ell}\varepsilon_{i'\ell} = n^{-1}\sum_{\ell\neq k,i'} Y^*_{k\ell}\varepsilon_{i'\ell} + n^{-1}(Y^*_{kk}\varepsilon_{i'k} + Y^*_{ki'}\varepsilon_{i'i'}). \tag{A.23}$$

Again, applying Corollary A.1 (conditional on Z_k and $Z_{i'}$) and the union bound, we obtain $\max_k\max_{i'}|n^{-1}\sum_{\ell\neq k,i'}Y^*_{k\ell}\varepsilon_{i'\ell}| = O_p\left((\ln n/n)^{1/2}\right)$. Similarly, since either $\varepsilon_{i'k}$ is (uniformly) sub-Gaussian (for $i' \neq k$) or is equal to $-Y^*_{kk}$ (which is bounded), we have $\max_k\max_{i'}|n^{-1}Y^*_{kk}\varepsilon_{i'k}| = O_p\left(n^{-1}(\ln n)^{1/2}\right)$. Finally, since $\varepsilon_{i'i'} = -Y^*_{i'i'}$ and Y^* is bounded, we have $\max_k\max_{i'}|n^{-1}Y^*_{ki'}\varepsilon_{i'i'}| \leqslant C/n$. These bounds, together with (A.23), imply that $\max_k\max_{i'}|n^{-1}\sum_\ell Y^*_{k\ell}\varepsilon_{i'\ell}| = O_p\left((\ln n/n)^{1/2}\right)$, which completes the proof. Q.E.D.

A.3.2 Proof of Theorem 3

First, we state and prove the following auxiliary lemma.

Lemma A.7. *Suppose $\mathcal{B} = \mathbb{R}^p$. Then, under Assumptions 1, 2, 7(i), we have $\max_{i\neq j}\left|d^2_{ij,n} - d^2_{ij}\right| = O_p\left((\ln n/n)^{1/2}\right)$, where $d^2_{ij,n} \coloneqq \min_{\beta\in\mathbb{R}_p} d^2_{ij,n}(\beta)$, and $d^2_{ij,n}(\beta)$ is given by (A.7).*

Proof of Lemma A.7. First, we argue that there exists some compact $\overline{\mathcal{B}}$ such that, for all pairs of agents, $d^2_{ij,n} \coloneqq \min_{\beta\in\mathbb{R}^p} d^2_{ij,n}(\beta) = \min_{\beta\in\overline{\mathcal{B}}} d^2_{ij,n}(\beta)$ with probability approaching one. For all $x, \tilde{x} \in \operatorname{supp}(X)$, let $\Delta w(X_k; x, \tilde{x}) \coloneqq w(x, X_k) - w(\tilde{x}, X_k)$ and

$$\hat{\mathcal{C}}(x,\tilde{x}) \coloneqq \frac{1}{n}\sum_{k=1}^n \Delta w(X_k; x,\tilde{x})\Delta w(X_k; x,\tilde{x})' = \sum_{r=1}^R \hat{p}_r H(x_r; x,\tilde{x})',$$

where $\hat{p}_r = n^{-1} \sum_{k=1}^n \mathbb{1}\{X_k = x_r\}$ and $H(x_r; x, \tilde{x}) \coloneqq \Delta w(x_r; x, \tilde{x}) \Delta w(x_r; x, \tilde{x})'$.

Notice that since $\operatorname{supp}(X) = \{x_1, \ldots, x_R\}$, there exists $C_\lambda > 0$ such that the minimal non-zero eigenvalue of $H(x_r; x, \tilde{x})$ is greater than C_λ uniformly over all $x_r, x, \tilde{x} \in \{x_1, \ldots, x_R\}$. Formally, we have $\lambda_{min,+}(H(x_r; x, \tilde{x})) > C_\lambda$ for all $x_r, x, \tilde{x} \in \{x_1, \ldots, x_R\}$ such that $H(x_r; x, \tilde{x})$ is non-zero, where $\lambda_{min,+}(H)$ denotes the minimal positive eigenvalue of a positive semidefinite (non-zero) matrix H. Next, notice that there exists $C_p > 0$ such that with probability approaching one, we have $\min_r \hat{p}_r > C_p$. Hence, we conclude that for some $C_{\mathcal{C}} > 0$ we have, with probability approaching one, $\lambda_{min,+}(\hat{\mathcal{C}}(x, \tilde{x})) > C_{\mathcal{C}}$ for all $x, \tilde{x} \in \{x_1, \ldots, x_R\}$ such that $\hat{\mathcal{C}}(x, \tilde{x})$ is non-zero. Next, for all $x, \tilde{x} \in \{x_1, \ldots, x_R\}$, let $\hat{O}(x, \tilde{x})$ be an orthogonal matrix, which diagonalizes $\hat{\mathcal{C}}(x, \tilde{x})$, i.e., $\hat{O}(x, \tilde{x}) \hat{\mathcal{C}}(x, \tilde{x}) \hat{O}(x, \tilde{x})' = \hat{\Lambda}(x, \tilde{x})$, where $\hat{\Lambda}(x, \tilde{x})$ is diagonal. Notice that non-zero elements $\hat{\Lambda}(x, \tilde{x})$ of are bounded away from zero by $C_{\mathcal{C}}$ with probability approaching (uniformly over $x, \tilde{x} \in \{x_1, \ldots, x_R\}$).

Take any pair of agents i and j. Since $\hat{O}(X_i, X_j)' \hat{O}(X_i, X_j)$ is an identity matrix,

$$
\begin{aligned}
d_{ij,n}^2 &= \min_{\beta \in \mathbb{R}_p} \frac{1}{n} \sum_k (Y_{ik}^* - Y_{jk}^* - \Delta w(X_k; X_i, X_j)' \hat{O}(X_i, X_j)' \hat{O}(X_i, X_j)\ \beta)^2 \\
&= \min_{b \in \mathbb{R}_p} \frac{1}{n} \sum_k (Y_{ik}^* - Y_{jk}^* - \Delta w(X_k; X_i, X_j)' \hat{O}(X_i, X_j)' b)^2,
\end{aligned}
$$

where the last equality follows from the change of the variable $b = \hat{O}(X_i, X_j)\beta$. The minimum is achieved at

$$
\hat{b}_{ij} = \left(\hat{\Lambda}(X_i, X_j)\right)^+ \left(\hat{O}(X_i, X_j) n^{-1} \sum_k \Delta w(X_k; X_i, X_j)(Y_{ik}^* - Y_{jk}^*)\right),
$$

where $+$ denotes the Moore–Penrose inverse. Notice that since the non-zero elements of diagonal matrix $\hat{\Lambda}(X_i, X_j)$ are bounded away from zero (uniformly in X_i and X_j) with probability approaching one, we conclude that $\max_{i \neq j} \lambda_{max}\left(\left(\hat{\Lambda}(X_i, X_j)\right)^+\right) < \overline{C}_\Lambda$ with probability approaching one. Since w and Y^* are bounded, we also have $\max_{i \neq j} \left\| \hat{O}(X_i, X_j) n^{-1} \sum_k w(X_k; X_i, X_j)(Y_{ik}^* - Y_{jk}^*) \right\| < C$, and thus we conclude that for some $C_\beta > 0$, we have with probability approaching one $\max_{i \neq j} \|\hat{b}_{ij}\| \leqslant C_\beta$.

Since $b = \hat{O}(X_i, X_j)\beta$, the minimum of $d_{ij,n}^2(\beta)$ is achieved at $\hat{\beta}_{ij} = \hat{O}(X_i, X_j)' \hat{b}_{ij}$, i.e., $d_{ij,n}^2 = d_{ij,n}^2(\hat{\beta}_{ij})$, and, with probability approaching one, we also have $\max_{i \neq j} \|\hat{\beta}_{ij}\| \leqslant C_\beta$. Then, if $\overline{\mathcal{B}}$ includes all β such that $\|\beta\| \leqslant C_\beta$, we have $d_{ij,n}^2 = d_{ij,n}^2(\hat{\beta}_{ij}) = \min_{\beta \in \overline{\mathcal{B}}} d_{ij,n}^2(\beta)$. Such a compact set $\overline{\mathcal{B}}$ clearly exists, the proof of the first part is complete.

Second, note that by the same argument, we have $d_{ij}^2 = \min_{\beta \in \mathbb{R}_p} d_{ij}^2(\beta) = \min_{\beta \in \overline{\mathcal{B}}} d_{ij}^2(\beta)$, where, without loss of generality, $\overline{\mathcal{B}}$ can be taken the same as before.

Finally, since $\overline{\mathcal{B}}$ is compact, we can apply the result of Lemma A.4, which guarantees that $\max_{i\neq j}\sup_{\beta\in\overline{\mathcal{B}}}\left|d^2_{ij,n}(\beta)-d^2_{ij}(\beta)\right| = O_p\left((\ln n/n)^{1/2}\right)$. This necessarily implies that

$$O_p\left((\ln n/n)^{1/2}\right) = \max_{i\neq j}|\min_{\beta\in\overline{\mathcal{B}}} d^2_{ij,n}(\beta) - \min_{\beta\in\overline{\mathcal{B}}} d^2_{ij}(\beta)| = \max_{i\neq j}\left|d^2_{ij,n}-d^2_{ij}\right|,$$

which completes the proof. Q.E.D.

Proof of Theorem 3. First, denote

$$\mathbf{W}_{i-j} = \begin{pmatrix}(W_{i1}-W_{j1})' \\ (W_{i2}-W_{j2})' \\ \cdots \\ (W_{in}-W_{jn})'\end{pmatrix}, \quad \mathbf{Y}^*_{i-j} = \begin{pmatrix}Y^*_{i1}-Y^*_{j1} \\ Y^*_{i2}-Y^*_{j2} \\ \cdots \\ Y^*_{in}-Y^*_{jn}\end{pmatrix}, \quad \hat{\mathbf{Y}}^*_{i-j} = \begin{pmatrix}\hat{Y}^*_{i1}-\hat{Y}^*_{j1} \\ \hat{Y}^*_{i2}-\hat{Y}^*_{j2} \\ \cdots \\ \hat{Y}^*_{in}-\hat{Y}^*_{jn}\end{pmatrix}.$$

Then

$$\hat{d^2_{ij}} = n^{-1}\left(\mathbf{Y}^{*\prime}_{i-j}\mathbf{P}_{i-j}\mathbf{Y}^*_{i-j} + \Delta\hat{\mathbf{Y}}^{*\prime}_{i-j}\mathbf{P}_{i-j}\mathbf{Y}^*_{i-j} + \mathbf{Y}^{*\prime}_{i-j}\mathbf{P}_{i-j}\Delta\hat{\mathbf{Y}}^*_{i-j} + \Delta\hat{\mathbf{Y}}^{*\prime}_{i-j}\mathbf{P}_{i-j}\Delta\hat{\mathbf{Y}}^*_{i-j}\right),$$

where $\Delta\hat{\mathbf{Y}}^*_{i-j} = \hat{\mathbf{Y}}^*_{i-j} - \mathbf{Y}^*_{i-j}$ and $\mathbf{P}_{i-j} = \mathbf{I}_n - \mathbf{W}_{i-j}(\mathbf{W}'_{i-j}\mathbf{W}_{i-j})^+\mathbf{W}'_{i-j}$, where $+$ stands for the Moore–Penrose inverse. Note that $n^{-1}\mathbf{Y}^{*\prime}_{i-j}\mathbf{P}_{i-j}\mathbf{Y}^*_{i-j} = d^2_{ij,n}$, where $d^2_{ij,n}$ is as defined in Lemma A.7, which ensures $\max_{i\neq j}\left|d^2_{ij,n}-d^2_{ij}\right| = O_p((\ln n/n)^{1/2})$. Hence, to complete the proof, it would be sufficient to show that $\max_{i\neq j}|n^{-1}\Delta\hat{\mathbf{Y}}^{*\prime}_{i-j}\mathbf{P}_{i-j}\mathbf{Y}^*_{i-j}| = O_p\left((\ln n/n)^{\frac{1}{2d_\xi}}\right)$, and $\max_{i\neq j}|n^{-1}\Delta\hat{\mathbf{Y}}^{*\prime}_{i-j}\mathbf{P}_{i-j}\Delta\hat{\mathbf{Y}}^*_{i-j}| = O_p\left((\ln n/n)^{\frac{1}{2d_\xi}}\right)$. To inspect these, note

$$\begin{aligned}\max_{i\neq j}|n^{-1}\Delta\hat{\mathbf{Y}}^{*\prime}_{i-j}\mathbf{P}_{i-j}\mathbf{Y}^*_{i-j}| &\leqslant \max_{i\neq j}|n^{-1}\Delta\hat{\mathbf{Y}}^{*\prime}_{i-j}\mathbf{Y}^*_{i-j}| = \max_{i\neq j}|n^{-1}\sum_\ell(Y^*_{i\ell}-Y^*_{j\ell})(\Delta\hat{Y}^*_{i\ell}-\Delta\hat{Y}^*_{j\ell})| \\ &\leqslant 4\max_k\max_i|n^{-1}\sum_\ell Y^*_{k\ell}(\hat{Y}^*_{i\ell}-Y^*_{i\ell})| = O_p\left((\ln n/n)^{\frac{1}{2d_\xi}}\right),\end{aligned}$$

where the last equality follows from Theorem 2(ii). Similarly,

$$\begin{aligned}\max_{i\neq j}|n^{-1}\Delta\hat{\mathbf{Y}}^{*\prime}_{i-j}\mathbf{P}_{i-j}\Delta\hat{\mathbf{Y}}^*_{i-j}| &\leqslant \max_{i\neq j}|n^{-1}\Delta\hat{\mathbf{Y}}^{*\prime}_{i-j}\Delta\hat{\mathbf{Y}}^*_{i-j}| = \max_{i\neq j}|n^{-1}\sum_\ell(\Delta\hat{Y}^*_{i\ell}-\Delta\hat{Y}^*_{j\ell})^2| \\ &\leqslant 4\max_i n^{-1}\sum_\ell(\hat{Y}^*_{i\ell}-Y^*_{i\ell})^2 = O_p\left((\ln n/n)^{\frac{1}{2d_\xi}}\right),\end{aligned}$$

where the last equality is due Theorem 2(i), which completes the proof. Q.E.D.

A.3.3 Proof of Lemma 2

Proof of Lemma 2. Let $\hat{d}^2_{ij}(\beta) \coloneqq \frac{1}{n}\sum_k(\hat{Y}^*_{ik} - \hat{Y}^*_{jk} - (W_{ik} - W_{jk})\beta)^2$, so $\hat{d}^2_{ij} = \min_{\beta\in\mathcal{B}} \hat{d}^2_{ij}(\beta)$.

$$\max_{i\neq j} |\hat{d}^2_{ij} - d^2_{ij}| = \max_{i\neq j} |\min_{\beta\in\mathcal{B}} \hat{d}^2_{ij}(\beta) - \min_{\beta\in\mathcal{B}} d^2_{ij}(\beta)| \leqslant 2 \max_{i\neq j} \max_{\beta\in\mathcal{B}} |\hat{d}^2_{ij}(\beta) - d^2_{ij}(\beta)|.$$

Hence, it suffices to show $\max_{i\neq j}\max_{\beta\in\mathcal{B}} \|\hat{d}^2_{ij}(\beta) - d^2_{ij}(\beta)\| = O_p\left((\ln n/n)^{1/2} + \mathcal{R}_n^{-1/2}\right)$. Let $\Delta\hat{Y}^*_{ik} \coloneqq \hat{Y}^*_{ik} - Y^*_{ik}$, and note that $\hat{d}^2_{ij}(\beta) - d_{ij}(\beta)$ can be decomposed as

$$\begin{aligned}
&\hat{d}^2_{ij}(\beta) - d_{ij}(\beta) = d^2_{ij,n}(\beta) - d_{ij}(\beta) \\
&+ \frac{2}{n}\sum_k \left(Y^*_{ik} - Y^*_{jk} + (W_{ik} - W_{jk})'\beta\right)(\Delta\hat{Y}^*_{ik} - \Delta\hat{Y}^*_{jk}) + \frac{1}{n}\sum_k (\Delta\hat{Y}^*_{ik} - \Delta\hat{Y}^*_{jk})^2. \qquad \text{(A.24)}
\end{aligned}$$

By the Cauchy-Schwartz inequality,

$$\max_{i\neq j} \frac{1}{n}\sum_k (\Delta\hat{Y}^*_{ik} - \Delta\hat{Y}^*_{jk})^2 \leqslant 4 \max_i \frac{1}{n}\sum_i (\hat{Y}^*_{ik} - Y^*_{ik})^2 = O_p(\mathcal{R}_n^{-1}),$$

$$\max_{i\neq j}\max_{\beta\in\mathcal{B}} \left|\frac{1}{n}\sum_k (Y^*_{ik} - Y^*_{jk} + (W_{ik} - W_{jk})'\beta)(\Delta\hat{Y}^*_{ik} - \Delta\hat{Y}^*_{jk})\right| \leqslant C\sqrt{\max_{i\neq j}\frac{1}{n}\sum_k (\Delta\hat{Y}^*_{ik} - \Delta\hat{Y}^*_{jk})^2} = O_p(\mathcal{R}_n^{-1/2}),$$

where the second inequality guaranteed by the fact that Y^*, W, and $\mathcal{B}$ are bounded. These bounds, paired with Lemma A.4 and (A.23), deliver the result. Q.E.D.

A.4 Proof of Theorem 4

Proof of Theorem 4. First, we decompose $\tilde{Y}^*_{ij}$ as

$$\tilde{Y}^*_{ij} = \frac{1}{m_{ij}} \sum_{(i',j')\in\hat{\mathcal{M}}_{ij}} Y_{i'j'} = Y^*_{ij} + \frac{1}{m_{ij}} \sum_{(i',j')\in\hat{\mathcal{M}}_{ij}} \left((g(\xi_{i'}, \xi_{j'}) - g(\xi_i, \xi_j)) + \varepsilon_{i'j'}\right).$$

To prove the result, it suffices to show that $\max_{i,j} |\frac{1}{m_{ij}} \sum_{(i',j')\in\hat{\mathcal{M}}_{ij}} (g(\xi_{i'}, \xi_{j'}) - g(\xi_i, \xi_j))|$ and $\max_{i,j} |\frac{1}{m_{ij}} \sum_{(i',j')\in\hat{\mathcal{M}}_{ij}} \varepsilon_{i'j'}|$ are both $o_p(1)$.

First, notice that Lemma A.6(iii) combined with the hypotheses of the theorem implies that $\max_i \max_{i'\in\hat{\mathcal{N}}_i(n_i)} \|\xi_i - \xi_{i'}\| = o_p(1)$. Hence, Using Assumption 2(iii),

$$\begin{aligned}
\max_{i,j} \left|\frac{1}{m_{ij}} \sum_{(i',j')\in\hat{\mathcal{M}}_{ij}} (g(\xi_{i'}, \xi_{j'}) - g(\xi_i, \xi_j))\right| &\leqslant \max_{i,j} \max_{i'\in\hat{\mathcal{N}}_i(n_i)} \max_{j'\in\hat{\mathcal{N}}_j(n_j)} |g(\xi_{i'}, \xi_{j'}) - g(\xi_i, \xi_j)| \\
&\leqslant 2\overline{G} \max_i \max_{i'\in\hat{\mathcal{N}}_i(n_i)} \|\xi_i - \xi_{i'}\| = o_p(1).
\end{aligned}$$

Next, let $\overline{\mathcal{M}}$ denote the set of all possible realizations of $\hat{\mathcal{M}}_{ij}$ formed according to (4.12). Unlike $\hat{\mathcal{M}}_{ij}$, $\overline{\mathcal{M}}$ is not random. Note that by construction

$$\max_{i,j}\left|\frac{1}{m_{ij}}\sum_{(i',j')\in\hat{\mathcal{M}}_{ij}}\varepsilon_{i'j'}\right| \leqslant \max_{\mathcal{M}\in\overline{\mathcal{M}}}\left|\frac{1}{|\mathcal{M}|}\sum_{(i',j')\in\mathcal{M}}\varepsilon_{i'j'}\right|. \tag{A.25}$$

Note that, for $\mathcal{M}\in\overline{\mathcal{M}}$, $|\mathcal{M}|\geqslant \underline{m}:=\underline{n}(\underline{n}-1)/2$, where $\underline{n}=\min_i n_i$. Applying Bernstein inequality A.2 conditional on $\{X_i,\xi_i\}_{i=1}^n$ and using uniform sub-Gaussianity of the errors, we conclude that there exist some positive constants C, a, and b such that for all $\epsilon>0$ and for all $\mathcal{M}\in\overline{\mathcal{M}}$

$$\mathbb{P}\left(\left|\frac{1}{|\mathcal{M}|}\sum_{(i',j')\in\mathcal{M}}\varepsilon_{i'j'}\right|>\epsilon\right)\leqslant C\exp\left(-\frac{|\mathcal{M}|\,\epsilon^2}{a+b\epsilon}\right)\leqslant C\exp\left(-\frac{\underline{m}\epsilon^2}{a+b\epsilon}\right). \tag{A.26}$$

To apply the union bound, we need to bound the cardinality of $\overline{\mathcal{M}}$. Note that $\left|\overline{\mathcal{M}}\right|\leqslant\left|\overline{\mathcal{N}}\right|^2$, where $\left|\overline{\mathcal{N}}\right|$ is the total number of possible realizations of $\hat{\mathcal{N}}_i(n_i)$ also allowing for possible variability in n_i. For a given n_i, the number of possible realizations of $\hat{\mathcal{N}}_i(n_i)$ is $\binom{n}{n_i}\leqslant\binom{n}{\overline{n}}$ with $\overline{n}=\max_i n_i$, where we also used monotonicity of $\binom{n}{k}$ for $k\leqslant\lceil n/2\rceil$. Hence, factoring in variability in n_i taking possible values from $\underline{n}$ to $\overline{n}$, we conclude that $\left|\overline{\mathcal{N}}\right|\leqslant\overline{n}^2\binom{n}{\overline{n}}$. Thus, combining the union bound with (A.26), we conclude

$$\mathbb{P}\left(\max_{\mathcal{M}\in\overline{\mathcal{M}}}\left|\frac{1}{|\mathcal{M}|}\sum_{(i',j')\in\mathcal{M}}\varepsilon_{i'j'}\right|>\epsilon\right)\leqslant\overline{n}^4\binom{n}{\overline{n}}^2 C\exp\left(-\frac{\underline{m}\epsilon^2}{a+b\epsilon}\right). \tag{A.27}$$

Since $\binom{n}{\overline{n}}<(\frac{n\times e}{\overline{n}})^{\overline{n}}$, we have $\ln\left(\binom{n}{\overline{n}}\right)<\overline{n}\ln(n)<\overline{C}n^{1/2}(\ln n)^{3/2}$ for sufficiently large $\overline{n}$. Since $\underline{m}>C_m n\ln n$ for some $C_m>0$, we conclude that the probability in (A.27) goes to zero for any fixed $\epsilon>0$, which, together with (A.25), completes the proof. Q.E.D.

A.5 Proofs of the results of Section 5.1

Proof of Lemma 3. The proof is by contradiction. Suppose that for some $\mu^*_{ij}(x,\tilde{x})\neq h(x,\tilde{x})$, we have $d^2_{ij}(x,\tilde{x})=d^2_{ij}(\mu^*_{ij}(x,\tilde{x}),x,\tilde{x})=0$, which implies

$$\begin{aligned}0&=\mathbb{E}\left[\left(Y^*_{ik}-Y^*_{jk}+\mu^*_{ij}(x,\tilde{x})\right)^2|X_i,\xi_i,X_j,\xi_j,X_k=x\right]\\&=\mathbb{E}\left[\left(-h(x,\tilde{x})+g(\xi_i,\xi_k)-g(\xi_j,\xi_k)+\mu^*_{ij}(x,\tilde{x})\right)^2|X_i,\xi_i,X_j,\xi_j,X_k=x\right].\end{aligned}$$

Since $\mathcal{E}_{x,\tilde{x}} \subseteq \text{supp}\,(\xi_k|X_k = x)$, the above implies

$$g(\xi_i, \xi_k) - g(\xi_j, \xi_k) = h(x, \tilde{x}) - \mu^*_{ij}(x, \tilde{x}) \neq 0 \tag{A.28}$$

for all $\xi_k \in \text{supp}\,(\xi_k|X_k = x)$. Next, consider

$$\begin{aligned}
&\mathbb{E}\left[\left(Y^*_{ik} - Y^*_{jk} - \mu^*_{ij}(x, \tilde{x})\right)^2 |X_i, \xi_i, X_j, \xi_j, X_k = \tilde{x}\right] \\
=&\mathbb{E}\left[\left(h(x, \tilde{x}) + g(\xi_i, \xi_k) - g(\xi_j, \xi_k) - \mu^*_{ij}(x, \tilde{x})\right)^2 |X_i, \xi_i, X_j, \xi_j, X_k = \tilde{x}\right] \\
\geqslant&\mathbb{P}(\xi_k \in \mathcal{E}_{x,\tilde{x}}|X_k = \tilde{x})\mathbb{E}\left[\left(h(x, \tilde{x}) + g(\xi_i, \xi_k) - g(\xi_j, \xi_k) - \mu^*_{ij}(x, \tilde{x})\right)^2 |X_i, \xi_i, X_j, \xi_j, X_k = \tilde{x}, \xi_k \in \mathcal{E}_{x,\tilde{x}}\right] \\
=&4\mathbb{P}(\xi_k \in \mathcal{E}_{x,\tilde{x}}|X_k = \tilde{x})(h(x, \tilde{x}) - \mu^*_{ij}(x, \tilde{x}))^2 > 0,
\end{aligned} \tag{A.29}$$

where the last equality follows from (A.28), and the last inequality follows from (A.28) and Assumption 8(ii). Note that (A.29) implies that $d^2_{ij}(x, \tilde{x}) = d^2_{ij}(\mu^*_{ij}(x, \tilde{x}), x, \tilde{x}) > 0$, which contradicts the initial hypothesis. Q.E.D.

A.6 Bernstein inequalities

In this section we specify the Bernstein inequalities, which we refer to in the proofs.

A.6.1 Bernstein inequality for bounded random variables

Theorem A.1 (Bernstein inequality; see, for example, Bennett (1962))**.** *Let $Z_1, \ldots, Z_n$ be mean zero independent random variables. Assume there exists a positive constant M such that $|Z_i| \leqslant M$ with probability one for each i. Also let $\sigma^2 := \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[Z_i^2\right]$. Then, for all $\epsilon > 0$, we have $\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| \geqslant \epsilon\right) \leqslant 2\exp\left(-\frac{n\epsilon^2}{2\left(\sigma^2 + \frac{1}{3}M\epsilon\right)}\right)$.*

A.6.2 Bernstein inequality for unbounded random variables

Lemma A.8 (Moments of a sub-Gaussian random variable)**.** *Let Z be a mean zero random variable satisfying $\mathbb{E}\left[e^{\lambda Z}\right] \leqslant e^{v\lambda^2}$ for all $\lambda \in \mathbb{R}$, for some $v > 0$. Then for every integer $q \geqslant 1$, we have $\mathbb{E}\left[Z^{2q}\right] \leqslant q!(4v)^q$.*

Proof of Lemma A.8. See Theorem 2.1 in Boucheron et al. (2013). Q.E.D.

Theorem A.2 (Bernstein inequality for unbounded random variables)**.** *Let $Z_1, \ldots, Z_n$ be independent random variables. Assume that there exist some positive constants ν and c such that $\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[Z_i^2\right] \leqslant \nu$ such that, for all integers $q \geqslant 3$, $\frac{1}{n}\sum \mathbb{E}\left[|Z_i|^q\right] \leqslant \frac{q!c^{q-2}}{2}\nu$. Then, for all $\epsilon > 0$, we have $\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n (Z_i - \mathbb{E}\left[Z_i\right])\right| \geqslant \epsilon\right) \leqslant 2\exp\left(-\frac{n\epsilon^2}{2(\nu + c\epsilon)}\right)$.*

Proof of Theorem A.2. See Corollary 2.11 in Boucheron et al. (2013). Q.E.D.

Specifically, we make use of the following corollary.

Corollary A.1. *Let $Z_1, \ldots, Z_n$ be mean zero independent random variables. Assume that there exists some $v > 0$ such that $\mathbb{E}\left[e^{\lambda Z_i}\right] \leqslant e^{v\lambda^2}$ for all $\lambda \in \mathbb{R}$ and for all $i \in \{1, \ldots, n\}$. Then, there exist some positive constants C, a, and b such that for all constants $\alpha_1, \ldots, \alpha_n$ satisfying $\max_i |\alpha_i| < \overline{\alpha}$ and for all $\epsilon > 0$,*

$$\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^{n}\alpha_i Z_i\right| \geqslant \epsilon\right) \leqslant C\exp\left(-\frac{n\epsilon^2}{a+b\epsilon}\right)$$

and

$$\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^{n}\alpha_i\left(Z_i^2 - \mathbb{E}\left[Z_i^2\right]\right)\right| \geqslant \epsilon\right) \leqslant C\exp\left(-\frac{n\epsilon^2}{a+b\epsilon}\right).$$

Proof of Corollary A.1. Follows from Lemma A.8 and Theorem A.2. Q.E.D.

Remark A.1. Note that the constants C, a, and b depend on v and $\overline{\alpha}$ only.

B Illustration of Assumption 6

Suppose ξ is scalar and $g(\xi_i, \xi_k) = \kappa\,|\xi_i - \xi_k|$. Note that, as a function of ξ_i, $g(\xi_i, \xi_k)$ is non-differentiable at $\xi_i = \xi_k$. When $\xi_i, \xi_j \leqslant \xi_k$, we have

$$g(\xi_i, \xi_k) - g(\xi_j, \xi_k) = \kappa(\xi_k - \xi_i) - \kappa(\xi_k - \xi_j) = -\kappa(\xi_i - \xi_j).$$

If $\xi_i, \xi_j \geqslant \xi_k$,

$$g(\xi_i, \xi_k) - g(\xi_j, \xi_k) = \kappa(\xi_i - \xi_k) - \kappa(\xi_j - \xi_k) = \kappa(\xi_i - \xi_j).$$

So, we can take

$$G(\xi_i, \xi_k) = \begin{cases} -\kappa, & \xi_i < \xi_k \\ \kappa, & \xi_i \geqslant \xi_k \end{cases}.$$

Then, the remainder $r_g(\xi_i, \xi_j, \xi_k) = 0$ when $\xi_i, \xi_j \leqslant \xi_k$ or $\xi_i, \xi_j \geqslant \xi_k$. However, if, for example, $\xi_i \leqslant \xi_k \leqslant \xi_j$,

$$g(\xi_i, \xi_k) - g(\xi_j, \xi_k) = \kappa(2\xi_k - \xi_i - \xi_j).$$

Since $G(\xi_i, \xi_k) = -\kappa$,

$$g(\xi_i, \xi_k) - g(\xi_j, \xi_k) = -\kappa(\xi_i - \xi_j) + 2\kappa(\xi_k - \xi_j) = G(\xi_i, \xi_k)(\xi_i - \xi_j) + r_g(\xi_i, \xi_j, \xi_k),$$

so $r_g(\xi_i, \xi_j, \xi_k) = 2\kappa(\xi_k - \xi_j)$ when $\xi_i \leqslant \xi_k \leqslant \xi_j$. Clearly, in this case, the linearization remainder is no longer $O(|\xi_i - \xi_j|^2)$. Importantly, Assumption 6 allows for this possibility: the remainder r_g is bounded by $C\delta_n^2$ only when $|\xi_i - \xi_k| > \delta_n$ and $|\xi_j - \xi_i| \leqslant \delta_n$. Under these restrictions, ξ_k cannot lie between ξ_i and ξ_j implying that in this case $r_g(\xi_i, \xi_j, \xi_k) = 0$, so the bound on $r_g(\xi_i, \xi_j, \xi_k)$ imposed by Assumption 6 is trivially satisfied.

C Imputation of $\hat{Y}_{ij}^*$ with Missing Data

In this section, we discuss how one can adjust the previously constructed estimator $\hat{Y}_{ij}^*$ when some interaction outcomes are missing. We will stick with the notation previously introduced in Section 5.2.

We start with constructing $\hat{d}_\infty^2(i, j)$ previously defined in (3.5). In the studied setting, $\hat{d}_\infty^2(i, j)$ can be computed as

$$\hat{d}_\infty^2(i, j) \coloneqq \max_{k \neq i, j} \left| |\mathcal{O}_{ijk}|^{-1} \sum_{\ell \in \mathcal{O}_{ijk}} (Y_{i\ell} - Y_{j\ell}) Y_{k\ell} \right|, \tag{C.1}$$

where $\mathcal{O}_{ijk} = \{\ell : D_{i\ell} = D_{j\ell} = D_{k\ell} = 1\}$. In practice, if $|\mathcal{O}_{ijk}|$ is too small for a given k, this k can be dropped from the calculation of $\hat{d}_\infty^2(i, j)$ above; it is sufficient to compute the maximum in (C.1) over a subset of all the other agents so long as it is not too small (otherwise, we can set $\hat{d}_\infty^2(i, j) = \infty$).

Next, we proceed with calculation of $\hat{Y}_{ij}^*$. To this end, we need to introduce a j-specific neighborhood of i denoted by $\hat{\mathcal{N}}_{ij}(n_{ij})$. As before, for simplicity, we will stick with discrete X. First, let us define a pool of candidate agents (donors) which can potentially be used to impute $\hat{Y}_{ij}^*$ denoted by $\mathcal{P}_{ij} \coloneqq \{i' : X_{i'} = X_i, D_{i'j} = 1\}$, where the requirement $D_{i'j} = 1$ ensures that $Y_{i'j}$ is observed. Then $\hat{\mathcal{N}}_{ij}(n_{ij})$ is constructed as a collection of n_{ij} agents from

$\mathcal{P}_{ij}$ closest to agent i in terms of $\hat{d}^2_\infty(i,i')$, i.e.,

$$\hat{\mathcal{N}}_{ij}(n_{ij}) \coloneqq \{i' \in \mathcal{P}_{ij} : \text{Rank}(\hat{d}^2_\infty(i,i')|\mathcal{P}_{ij}) \leqslant n_i\},$$

and $\hat{Y}^*_{ij}$ can be estimated as

$$\hat{Y}^*_{ij} = \frac{\sum_{i' \in \hat{\mathcal{N}}_{ij}(n_{ij})} Y_{i'j}}{n_{ij}}.$$

When the observed matrix Y is so sparse and $\mathcal{P}_{ij}$ is so limited in the first place to find enough decent matches for imputation of Y^*_{ij}, one can try to sequentially impute the elements of Y^* with a sufficient number of candidate donors in $\mathcal{P}$ first, and then use the newly imputed $\hat{Y}^*_{i'j}$'s as observed outcomes $Y_{i'j}$ to expand the previously thin pool of potential donors $\mathcal{P}_{ij}$. One can apply this procedure sequentially if needed to impute the whole Y^* or a part of it.

Once $\hat{Y}^*$ is constructed, we can estimate $\hat{d}^2_{ij}$ as in (3.4), or if some elements of $\hat{Y}^*$ are still missing, we can modify (3.4) in the same fashion as in (5.8), where the overlaps should be defined based on whether the imputed $\hat{Y}^*$'s are missing or not. Once $\hat{d}^2_{ij}$'s are obtained, we can calculate $\hat{\beta}$ as in (5.9), or even use the imputed outcomes instead of the plain outcomes in (5.9), including in the definition of the overlap $\mathcal{O}_{ij}$ again.

References

Ahn, H. and Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1-2):3–29.

Arcones, M. A. (1995). A bernstein-type inequality for u-statistics and u-processes. *Statistics & Probability Letters*, 22(3):239–247.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57(297):33–45.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.