Identification of a rank-dependent peer effect model

Eyo I. Herstad* Myungkou Shin†

October 17th, 2024

Abstract

This paper develops an econometric model to analyse heterogeneity in peer effects in network data with endogenous spillover across units. We introduce a rank-dependent peer effect model that captures how the relative ranking of a peer outcome shapes the influence units have on one another, by modeling the peer effect to be linear in ordered peer outcomes. In contrast to the traditional linear-in-means model, our approach allows for greater flexibility in peer effect by accounting for the distribution of peer outcomes as well as the size of peer groups. Under a minimal condition, the rank-dependent peer effect model admits a unique equilibrium and is therefore tractable. Our simulations show that that estimation performs well in finite samples given sufficient covariate strength. We then apply our model to educational data from Norway, where we see that higher-performing students disproportionately drive GPA spillovers.

*School of Economics, Erasmus University Rotterdam and Tindbergen Institute. email: herstad@ese.eur.nl

†School of Social Sciences, University of Surrey. email: m.shin@surrey.ac.uk

1 Introduction

This paper introduces an econometric model designed to investigate heterogeneous peer effects within network data. Network datasets capture the binary links between units, such as friendships among students or business transactions between firms. These connections raise important questions about how linked units influence one another through spillovers, known as peer effects. The goal of this paper is to develop an econometric framework that allows for flexible patterns of peer effects, with which empirical researchers can uncover variations in the structure of peer effects across different settings. Specifically, we will focus on variation based on peers outcomes relative to each other: a rank-dependent peer effect model.

In different empirical contexts, we anticipate varying patterns of peer effects. For instance, when students in a school interact in ways that allow them to receive help from higher-performing peers, the outcome of your highest-performing friend may determine the spillover effect. On the other hand, in a manufacturing setting where workers operate on a conveyor-belt-style assembly line, the performance of the slowest worker might be the key factor determining the spillover effect. The rank-dependent peer effect model allows us to identify these rich patterns of peer effects.

The current literature of peer effects primarily focus on estimating a scalar spillover parameter. Following the seminal works of Manski (1993); Bramoullé, Djebbari, and Fortin (2009), the canonical linear-in-means (LIM) peer effect model is as follows: for $i = 1, \cdots, n$,

$$Y_i = \beta \bar{Y}_i + {x_i}^\intercal \gamma + \varepsilon_i. \tag{1}$$

In the example of the spillover in the exam score, Y_i is the exam score for student i, $\bar{Y}_i$

is the mean of the student i's peers' exam scores and x_i is the vector of student-level control covariates for student i. x_i may include variables for contextual effects: e.g., the mean of the peers' socioeconomic variables for student i. In the LIM model, the peer effect is modeled to be a product of $\bar{Y}_i$ and a scalar peer effect parameter β.

However, the LIM model has several caveats due to assuming linearity and uniformity, limiting the patterns of spillover it can analyse. Firstly, the LIM model assumes that inputs from different peers are perfectly substitutable; e.g., there is no compositional effect from having homogeneous peers compared to having heterogeneous peers with equal mean. Moreover, the LIM model imposes that the number of peers should not matter. This is in contrast with the recent findings in the empirical economics literature that the centrality measure such as number of peers is an important determinant in one's outcome.

In light of these limitations, we develop an econometric model where the inputs from peers affect an individual's outcome differently, depending on the peer's position in the individual's peer group: for $i = 1, \ldots, n$,

$$Y_i = \sum_{k=1}^{d_i} \tilde{Y}_{i,k}\beta_{k,d_i} + {x_i}^\intercal\gamma + \varepsilon_i. \tag{2}$$

d_i is the number of peers for individual i and $\tilde{Y}_{i,k}$ is the outcome of the k-th lowest performing peer of individual i; $\tilde{Y}_{i,k}$ takes the k-th lowest value from the set

$$\{Y_j : \text{individual } j \text{ is a peer of individual } i.\}.$$

Note that the model (2) does not impose linear separability on inputs from peers. For example, suppose that individual 1 and 2 are peers. The peer effect coefficient on individual 1's outcome for individual 2 depends on the outcomes of other peers in individual 2's peer group, through the ranking of individual 1 in the group. In this sense, the peer effect coefficients are rank-dependent. We call the model (2) a rank-dependent peer effect model.

The rank-dependent peer effect model has several desirable properties. Firstly, the model does not impose perfect substitution; with the rank-dependent peer effect model, we can discuss whether the interaction among peers in terms of the peer effect show a pattern closer to complementarity or substitutability. Secondly, the size of a peer group matters in the model; we can see if an individual more connected to other individuals are systemically different from individuals with less connections. These observations are a part of the broader point that the peer effect construction is much more flexible in the rank-dependent peer effect model than in the LIM model. For example, an individual can only be affected by their lowest-performing or median or highest-performing peer:

$$Y_i = \beta_{\min}\tilde{Y}_{i,1} + \beta_{\text{median}}\tilde{Y}_{i,\lceil d_i/2\rceil} + \beta_{\max}\tilde{Y}_{i,d_i} + {x_i}^{\intercal}\gamma + \varepsilon_i. \tag{3}$$

Alternatively, the mean of peer outcomes can be the sole determining factor in the peer effect as in the LIM model, while the coefficient depends on the size of the peer group:

$$Y_i = \beta_{d_i}\bar{Y}_i + {x_i}^{\intercal}\gamma + \varepsilon_i.$$

These are merely two examples of peer effect heterogeneity patterns that the rank-dependent peer effect model can accommodate.

Moreover, the rank-dependent peer effect model can be motivated as a general regression of an individual's outcome on the distribution of their peers' outcomes. As will be further discussed in Section 2, ordered peer outcomes have one-to-one relationship with the (empirical) distribution of peer outcomes. As such, our model allows for a wide variety of heterogeneous patterns in the peer effect in relation to the distribution. This approach relies on the key assumption of exchangeability amongst peers in the peer effect model, meaning that the spillover an individual experiences does not depend on the identity of their peers.

The primary contribution of this paper is the identification of the rank-dependent peer effect model in (2). Following the literature on the peer effect in a network, our identification is in terms of the conditional distribution of the outcomes given the network connections and the individual characteristics; the network is assumed to be exogenous. The steps of the identification argument are as follows. Firstly, we show that the rank-dependent peer effect model admits a unique equilibrium when the magnitude of the peer effect is bounded by one. Then, using the unique equilibrium result, we derive a reduced-form representation of the model (2). Given the reduced-form representation, we construct moment conditions to identify the peer effect coefficients $\{\beta_{k,d}\}_{k,d}$. The instrument relevance condition assumed for the identification can easily be related to that in the LIM model.

In implementation, we use the Two-stage Least-squares (TSLS) estimation to estimate the parameters of the rank-dependent peer effect model. Our simulation exercises with a simple DGP show how the finite-sample performance of the TSLS estimator fares as the the rank-dependent peer effect model becomes more complex. To show the applicability of our model in an empirical context, we apply the TSLS estimator to estimate spillovers in learning between students at two Norwegian High-Schools. Our estimates show that higher-GPA peers are the driving factor in GPA spillover, supporting the empirical relevance of the rank-dependent peer effect model compared to the canonical models in the literature.

This paper contributes to the literature on peer effects in a network by developing a model where the peer effect is a flexible function of peers' outcomes in a way that how an individual affects their peer depends on their rank in the peer group. The most comparable to ours in the literature is Boucher, Rendall, Ushchev, and Zenou (2024). They model that the peer outcomes be aggregated to a scalar summary measure for the construction of the peer effect, while using a class of aggregators that nests the mean used in the LIM model. Thus, unlike our model, they do not discuss returns to centrality. In two extreme cases of their model, Boucher, Rendall, Ushchev, and Zenou (2024) construct the peer effect to be linear in the minimum peer outcome or the maximum peer outcome alone, both of which are allowed in our model as well. Furthermore, some empirical papers use specifications close in spirit to our model (Foster, 2006; Han and Li, 2009; Guryan, Kroft, and Notowidigdo, 2009), though to our best knowledge, there is yet to be a theoretical discussion of these models.

The rest of this paper is structured as follows. Section 2 discusses the model and its implications. Section 3 formally identifies the parameters of the rank-dependent peer effect model. Section 4 discusses finite-sample performances of a TSLS estimator in the rank-dependent peer effect model, using simulations. Finally, Section 5 applies our methodology to educational peer effects in Norwegian schools.

2 Model

In our setting, a researcher observes individual-level outcomes and individual-level control covariates $\{Y_i, x_i\}_{i=1}^n$ and the network adjacency matrix $\{A_{i,j}\}_{1\leq i,j\leq n}$. We let $A_{i,j} = 1$ indicate that individual i is linked to individual j, and $A_{i,j} = 0$ otherwise. We assume there are no self-loops in the network, so $A_{i,i} = 0$, and links may be directed, i.e., $A_{i,j}$ and $A_{j,i}$ are allowed to be different.

The rank-dependent peer effect model is as follows: for $i = 1, \cdots, n$,

$$Y_i = \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i} + x_i{}^\intercal \gamma + \varepsilon_i \tag{4}$$

where

$$\begin{aligned} d_i &= \sum_{j=1}^{n} A_{i,j} = \text{number of peers for individual } i \\ \tilde{Y}_{i,k} &= \text{outcome of the } k\text{-th lowest performing peer of individual } i \\ &= k\text{-th ordered statistic from } \{Y_j : A_{i,j} = 1\}. \end{aligned}$$

Note that $\tilde{Y}_{i,1} \leq \cdots \leq \tilde{Y}_{i,d_i}$ for any i and $\tilde{Y}_{i,k} = Y_j$ for some $j \neq i$. The individual-level control covariates x_i may include functions of the network adjacency matrix $\{A_{i,j}\}_{1 \leq i,j \leq n}$ and the peer characteristics. For example, given some individual-specific characteristics $\{w_i\}_{i=1}^n$, the control covariate x_i could be functions of $\{w_i\}_{i=1}^n$ and $\{A_{i,j}\}_{1 \leq i,j \leq n}$, allowing for contextual effects:

$$Y_i = \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i} + \begin{pmatrix} w_i & \frac{1}{d_i} \sum_{j \neq i} A_{i,j} w_j \end{pmatrix}^\intercal \gamma + \varepsilon_i =: \sum_{k=1}^{d_i} \tilde{Y}_{i,k} \beta_{k,d_i} + x_i \gamma + \varepsilon_i.$$

The rank-dependent peer effect model (4) assumes that the outcomes of individual i's peers affect individual i in a way that the coefficient on a peer outcome depends on the peer's rank among individual i's peer group. $\beta_{k,d}$ is the coefficient for the k-th lowest performing peer's outcome, when the given individual has d peers in total. Note that the model allows a individual to affect different peers differently. Suppose that individual 1 is friends with both individual 2 and 3. If individual 1 is the lowest-performing peer of individual 2 while being the highest-performing peer of individual 3, the effect of individual 1's outcome will be β_{1,d_2} for individual 2 and β_{d_3,d_3} for individual 3.

The model nests both the linear-in-means (LIM) model and the linear-in-sums (LIS) model. By letting $\beta_{k,d} = \beta/d$, the peer effect term in (4) becomes the LIM peer effect:

$$\sum_{k=1}^{d_i} \tilde{Y}_{i,k}\beta_{k,d_i} = \frac{1}{d_i}\sum_{k=1}^{d_i} \tilde{Y}_{i,k}\beta = \beta\bar{Y}_i \tag{5}$$

with $\bar{Y}_i = \frac{1}{d_i}\sum_{k=1}^{d_i} \tilde{Y}_{i,k} = \frac{\sum_{j=1}^{n} A_{i,j}Y_j}{\sum_{j=1}^{n} A_{i,j}}$. By letting $\beta_{k,d} = \beta$, the peer effect term becomes the LIS peer effect:

$$\sum_{k=1}^{d_i} \tilde{Y}_{i,k}\beta_{k,d_i} = \sum_{k=1}^{d_i} \tilde{Y}_{i,k}\beta = \beta\sum_{j=1}^{n} A_{i,j}Y_j. \tag{6}$$

Though widely used, the LIM model and the LIS model have a caveat that using only a mean or a sum may ignore important information contained in the second or higher moments of the peer outcomes.[1] Thus, researchers often choose to include additional parametric terms in the peer effect. A standard deviation of peer outcomes is a popular choice. Other examples include minimum/maximum peer outcome, as in Foster (2006); Han and Li (2009); Guryan, Kroft, and Notowidigdo (2009), and proportion of peers whose outcomes are below/above a cutoff, as in Kang (2007). Our model shares the same spirit and allows for more generalized patterns of peer effect.

In fact, the peer effect construction in our model can be understood as a linear special case of a nonlinear peer effect where the peer outcome inputs are exchangeable. Consider a model with nonlinear peer effect:

$$Y_i = h\big(\{A_{i,j}Y_j\}_{j=1}^{n}\big) + x_i{}^{\intercal}\gamma + \varepsilon_i.$$

[1]A notable exception is where the peer effect occurs through a binary variable, as in (Gazze, Persico, and Spirovska, 2024)

We call $h(\cdot)$ the peer effect function, or simply the peer effect, which takes a $n \times 1$ vector and produces a scalar peer effect. When the inputs of $h(\cdot)$ are exchangeable, i.e.

$$h\big(\{A_{i,j}Y_j\}_{j=1}^n\big) = h\big(\{A_{i,\pi(j)}Y_{\pi(j)}\}_{j=1}^n\big)$$

for any permutation π on $\{1, \cdots, n\}\setminus\{i\}$, we can write $h\big(\{A_{i,j}Y_j\}_{j=1}^n\big) = h\big(\{\tilde{Y}_{i,k}\}_{k=1}^{d_i}\big)$ with some abuse of notation. Note that the vector of ordered peer outcomes $\{\tilde{Y}_{i,k}\}_{k=1}^{d_i}$ is bijective with the number of peers d_i and the empirical distribution function of peer outcomes:

$$\mathbf{F}_i(y) = \frac{1}{d_i}\sum_{j=1}^{n} A_{i,j}\mathbf{1}\{Y_j \leq y\}.$$

In this sense, the construction of the peer effect $h(\cdot)$ can be thought of as a function that takes a distribution of peer outcomes as an input. This characterization covers all of the parametrizations discussed in the previous paragraph and many more: standard deviation, minimum, maximum, etc. The model (4) additionally assumes that $h(\cdot)$ is linear in $\tilde{Y}_{i,k}$.[2] In this sense, a key intuition for the rank-dependent peer effect model is that the names of individuals do not matter and therefore the peer effect is a function of the peer outcome distribution.

The flexibility of our peer effect model (4) allows us to address key questions in the peer effect literature. First, because the coefficients are rank-dependent, our model enables the analysis of complementarity versus substitutability, as has been discussed in Kang (2007) amongst others. For instance, when $\beta_{1,d} \neq 0$ and $\beta_{k,d} = 0$ for every $k \geq 2$ and d, the peer effect reflects perfect complementarity, meaning that only the lowest-performing peer influences the outcome. On the other hand, when $\beta_{k,d} = \beta_{k',d}$ for all k, k', the peer effect exhibits global perfect substitutability, where each peer

[2]Note that the linearity is not on the $\{A_{i,j}Y_j\}_{j=1}^n$ but on $\{\tilde{Y}_{i,k}\}_{k=1}^{d_i}$; the peer outcomes are not linearly separable when unordered.

contributes equally. It is important to note that the linear structure of our model imposes local perfect substitutability. Specifically, all peer outcomes are perfectly substitutable with a constant marginal rate of substitution for small changes in $\tilde{Y}_{i,k}$ such that $\tilde{Y}_{i,k-1} \leq \tilde{Y}_{i,k} + \Delta \leq \tilde{Y}_{i,k+1}$. However, when the change in $\tilde{Y}_{i,k}$ is large enough to alter its rank, the slope of the isoquant curve adjusts accordingly.

Secondly, because the coefficients also depend on d_i, the size of the peer group, our peer effect model enables the exploration of returns to centrality. For simplicity, assume $Y_j = y$ for every j in the peer group of individual i. In the LIM model, there is zero return to centrality, meaning that the peer effect remains unchanged when an additional peer is added. The LIS model imposes a constant return to centrality. In contrast, our model we can document the return to centrality in a flexible manner, by plotting $d \mapsto \sum_{k=1}^{d} \beta_{k,d} y$. These two features of peer effect heterogeneity are directly related to several empirically important questions: does a student benefit from a high-performing peer or get negatively influenced by a low-performing one? how substitutable are peer outcomes? does the number of friends impact the spillover effect?

While the rank-dependent peer effect model (4) allows for nonlinearity in the peer effect, allowing for the peer effect of one peer to depend on the outcomes of all peers, it still admits a unique equilibrium under a natural extension of the assumption used in LIM models (Bramoullé, Djebbari, and Fortin, 2009).

Assumption 1. (Bounded Peer Effect) *There exists some $\bar{d}$ such that* $\max_i d_i \leq \bar{d}$. *For each $d = 1, \cdots, \bar{d}$,*

$$\sum_{k=1}^{d} |\beta_{k,d}| < 1.$$

Assumption 1 assumes that the sum of the absolute values of peer effect coefficients for each individual is less than 1. Proposition 1 states our unique equilibrium result.

Proposition 1. *Assumption 1 holds. Then, for any realization of* $\{x_i, A_{i,j}, \epsilon_i\}_{1\leq i,j\leq n}$*, the peer effect model* (4) *admits a unique equilibrium.*

Proof. See Appendix B.1 □

Since the peer effect model admits a unique equilibrium, there exists a well-defined function from $\{\varepsilon_i\}_{i=1}^n$ to $\{Y_i\}_{i=1}^n$ given $\{x_i, A_{i,j}\}_{1\leq i,j\leq n}$.

The unique equilibrium result from Proposition 1 allows us to introduce a key ingredient in the identification of the peer effect parameters: the conditional distribution of $\{Y_i\}_{i=1}^n$ given $\{x_i, A_{i,j}\}_{1\leq i,j\leq n}$. Without a unique equilibrium result as in Proposition 1, we would have to assume an (often arbitrary) equilibrium selection mechanism, if we were to derive a conditional distribution of $\{Y_i\}_{i=1}^n$ given $\{x_i, A_{i,j}\}_{1\leq i,j\leq n}$ from a conditional distribution of $\{\varepsilon_i\}_{i=1}^n$ given $\{x_i, A_{i,j}\}_{1\leq i,j\leq n}$.[3] However, thanks to Proposition 1, we can assume Assumption 1 and additional conditions on the distribution of $\{\varepsilon_i\}_{i=1}^n$, to construct moment conditions with $\{Y_i\}_{i=1}^n$.

3 Identification

To simplify notation, we use $\mathbf{E}_n$ to denote the conditional expectation of a random variable given the exogenous variables throughout the paper. That is for any random variable W,

$$\mathbf{E}_n\left[W\right] := \mathbf{E}\left[W|\{x_i, A_{i,j}\}_{1\leq i,j\leq n}\right].$$

[3]Alternatively, identification arguments that do not rely on knowing the entire distribution of the sample have been proposed in the literature of games with multiple equilibria: see De Paula (2013) for more. For example, we may limit our attention to sets of $\{y_i\}_{i=1}^n$ that can only be an outcome of a single realization of $\{\epsilon_i\}_{i=1}^n$, thus adding an additional conditioning variable to the conditional distribution we discuss identification with. Or, we may focus on extreme values of $\{x_i\}_{i=1}^n$ and discuss identification for samples with extreme values only. Though each of these approaches have unique appeals, it is hard to implement them in our setup since multiple equilibria in the rank-dependent peer effect model mean that we need to consider $n!$ different possible orderings to find values of $\{y_i\}_{i=1}^n$ consistent with $\{\epsilon_i\}_{i=1}^n$ and thus the correspondence that maps $\{x_i, A_{i,j}, \epsilon_i\}_{1\leq i,j\leq n}$ to the set of $\{y_i\}_{i=1}^n$ is not straightforward to pin down.

We will follow the literature by assuming that the error term ε_i is exogenous to all individual characteristics and the network.

Assumption 2. (Exogenous Network and Covariates) *For each* $i = 1, \ldots, n$

$$\mathbf{E}\left[\varepsilon_i | \{x_i, A_{i,j}\}_{1 \leq i,j \leq n}\right] = 0.$$

Assumption 2 rules out endogenous network formation and treats the network structure as fixed, meaning it does not depend on the randomness in $\{\varepsilon_i\}_{i=1}^n$. While there has been advances in analysing peer effect models with endogenous network formation (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Johnsson and Moon, 2021; Jochmans, 2023), we instead focus on endogeneity in peer effects conditional on the network.[4]

Assumption 2 allows us to construct a reduced-form representation of $\mathbf{E}_n[Y_i]$, though unlike the LIM model this reduced form is not linear in $\{x_i\}_{i=1}^n$. Instead, the rank-dependence creates a nonlinear reduced-form representation as shown in Corollary 1. To discuss the coefficients in the corollary, let us introduce the following notations: π denotes an ordering on $\{1, \ldots, n\}$ in terms of $\{y_i\}_{i=1}^n$ and $\mathbb{B}(\pi)$ denotes a $n \times n$ peer effect coefficient matrix such that its i-th row j-th column component corresponds to the peer effect coefficient for individual j's outcome on individual i's outcome, given the ordering π. For any realization of $\{\epsilon_i\}_{i=1}^n$, Proposition 1 gives us

[4]To see this, note that we can view our model as a second step of a two-stage process where firstly, the network is formed and secondly, the peer effect is determined based on the connections established in the first step. Models of endogenous network formation address endogeneity in the first step, whereas our model addresses endogeneity in the second step by allowing peer effect coefficients to depend on $\{\varepsilon_i\}_{i=1}^n$.

a unique equilibrium $\{y_i\}_{i=1}^n$ and thus a unique π such that

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \mathbb{B}(\pi) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} + \begin{pmatrix} x_1{}^\intercal\gamma + \epsilon_1 \\ \vdots \\ x_n{}^\intercal\gamma + \epsilon_n \end{pmatrix}.$$

Corollary 1 takes expectations over the set of every possible ordering, Π, and gives us a reduced-form representation of $\mathbf{E}_n[Y_i]$.

Corollary 1. *Assumptions 1 and 2 hold. Then, the peer effects model* (4) *admits a reduced-form relationship with residual terms η_i:*

$$\mathbf{E}_n\left[Y_i\right] = \sum_{j=1}^n \theta_{i,j} x_j{}^\intercal\gamma + \eta_i$$

where

$$\theta_{i,j} = \sum_{\pi\in\Pi} \theta_{i,j}(\pi)\mathrm{Pr}_n\left\{\pi\right\}$$

$$\eta_i = \sum_{\pi\in\Pi}\sum_{j=1}^n \theta_{i,j}(\pi)\mathbf{E}_n\left[\varepsilon_j|\pi\right]\cdot\mathrm{Pr}_n\left\{\pi\right\}$$

and $\Big(\theta_{i,1}(\pi),\cdots,\theta_{i,n}(\pi)\Big)$ *is the i-th row of the* $n\times n$ *matrix* $\left(I_n - \mathbb{B}(\pi)\right)^{-1}$*. Moreover, there also exists a reduced-form relationship between the ordered peer outcome* $\tilde{Y}_{i,k}$ *and* $\{x_j\}_{j=1}^n$*, with residual terms* $\{\tilde{\eta}_{j,k}\}_{j=1}^n$*:*

$$\mathbf{E}_n\left[\tilde{Y}_{i,k}\right] = \sum_{j=1}^n \tilde{\theta}_{i,k,j} x_j{}^\intercal\gamma + \tilde{\eta}_{i,k}$$

where

$$\tilde{\theta}_{i,k,j} = \sum_{\pi\in\Pi} \tilde{\theta}_{i,k,j}(\pi)\mathrm{Pr}_n\{\pi\}$$
$$\tilde{\eta}_{i,k} = \sum_{\pi\in\Pi}\sum_{j=1}^{n} \tilde{\theta}_{i,k,j}(\pi)\mathbf{E}_n[\varepsilon_j|\pi]\cdot\mathrm{Pr}_n\{\pi\}.$$

and $\left(\tilde{\theta}_{i,k,1}(\pi),\cdots,\tilde{\theta}_{i,k,n}(\pi)\right)$ *is one row of the* $n\times n$ *matrix* $(I_n - \mathbb{B}(\pi))^{-1}$*, which corresponds to the* k*-th lowest performing peer of individual* i*.*

Proof. See Appendix B.2 □

Note that the residual terms η_i and $\tilde{\eta}_{i,k}$ may not be linear in $\{x_i\}_{i=1}^n$.

To compare the reduced form from Corollary 1 with more familiar models, such as the LIM model in (1) and the LIS model, observe that for these models, Assumptions 1 and 2 directly yield a reduced-form representation of $\{Y_i\}_{i=1}^n$:

$$\begin{pmatrix}\mathbf{E}_n[Y_1]\\ \vdots \\ \mathbf{E}_n[Y_n]\end{pmatrix} = (I_n - \mathbb{B})^{-1}\begin{pmatrix}x_1{}^\intercal\gamma\\ \vdots \\ x_n{}^\intercal\gamma\end{pmatrix} \tag{7}$$

where the $n\times n$ matrix $\mathbb{B}$ is $\beta G := \beta\cdot\left(A_{i,j}/d_i\right)_{i,j}$ in the LIM model and $\beta A := \beta\cdot\left(A_{i,j}\right)_{i,j}$ in the LIS model. Importantly, this equation is linear in $\{x_i\}_{i=1}^n$.

However, this linearity does not hold in our model. Instead, the rank-dependent peer effect coefficients introduce nonlinearities in the conditional expectations due to the connection between ordering and $\{x_i\}_{i=1}^n$. Consider a simple case with three individuals where individual 1 is connected to both individuals 2 and 3, but individuals

2 and 3 are not connected to each other. From Assumption 2, we have:

$$\begin{aligned}\mathbf{E}_n\left[Y_1\right] = &\beta_{12}\mathbf{E}_n\left[Y_2\mathbf{1}\{Y_2 \leq Y_3\} + Y_3\mathbf{1}\{Y_2 > Y_3\}\right] \\ &+ \beta_{22}\mathbf{E}_n\left[Y_3\mathbf{1}\{Y_2 \leq Y_3\} + Y_2\mathbf{1}\{Y_2 > Y_3\}\right] + x_1{}^\intercal\gamma.\end{aligned}$$

Substituting Y_2 and Y_3 using their reduced forms, we get:

$$\begin{aligned}\tilde{\beta}\mathbf{E}_n\left[Y_1|\{x_i, A_{i,j}\}_{1\leq i,j\leq n}\right] = &\beta_{12}\mathbf{E}_n\left[\varepsilon_2\mathbf{1}\{Y_2 \leq Y_3\} + \varepsilon_3\mathbf{1}\{Y_2 > Y_3\}\right] \\ &+ \beta_{22}\mathbf{E}_n\left[\varepsilon_3\mathbf{1}\{Y_2 \leq Y_3\} + \varepsilon_2\mathbf{1}\{Y_2 > Y_3\}\right] + \sum_{i=1}^{3} x_i{}^\intercal\tilde{\gamma}_i\end{aligned}$$

for some $\tilde{\beta}$, $\tilde{\gamma}_1$, $\tilde{\gamma}_2$, and $\tilde{\gamma}_3$. Without further assumptions on the distribution of $\{\varepsilon_i\}_{i=1}^n$, the conditional expectations involving terms like $\mathbf{E}_n\left[\varepsilon_2\mathbf{1}\{Y_2 \leq Y_3\}\right]$, which may not be zero or linear in x_i. An exception occurs when $\beta_{12} = \beta_{22}$, in which case the conditional expectations cancel out, resulting in a form similar to the LIM or LIS model.

These complications affect the relevance conditions we use to construct valid moment conditions. In the standard LIM and LIS models, the relevance condition on a set of instruments $\{z_i\}_{i=1}^n$ for identifying β is given by ensuring that the matrix

$$\begin{pmatrix} z_1 & \dots & z_n \end{pmatrix} \left(\mathbb{B}\left(I_n - \mathbb{B}\right)^{-1} \begin{pmatrix} x_1{}^\intercal\gamma \\ \vdots \\ x_n{}^\intercal\gamma \end{pmatrix} \quad \begin{pmatrix} x_1{}^\intercal \\ \vdots \\ x_n{}^\intercal \end{pmatrix} \right) \tag{8}$$

has full rank. Bramoullé, Djebbari, and Fortin (2009) discusses a similar condition in the LIM model when instruments are based on the mean covariates of peers or peers of peers.

Our conditions are different and will instead relate to the reduced forms presented in Corollary 1. These conditions are given in Assumption 3.

Assumption 3.

a. (EXOGENEITY) $\left\{z_{i,1}, \cdots, z_{i,\bar{d}}\right\}_{i=1}^{n}$ *are known, predetermined functions of* $\{x_i, A_{i,j}\}_{1\leq i,j\leq n}$.

b. (RELEVANCE) *The construction of the instrument* z_i *and the reduced-form representation of*

$$\mathbf{E}_n\left[\tilde{Y}_{i,k}\right] = \sum_{j=1}^{n} \tilde{\theta}_{i,k,j} x_j{}^{\intercal}\gamma + \tilde{\eta}_{i,k}$$

from Corollary 1 satisfy that

$$\sum_{i=1}^{n} \begin{pmatrix} z_{i,d} \\ x_i \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n} \tilde{\theta}_{i,1,j} x_j{}^{\intercal}\gamma + \tilde{\eta}_{i,1} \\ \vdots \\ \sum_{j=1}^{n} \tilde{\theta}_{i,d_i,j} x_j{}^{\intercal}\gamma + \tilde{\eta}_{i,d_i} \\ x_i \end{pmatrix}^{\intercal} \mathbf{1}\{d_i = d\}$$

has full rank, for each $d = 1, \cdots, \bar{d}$.

Part **a.** of Assumption 3, together with Assumption 2, provides the standard instrument exogeneity assumption commonly found in the literature. This includes, for example, instruments based on the average covariates of peers and peers-of-peers, as introduced by Bramoullé, Djebbari, and Fortin (2009) for estimating the effects of peers' average outcomes. Given the increased number of endogenous variables in our model compared to the LIM model, a larger set of instruments is required. Beyond the instruments discussed in Bramoullé, Djebbari, and Fortin (2009), we also consider alternatives such as ordered peer covariates or higher moments of peer covariates. In the following discussion, we do not commit to a specific set of instruments, we focus on outlining the necessary conditions that any valid instruments must satisfy for identification.

Assumption 3-**b.** establishes a rank condition analogous to those found in Bramoullé, Djebbari, and Fortin (2009). The matrix in Assumption 3-**b.** decomposes into two components:

$$\sum_{i:d_i=d} \begin{pmatrix} z_{i,d} \\ x_i \end{pmatrix} \begin{pmatrix} \sum_{j=1}^n \tilde{\theta}_{i,1,j} x_j^\intercal \gamma \\ \vdots \\ \sum_{j=1}^n \tilde{\theta}_{i,d,j} x_j^\intercal \gamma \\ x_i \end{pmatrix}^\intercal \quad \text{and} \quad \sum_{i:d_i=d} \begin{pmatrix} z_{i,d} \\ x_i \end{pmatrix} \begin{pmatrix} \tilde{\eta}_{i,1} \\ \vdots \\ \tilde{\eta}_{i,d} \\ \mathbf{0}_l \end{pmatrix}^\intercal .$$

Constructing instruments $\{z_{i,d}\}_{i=1}^n$ such that the first matrix is full rank is straightforward as long as there is sufficient non-transitivity in the network and variation in the covariates $\{x_i\}_{i=1}^n$. Specifically, we need the columns of $\mathbb{X} = \left(x_1, \ldots, x_n\right)^\intercal$ and

$$\begin{pmatrix} \vdots \\ \sum_{j=1}^n \tilde{\theta}_{i,1,j} x_j^\intercal \gamma \\ \vdots \end{pmatrix} \quad \cdots \quad \begin{pmatrix} \vdots \\ \sum_{j=1}^n \tilde{\theta}_{i,d,j} x_j^\intercal \gamma \\ \vdots \end{pmatrix}$$

to be linearly independent. This condition is closely related to that found in Bramoullé, Djebbari, and Fortin (2009), where there are no residual terms $\{\tilde{\eta}_{i,k}\}_{i,k}$. As such, when $\{\tilde{\eta}_{i,k}\}_{i,k}$ are close to zero, the sum of the two matrices will also have full rank, satisfying our relevance condition. In cases where $\{\tilde{\eta}_{i,k}\}_{i,k}$ are nonzero, it is challenging to interpret the rank condition, but we can empirically check whether or not it holds in a given sample.

To simplify the notation of our identification results, let

$$\beta = \begin{pmatrix} \beta_{11} & \beta_{21} & \beta_{22} & \cdots & \beta_{1\bar{d}} & \cdots & \beta_{\bar{d}\bar{d}} \end{pmatrix}^\intercal$$
$$\tilde{Y}_i = \begin{pmatrix} 0 & \cdots & 0 & \tilde{Y}_{i,1} & \cdots & \tilde{Y}_{i,d_i} & 0 & \cdots 0 \end{pmatrix}^\intercal$$

so that the rank-dependent peer effect model (4) can be written as follows:

$$Y_i = \tilde{Y}_i^\intercal \beta + x_i{}^\intercal \gamma + \varepsilon_i = W_i{}^\intercal \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \varepsilon_i. \tag{9}$$

Let $W_i = \left(\tilde{Y}_i^\intercal, x_i{}^\intercal\right)^\intercal \in \mathbb{R}^{\frac{\bar{d}(\bar{d}+1)}{2}+l}$ and $z_i = \left(z_{i,1}{}^\intercal, \cdots, z_{i,\bar{d}}{}^\intercal, x_i{}^\intercal\right)^\intercal$. Lastly, let $\mathbb{Y}$, $\mathbb{W}$ and $\mathbb{Z}$ denote the row-stacked matrices of Y_i, W_i and z_i. Theorem 1 provides the identification result.

Theorem 1. *Suppose that Assumptions 1-3 hold and* $n \geq \frac{\bar{d}(\bar{d}+1)}{2} + l$. *Then,* β *and* γ *are identified from the moment condition below:*

$$\mathbf{E}_n \left[\mathbb{Z}^\intercal \mathbb{W}\right] \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \mathbf{E}_n \left[\mathbb{Z}^\intercal \mathbb{Y}\right]. \tag{10}$$

Proof. See Appendix B.3. □

For identification, we consider a fixed n and $\bar{d}$ environment. The minimal condition $n \geq \frac{\bar{d}(\bar{d}+1)}{2} + l$, along with Assumption 3, guarantees that the matrix $\mathbf{E}_n \left[\mathbb{Z}^\intercal \mathbb{W}\right]$ has a left inverse. However, in practice, when $\bar{d}$ is large relative to n, this can pose challenges. In such cases, it may be beneficial to consider an asymptotic framework where $\bar{d}$ grows with n and thus the dimension of β also grows with n. If the parameter β satisfies some sparsity or smoothness restrictions, it may be possible to take advantage of recent developments for high-dimensional models with many endogenous variables. See for example Belloni, Hansen, and Newey (2022). However, this literature primarily analyses cross-sectional data, and it is unclear if these results will extend to our setting due to the dependent nature of network data.

The moment condition (10) consists of n different equations, one for each student $i = 1, \cdots, n$. Since $\tilde{Y}_i$ concatenates $\bar{d}$ different vectors of peer outcomes where only one out of the $\bar{d}$ vectors is nonzero for a given student, the $n \times \left(\frac{\bar{d}(\bar{d}+1)}{2} + l\right)$ matrix $\mathbf{E}_n[\mathbb{Z}^\intercal \mathbb{W}]$ can also be decomposed into $\bar{d}$ submatrices. Part **b.** of Assumption 3 gives us each of the $\bar{d}$ submatrices having full rank, leading to $\mathbf{E}_n[\mathbb{Z}^\intercal \mathbb{W}]$ having full rank as a result. Given that $\mathbf{E}_n\left[\mathbb{Z}^\intercal \mathbb{W}\right]$ has full rank, it is trivial to write β and γ as functions of the conditional moments of $\{Y_i\}_{i=1}^n$. In the case of the TSLS estimand, the first stage coefficient matrix from regressing W_i on z_i is used. Let Γ denote the first stage coefficient matrix. Then,

$$\begin{pmatrix} \beta^{TSLS} \\ \gamma^{TSLS} \end{pmatrix} := \left(\mathbf{E}_n\left[(\mathbb{Z}\Gamma)^\intercal \mathbb{W}\right]\right)^{-1} \mathbf{E}_n\left[(\mathbb{Z}\Gamma)^\intercal \mathbb{Y}\right] = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}.$$

For Sections 4 and 5, we use the sample analogue of the TSLS estimand as our estimator.

3.1 Comparison to other peer effect estimands

A natural question, given that our model generalizes the standard LIM model, is to what extent existing estimands recover the key parameters of interest from our model. A commonly accepted minimal standard for such estimands is that they represent weighted averages of the underlying heterogeneity, such as in the analysis of instrument variables (Mogstad and Torgovitsky, 2024) or the analysis of Difference-in-

Difference estimands (De Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021). To explore this, we first assume instruments that satisfy the relevance condition used in the LIM model and define a LIM estimand as follows:

$$\beta^{\text{LIM}} := \left(\mathbf{E}_n\left[\mathbb{Z}^\intercal \mathbb{M}_X G \mathbb{Y}\right]\right)^{-1} \mathbf{E}_n\left[\mathbb{Z}^\intercal \mathbb{M}_X \mathbb{Y}\right]$$

where $G = \left(A_{i,j}/d_i\right)_{i,j}$ and $\mathbb{M}_X = I_n - \mathbb{X}\left(\mathbb{X}^\intercal\mathbb{X}\right)^{-1}\mathbb{X}^\intercal$ with $\mathbb{X}$, a row-stacked matrix constructed with $\{x_i\}_{i=1}^n$. Similarly, for the LIS model, we define

$$\beta^{\text{LIS}} := \left(\mathbf{E}_n\left[\mathbb{Z}^\intercal \mathbb{M}_X A \mathbb{Y}\right]\right)^{-1} \mathbf{E}_n\left[\mathbb{Z}^\intercal \mathbb{M}_X \mathbb{Y}\right]$$

where $A = \left(A_{i,j}\right)_{i,j}$.

Assumption 4.

i *There exist an instrument $\{z_i\}_{i=1}^n$, a known, predetermined function of $\{x_i, A_{i,j}\}_{1\leq i,j\leq n}$, that satisfies $\mathbf{E}_n\left[\mathbb{Z}^\intercal \mathbb{M}_X G \mathbb{Y}\right] \neq 0$.*

ii *There exist an instrument $\{z_i\}_{i=1}^n$, a known, predetermined function of $\{x_i, A_{i,j}\}_{1\leq i,j\leq n}$, that satisfies $\mathbf{E}_n\left[\mathbb{Z}^\intercal \mathbb{M}_X A \mathbb{Y}\right] \neq 0$.*

These conditions are the same instrument relevance conditions used in the LIM model and the LIS model, as in Bramoullé, Djebbari, and Fortin (2009).

Proposition 2. *Under Assumption 1, 2 and 4-i,*

$$\beta^{\text{LIM}} = \sum_{d=1}^{\bar{d}} \sum_{k=1}^{d} w_{k,d}^{\text{LIM}} d\beta_{k,d}$$

where $\sum_{d=1}^{\bar{d}} \sum_{k=1}^{d} w_{k,d}^{\text{LIM}} = 1$ and $w_{k,d}^{\text{LIM}} \lesseqgtr 0$. Similarly, under Assumptions 1, 2 and

4-ii,

$$\beta^{\text{LIS}} = \sum_{d=1}^{\bar{d}} \sum_{k=1}^{d} w_{k,d}^{\text{LIS}} \beta_{k,d}$$

where $\sum_{d=1}^{\bar{d}} \sum_{k=1}^{d} w_{k,d}^{\text{LIS}} = 1$ *and* $w_{k,d}^{\text{LIS}} \lesseqgtr 0$.

Proof. See Appendix B.4. □

Note that, in the case of LIM misspecification, the weights are applied not directly to $\beta_{k,d}$ but to $d\beta_{k,d}$. The rank-dependent peer effect coefficient $\beta_{k,d}$ represents the impact of a single peer outcome when there are d peers in total, so it should be rescaled by d to interpret it as a coefficient on a 'representative' or'average' peer. For instance, when $\beta_{1,d} = \ldots = \beta_{d,d} = \beta$, we have:

$$\sum_{k=1}^{d} \beta_{k,d} \tilde{y}_{i,k} = d\beta \bar{y}_i,$$

where the coefficient on the average peer outcome is $d\beta_{k,d}$.

Proposition 2 demonstrates that both the LIM estimand β^{LIM} and the LIS estimand β^{LIS} are weighted sums of the rank-dependent peer effect coefficients $\{\beta_{k,d}\}_{k,d}$. While the weights sum to one, which is reasonable, there is no guarantee that the weights have the same sign. As a result, it is possible for all $\beta_{k,d}$ to be positive, while the LIM estimand β^{LIM} is negative, and vice versa. The expressions for the weights are provided in the Appendix.

4 Simulation

To investigate the finite sample performance of the TSLS estimator for the rank-dependent peer effects, we simulated 1000 samples from the rank-dependent peer effect model, while varying the model specifications. To fully investigate the small sample properties of the estimator, we consider sample sizes of either $n = 100$ or $n = 200$. The DGP is as follows: for $i = 1, \ldots, n$,

$$Y_i = \gamma_0 + \gamma_1 X_i + \tilde{Y}_i^\intercal \beta + \varepsilon_i$$

and

$$\begin{pmatrix} X_i \\ \varepsilon_i \end{pmatrix} \overset{\text{iid}}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

To simulate the network adjacency matrix A, we let

$$A_{i,j} = \mathbf{1}\{V_{i,j} \geq \tau\}.$$

with $V_{i,j} \overset{\text{iid}}{\sim} \text{unif}[0,1]$. $\{V_{i,j}\}_{i \neq j}$ are drawn independently of $\{X_i, \varepsilon_i\}_{i=1}^n$ and the threshold τ is decided after drawing $\{V_{i,j}\}_{i \neq j}$ to ensure that the maximum number of peers is capped at a predetermined $\bar{d}$. In cases where the network has less than $\bar{d} + 2$ students for each level of $d_i = 0, \cdots, \bar{d}$ we redraw the network.

The instruments $\left\{z_i = \left(z_{i,1}, \ldots, z_{i,\bar{d}}{}^\intercal\right)^\intercal\right\}_{i=1}^n$ used in the TSLS estimation are constructed by taking ordered statistics of peers' covariate $\{X_j\}_{j:A_{i,j}=1}$ for each student i. $z_{i,d}$ is a d-dimensional vector that orders $\{X_j\}_{j:A_{i,j}=1}$ when $d_i = d$ and contains zeros when $d_i \neq d$. To avoid the contamination bias discussed in Goldsmith-Pinkham, Hull, and Kolesár (2024+), we stratify our estimator. We do this by splitting the dataset into $\bar{d}+1$ strata of $\{i : d_i = 0\}, \ldots, \{i : d_i = \bar{d}\}$ and the sets of the coefficients $\left(\gamma_0, \gamma_1\right), \beta_{1,1}, \ldots, \left(\beta_{1,\bar{d}}, \ldots, \beta_{\bar{d},\bar{d}}\right)$ are separately estimated from each strata.

First, we set $\bar{d} = 2$. The bias and MSE of the TSLS estimator are compared to those of the OLS estimator in Table I. In this table, two key model parameters,

Table I
Comparison between the OLS estimator and the TSLS estimator

	bias				MSE			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\gamma}_1^{OLS}$	-0.006	-0.010	-0.006	-0.008	0.011	0.011	0.005	0.006
$\hat{\beta}_{1,2}^{OLS}$	0.118	0.051	0.138	0.050	0.144	0.047	0.125	0.036
$\hat{\gamma}_1^{TSLS}$	0.000	0.001	0.000	-0.002	0.019	0.018	0.008	0.008
$\hat{\beta}_{1,2}^{TSLS}$	0.042	0.012	0.039	0.004	0.409	0.081	0.379	0.053
γ_1	1	2	1	2	1	2	1	2
n	100	100	200	200	100	100	200	200

Notes: The values of the coefficients used in DGP are as follows: $\gamma_0 = 1$ and $\beta = \big(0.3, 0.1, 0.2\big)$. The instruments used in TSLS estimation are ordered peers' covariates and the TSLS estimation is stratified in the sense that the dataset is separated into three subdata—$\{i : d_i = 0\}$, $\{i : d_i = 1\}$ and $\{i : d_i = 2\}$—and the sets of coefficients—$\big(\gamma_0, \gamma_1\big)$, $\beta_{1,1}$ and $\big(\beta_{1,2}, \beta_{2,2}\big)$—are estimated from each subdata.

γ_1 and n, vary across the columns. As expected, the OLS estimator shows a larger bias compared to the TSLS estimator. However, the TSLS estimator has a higher variance than the OLS estimator. Notably, as $|\gamma_1|$ increases, the variance of the TSLS estimator decreases. This indicates the finite sample performance of the estimator depends on the existence of strong covariates.

Additionally, as $|\gamma_1|$ increases, the performance of the OLS estimator also improves in terms of both bias and MSE. A larger $|\gamma_1|$ means that the ordering of $\{Y_i\}_{i=1}^n$ relies less on the variation in $\{\varepsilon_i\}_{i=1}^n$ and more on the variation in $\{\gamma_1 x_i\}_{i=1}^n$. If the endogeneity problem is exacerbated by the randomness in the mapping $\{Y_j\}_{j=1}^n \mapsto \tilde{Y}_{i,k}$, then a larger $|\gamma_1|$ may help reduce bias by decreasing this randomness.

Next, we increase the parameter space by setting $\bar{d} = 5$. Since the dimension of β is

TABLE II
THE TSLS ESTIMATION WHEN THE NETWORK IS MORE CONNECTED

	bias					MSE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\hat{\beta}_{1,2}$	0.042	0.015	-0.011	-0.006	0.002	0.409	0.348	0.034	0.013	0.004
$\hat{\beta}_{2,2}$	-0.018	-0.003	0.011	-0.001	-0.004	0.089	0.088	0.016	0.008	0.003
$\hat{\beta}_{1,5}$		0.135	0.023	0.005	0.017		1.790	0.837	0.265	0.175
$\hat{\beta}_{2,5}$		-0.078	0.001	0.005	0.004		3.967	1.898	0.936	0.358
$\hat{\beta}_{3,5}$		0.082	-0.013	0.019	-0.027		5.587	2.257	1.231	0.394
$\hat{\beta}_{4,5}$		0.025	-0.006	-0.003	0.022		4.837	1.378	0.836	0.342
$\hat{\beta}_{5,5}$		-0.024	0.013	-0.003	-0.010		1.576	0.458	0.253	0.097
γ_1	1	1	2	3	5	1	1	2	3	5
$\bar{d}$	2	5	5	5	5	2	5	5	5	5

Notes: The values of the coefficients used in DGP are as follows: $n = 100$, $\gamma_0 = 1$, $\beta_{1,1} = 0.3$ and

$$\beta_{k,d} = 0.2\mathbf{1}\{k = d\} + \frac{0.1}{d-1}\mathbf{1}\{k < d\}$$

for $d = 2, \cdots, 5$. The instruments used in TSLS estimation are ordered peers' covariates and the TSLS estimation is stratified in the sense that the dataset is separated into three subdata—$\{i : d_i = 0\}, \cdots, \{i : d_i = 5\}$—and the sets of coefficients—$(\gamma_0, \gamma_1), \cdots, (\beta_{1,5}, \cdots, \beta_{5,5})$—are estimated from each subdata.

proportional to $\bar{d}^2$, this introduces 15 different coefficients for the rank-dependent peer effect in our model. Table II presents the bias and MSE of the TSLS estimator under this setting. We observe that the TSLS estimator suffers from significant bias and highly volatile variance when $\gamma_1 = 1$ (columns 2 and 7). However, as $|\gamma_1|$ increases, the variance of the TSLS estimator decreases, as shown in columns 8 through 10.

Interestingly, the MSE is larger for the coefficients $\beta_{k,5}$ where $k = 2, 3, 4$, compared to $k = 1, 5$. This may be related to our earlier observation that larger $|\gamma_1|$ improves the

TABLE III
R^2 IN THE FIRST STAGE REGRESSION

	average first stage R^2				$\Pr\{R^2 \geq 0.8\}$			
LHS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\tilde{Y}_{1,5}$	0.833	0.897	0.937	0.960	0.662	0.831	0.936	0.978
$\tilde{Y}_{2,5}$	0.847	0.896	0.930	0.955	0.715	0.835	0.928	0.974
$\tilde{Y}_{3,5}$	0.839	0.895	0.930	0.953	0.687	0.834	0.930	0.974
$\tilde{Y}_{4,5}$	0.834	0.898	0.926	0.960	0.684	0.841	0.906	0.984
$\tilde{Y}_{5,5}$	0.824	0.900	0.930	0.960	0.663	0.855	0.925	0.982
γ_1	1	2	3	5	1	2	3	5

Notes: The values of the coefficients used in DGP are as follows: $n = 100$, $\bar{d} = 5$, $\gamma_0 = 1$, $\beta_{1,1} = 0.3$ and

$$\beta_{k,d} = 0.2\mathbf{1}\{k = d\} + \frac{0.1}{d-1}\mathbf{1}\{k < d\}$$

for $d = 2, \cdots, 5$. The instruments used in TSLS estimation are ordered peers' covariates and the TSLS estimation is stratified in the sense that the dataset is separated into three subdata—$\{i : d_i = 0\}, \cdots, \{i : d_i = 5\}$—and the sets of coefficients—$(\gamma_0, \gamma_1), \cdots, (\beta_{1,5}, \cdots, \beta_{5,5})$—are estimated from each subdata.

performance of the OLS estimator. The identity of the k-th lowest-performing peer tends to fluctuate more frequently for $k = 2, 3, 4$, leading to a more severe endogeneity problem from the randomness in the mapping $\{Y_j\}_{j=1}^n \mapsto \tilde{Y}_{i,k}$. In contrast, this issue is less pronounced for $k = 1, 5$. Therefore, it may be reasonable to impose smoothness restrictions on $\beta_{k,d}$ for values of k that are not close to 1 or d.

Finally, Table III reports the average first-stage R^2 and the probability of the R^2 exceeding 0.8 for each of the endogenous variables $\tilde{Y}_{1,5}, \cdots, \tilde{Y}_{5,5}$. The results show that a weak first-stage regression exacerbates the performance of the TSLS estimator. This suggests that the potential costs of allowing for a more complex parameter space in terms of peer effect coefficients can be mitigated by carefully selecting the instruments

$\{z_i\}_{i=1}^n$ to ensure a sufficiently strong first-stage regression.

5 Empirical Illustration

To see how our model behaves with real data, we apply it to a dataset from Herstad (2023), which is collected in two Norwegian middle schools and contains school grades, test scores for a nationwide test ("National Test") taken prior to entering middle school, parental background variables, and friendship network of the students. We follow the regression specification of Alne, Herstad, and Myhre (2024); the GPA is the outcome variable with possible spillovers and the National Test score and other socioeconomic and demographic variables are the control covariates. The instruments are constructed with the National Test score.

In this paper, we expand the LIM model to allow for ordered peer outcomes to affect a student differently, as in (3). Specifically, we focus on the highest-performing peer and the lowest-performing peer. Furthermore, whenever well-defined, we let

$$\bar{Y}_{i,-1-d_i} = \frac{1}{d_i - 2} \sum_{k=2}^{d_i - 1} \tilde{Y}_{i,k}$$

$$\bar{Y}_{i,-1} = \frac{1}{d_i - 1} \sum_{k=2}^{d_i} \tilde{Y}_{i,k}$$

$$\bar{Y}_{i,-d_i} = \frac{1}{d_i - 1} \sum_{k=1}^{d_i - 1} \tilde{Y}_{i,k}.$$

As $\tilde{Y}_{i,k}$ is the k-th lowest GPA among student i's peers, $\bar{Y}_{i,-1-d_i}$ is the average GPA of student i's peers minus the one with the lowest GPA and the one with the highest GPA, and so on. For the instruments, we construct $z_{i,1}$, z_{i,d_i}, $\bar{z}_{i,-1-d_i}$, $\bar{z}_{i,-1}$ and $\bar{z}_{i,-d_i}$ accordingly, using the National Test score. For each specification, we use the corresponding instruments to the $\tilde{Y}_i$ variables contained in the specification.

Table IV

Min/max models for educational peer effects

	Dependent variable: GPA				
	(1)	(2)	(3)	(4)	(5)
$\bar{Y}$	0.483***				
	(0.131)				
$\bar{Y}_{-d_i}$			-0.003		
			(0.278)		
$\bar{Y}_{-1}$		0.678**			
		(0.295)			
$\bar{Y}_{-1,-d_i}$				0.156	
				(0.475)	
$\tilde{Y}_{d_i}$			0.43	0.303	0.407***
			(0.313)	*(0.391)*	*(0.151)*
$\tilde{Y}_1$		-0.067		0.138	0.18*
		(0.221)		*(0.202)*	*(0.106)*
Parent Income	0.146**	0.15**	0.148**	0.135**	0.136**
	(0.066)	*(0.068)*	*(0.069)*	*(0.065)*	*(0.066)*
Born in Norway	-0.277**	-0.26*	-0.363**	-0.298**	-0.321**
	(0.135)	*(0.136)*	*(0.148)*	*(0.149)*	*(0.139)*
National Test	0.388***	0.383***	0.401***	0.381***	0.381***
	(0.053)	*(0.053)*	*(0.052)*	*(0.053)*	*(0.053)*
Parental Education	0.053***	0.051***	0.049***	0.05***	0.05***
	(0.016)	*(0.016)*	*(0.017)*	*(0.016)*	*(0.016)*
Distance to School	-0.011	-0.015	0	-0.022	-0.024
	(0.048)	*(0.051)*	*(0.052)*	*(0.052)*	*(0.05)*
Female	0.468***	0.4***	0.586***	0.442***	0.466***
	(0.101)	*(0.093)*	*(0.103)*	*(0.093)*	*(0.104)*
Constant	-1.377***	-1.565***	-1.83***	-1.434***	-1.488***
	(0.345)	*(0.454)*	*(0.477)*	*(0.382)*	*(0.387)*
Observations	529	529	529	529	529
Adjusted R^2	0.476	0.474	0.450	0.480	0.476

Notes: $^{*}p < 0.01$; $^{**}p < 0.05$; $^{***}p < 0.01$. The standard errors are computed from the asymptotic variance formula discussed in Section A of the appendix.

As controls for student background, we use the log of household income, whether the student was born in Norway, the average number of years the parents have been educated, the distance the students live from school as well as the student's gender.

Table IV shows the estimation results. There are five specifications we consider. The first is the classic LIM model. The second and third specifications separate out the lowest and highest friends GPA respectively. In the fourth we separate out both the highest and lowest GPA, while in the final specification we only include the highest and lowest GPA friends, not using the average GPA of the peers inbetween.

While the estimates are noisy, there are indications that the majority of estimated peer effect in the LIM model is driven by $\tilde{Y}_{d_i}$, the friend with the highest grade. We see that the coefficient on $\tilde{Y}_1$ is consistently smaller than the coefficient on $\tilde{Y}_{d_i}$. In our specification where we only separate out $\tilde{Y}_{d_i}$, the coefficient on the remaining LIM component is almost zero. However, the estimate of the coefficient on $\tilde{Y}_{d_i}$ is not significant in this specification. In total, this exercise shows that there may be large heterogeneity in peer spillovers in education masked by the estimated effect of the LIM model, though we do not have the power in this sample to conclude anything concrete.

6 Conclusion

This paper introduced the rank-dependent peer effect model. In this model, how you affect your peer depends not only on your own outcome level but also on the outcomes of the other peers that they have. The construction of the peer effect in the model is flexible enough to allow us to discuss many interesting questions regarding how the composition of a peer group affects the peer effect: complementarity v. substitutability, returns to centrality, etc. Traditional peer effect models, such as the linear-in-means (LIM) model, tend to oversimplify the dynamics of peer influence by focusing on a single scalar parameter and assuming perfect substitutability among peers. This oversimplification neglects the heterogeneity within peer effects, where individuals may experience varying degrees of influence depending on the structure and nature of their interactions, across different empirical contexts. By developing the rank-dependent peer effect model, this paper allows researchers to investigate richer patterns of peer interactions, with specific peers—such as the highest or lowest performers—having disproportionately larger effects on outcomes in certain contexts.

While we have focused on identification and estimation through TSLS, there are multiple avenues to establish estimators with better finite sample properties. For example, the model can benefit from having an estimator that has data-driven smoothness or sparsity property. Future versions of this paper will analyse some of these avenues.

Finally, this paper has only considered settings with exogenous networks. As work on incorporating endogenous networks into peer effect models continue to develop, it would similarly be interesting to see if similar approaches can be applied to the rank-dependent peer effect model.

References

Alne, R., E. I. Herstad, and A. Myhre (2024): "Classroom connections: how school environment shapes friendships beyond socioeconomic status," unpublished working paper, available on request.

Belloni, A., C. Hansen, and W. Newey (2022): "High-dimensional linear models with many endogenous variables," Journal of Econometrics, 228(1), 4–26.

Boucher, V., M. Rendall, P. Ushchev, and Y. Zenou (2024): "Toward a general theory of peer effects," Econometrica, 92(2), 543–565.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," Journal of econometrics, 150(1), 41–55.

De Chaisemartin, C., and X. d'Haultfoeuille (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," American economic review, 110(9), 2964–2996.

De Paula, A. (2013): "Econometric analysis of games with multiple equilibria," Annu. Rev. Econ., 5(1), 107–131.

Foster, G. (2006): "It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism," Journal of public Economics, 90(8-9), 1455–1475.

Gazze, L., C. Persico, and S. Spirovska (2024): "The long-run spillover effects of pollution: How exposure to lead affects everyone in the classroom," Journal of Labor Economics, 42(2), 357–394.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2024+): "Contamination bias in linear regressions," The American Economic Review.

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

Goodman-Bacon, A. (2021): "Difference-in-differences with variation in treatment timing," Journal of econometrics, 225(2), 254–277.

Guryan, J., K. Kroft, and M. J. Notowidigdo (2009): "Peer effects in the workplace: Evidence from random groupings in professional golf tournaments," American Economic Journal: Applied Economics, 1(4), 34–68.

Han, L., and T. Li (2009): "The gender difference of peer influence in higher education," Economics of Education review, 28(1), 129–134.

Herstad, E. (2023): "Estimating peer effects and network formation models with missing links," available via www.eyoherstad.com.

Hsieh, C.-S., and L. F. Lee (2016): "A social interactions model with endogenous friendship formation and selectivity," Journal of Applied Econometrics, 31(2), 301–319.

Jochmans, K. (2023): "Peer effects and endogenous social interactions," Journal of Econometrics, 235(2), 1203–1214.

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

Kang, C. (2007): "Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea," Journal of Urban Economics, 61(3), 458–495.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," The review of economic studies, 60(3), 531–542.

Mogstad, M., and A. Torgovitsky (2024): "Instrumental Variables with Unobserved Heterogeneity in Treatment Effects," Discussion paper, National Bureau of Economic Research.

APPENDIX

A Asymptotic theory for the TSLS estimator

Assumption 5 gives a set of sufficient conditions for the asymptotic normality of the TSLS estimator. Recall the peer effect equation (9).

Assumption 5.

a. *The two matrices*

$$\frac{1}{n}\sum_{i=1}^{n} z_i W_i^{\intercal} \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} z_i z_i^{\intercal}$$

converge to full rank matrices (in probability) as $n \to \infty$.

b. $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i \varepsilon_i \xrightarrow{d} \mathcal{N}(0, \Sigma_W)$ *with some $\Sigma_W > 0$ as $n \to \infty$.*

Part **a.** of Assumption 5 is the sample counterpart of the relevance conditions discussed in Section 3. Part **b.** assumes that $\{\varepsilon_i\}_{i=1}^{n}$ satisfies some weak dependence condition and $\{z_i\}_{i=1}^{n}$ does not explode so that we have asymptotic normality on $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i \varepsilon_i$.

Theorem 2. *Assumptions 1-3 and 5 hold. Then,*

$$\sqrt{n}\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$$

as $n \to \infty$, with some consistently estimable positive definite matrix Σ.

Proof. The TSLS estimator is

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \left(\frac{1}{n} \sum_{i=1}^{n} W_i {z_i}^\intercal \left(\frac{1}{n} \sum_{i=1}^{n} z_i {z_i}^\intercal \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} z_i {W_i}^\intercal \right)^{-1} \cdot \frac{1}{n} \sum_{i=1}^{n} W_i {z_i}^\intercal \left(\frac{1}{n} \sum_{i=1}^{n} z_i {z_i}^\intercal \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i Y_i.$$

Then,

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} = \left(\frac{1}{n} \sum_{i=1}^{n} W_i {z_i}^\intercal \left(\frac{1}{n} \sum_{i=1}^{n} z_i {z_i}^\intercal \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} z_i {W_i}^\intercal \right)^{-1} \cdot \frac{1}{n} \sum_{i=1}^{n} W_i {z_i}^\intercal \left(\frac{1}{n} \sum_{i=1}^{n} z_i {z_i}^\intercal \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i \varepsilon_i.$$

By Assumption 5 $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} z_i \varepsilon_i \xrightarrow{d} \mathcal{N}(0, \Sigma_W)$, and the matrices

$$C = \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} z_i {W_i}^\intercal \quad \text{and} \quad B = \operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} z_i {z_i}^\intercal$$

exist and are full rank. Therefore by the continuous mapping theorem we get

$$\sqrt{n} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

With $\Sigma = (CB^{-1}C)^{-1} CB^{-1} \Sigma_W \left((CB^{-1}C)^{-1} CB^{-1} \right)^\intercal$

□

B Proofs

B.1 Proof for Proposition 1

WTS for any $\{x_i, \epsilon_i\}_{i=1}^n$, the peer effect model (4) admits a unique equilibrium $\{y_i\}_{i=1}^n$ such that for $i = 1, \cdots, n$,

$$y_i = \sum_{k=1}^{d_i} \tilde{y}_{i,k} \beta_{kd_i} + x_i{}^\intercal \gamma + \epsilon_i.$$

$\tilde{y}_{i,k}$ is constructed in the same way as $\tilde{Y}_{i,k}$.

Consider a function $g : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$g(y) = \begin{pmatrix} g_1(y) \\ \vdots \\ g_n(y) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{d_1} y_{1k}^p \beta_{kd_1} + x_1{}^\intercal \gamma + \epsilon_1 \\ \vdots \\ \sum_{k=1}^{d_n} y_{nk}^p \beta_{kd_n} + x_n{}^\intercal \gamma + \epsilon_n \end{pmatrix}$$

when $y = \left(y_1, \cdots, y_n\right)^\intercal$. Note that the function g depends on $\{x_i, A_{ij}, \epsilon_i\}_{1 \le i,j \le n}$ in addition to $\{\beta_{k,d}\}_{k,d}$ and γ. We show that the function g is a contraction mapping with the supremum norm: for any $y, y' \in \mathbb{R}^n$,

$$\|g(y) - g(y')\|_\infty \le \bar{\beta} \, \|y - y'\|_\infty .$$

Firstly, suppose that y and y' have the same order. The inequality holds trivially:

for each i,

$$\begin{aligned}|g_i(y) - g_i(y')| &= \left|\sum_{k=1}^{d_i} \left(\tilde{y}_{i,k} - \tilde{y}'_{i,k}\right) \beta_{k,d_i}\right| \\ &\leq \sum_{k=1}^{d_i} |\beta_{kd_i}| \cdot \|y - y'\|_\infty \leq \bar{\beta} \|y - y'\|_\infty\end{aligned}$$

with some $\bar{\beta} = \max_d \sum_{k=1}^{d} |\beta_{k,d}| < 1$. The first inequality holds since y and y' have the same order and therefore $\tilde{y}_{i,k}$ and $\tilde{y}'_{i,k}$ are the outcome of the same individual: $\left|\tilde{y}_{i,k} - \tilde{y}'_{i,k}\right|$ is bounded by $\|y - y'\|_\infty$. The second equality is from Assumption 1.

Secondly, suppose that y and y' do not have the same order.[5] Fix i and k and consider

$$\left|\left(\tilde{y}_{i,k} - \tilde{y}'_{i,k}\right) \beta_{k,d_i}\right|.$$

When individual i's k-th lowest performing peer under the outcome vector y is the same individual with that under the outcome vector y', the quantity above is bounded by $|\beta_{kd_i}| \cdot \|y - y'\|_\infty$. Suppose that the order change affects individual i's k-th lowest performing peer and WLOG let $\tilde{y}_{i,k} = y_1$ and $\tilde{y}'_{i,k} = y_2'$; individual i's k-th lowest performing peer is individual 1 under the outcome vector y and is individual 2 under the outcome vector y'. Then,

$$\left|\left(\tilde{y}_{i,k} - \tilde{y}'_{i,k}\right) \beta_{k,d_i}\right| \leq |\beta_{kd_i}| \cdot |y_1 - y_2'|$$

and individual 1's rank among individual i's peers under the outcome vector y' is either higher than k or lower than k.

[5] It is assumed throughout the proof that some tie-breaking rule is applied here so that there is one-to-one relationship between an individual's peers and the ordered outcomes $y^p_{i1}, \cdots, y^p_{id_i}$.

Case 1. Individual 1's rank among individual i's peers under the outcome vector y' is higher than k: there exists some $k' > k$ such that ${y_1}' = {y^p_{i,k'}}'$. Then, we have that

$${y_1}' = {y^p_{i,k'}}' \geq \tilde{y}'_{i,k} = {y_2}'$$

and therefore

$$y_1 - {y_2}' \geq {y_1}' - \|y - y'\|_\infty - {y_2}' \geq -\|y - y'\|_\infty.$$

Since individual 1 is ranked higher under the outcome vector y' compared to the outcome vector y where his rank was k, there must be another individual $j \neq 1$ whose rank is lower than or equal to k under the outcome vector y' and is higher than k under the outcome vector y:

$${y_j}' \leq \tilde{y}'_{i,k} = {y_2}' \quad \text{and} \quad y_j \geq \tilde{y}_{i,k} = y_1$$

and therefore

$$y_1 - {y_2}' \leq y_j - {y_2}' \leq {y_j}' + \|y - y'\|_\infty - {y_2}' \leq \|y - y'\|_\infty.$$

By combining the two inequalities, we get $|y_1 - {y_2}'| \leq \|y - y'\|_\infty$.

Case 2. Individual 1's rank among individual i's peers under the outcome vector y' is lower than k: there exists some $k' < k$ such that ${y_1}' = {y^p_{i,k'}}'$. Then, we have that

$${y_1}' = {y^p_{i,k'}}' \leq \tilde{y}'_{i,k} = {y_2}'$$

and therefore

$$y_1 - {y_2}' \leq {y_1}' + \|y - y'\|_\infty - {y_2}' \leq \|y - y'\|_\infty.$$

Since individual 1 is ranked lower under the outcome vector y' compared to the outcome vector y where his rank was k, there must be another individual $j \neq 1$ whose rank is higher than or equal to k under the outcome vector y' and is lower than k under the outcome vector y:

$${y_j}' \geq \tilde{y}'_{i,k} = {y_2}' \quad \text{and} \quad y_j \leq \tilde{y}_{i,k} = y_1$$

and therefore

$$y_1 - {y_2}' \geq y_j - {y_2}' \geq {y_j}' - \|y - y'\|_\infty - {y_2}' \geq -\|y - y'\|_\infty.$$

By combining the two inequalities, we get $|y_1 - {y_2}'| \leq \|y - y'\|_\infty$.

By aggregating across k for the fixed i and then aggregating across i, we get

$$\|g(y) - g(y')\|_\infty \leq \bar{\beta} \, \|y - y'\|_\infty \, .$$

Assume to the contrary that there exist $y, y' \in \mathbb{R}^n$ such that $y \neq y'$, $y = g(y)$ and $y' = g(y')$. Let $c = \|y - y'\|_\infty > 0$. Then,

$$c = \|y - y'\|_\infty = \|g(y) - g(y')\|_\infty \leq \bar{\beta}\|y - y'\|_\infty \leq \bar{\beta}c < c$$

from Assumption 1, leading to a contradiction.

□

B.2 Proof for Corollary 1

Let π denote a ordering on $\{Y_i, \cdots, Y_n\}$ with some tiebreaking rule.

$$\pi : \{1, \cdots, n\} \mapsto \{1, \cdots, n\}$$

and satisfies that $Y_{\pi(1)} \leq \cdots \leq Y_{\pi(n)}$. Note that π is a function of $\{\varepsilon_i\}_{i=1}^n$; from Proposition 1, we have shown that there exists a unique equilibrium and thus a unique ordering for each realization $\{\epsilon_i\}_{i=1}^n$. Let Π denote the set of all possible ordering on $\{Y_1, \cdots, Y_n\}$. Π is a function of $\{x_i, A_{ij}\}_{i,j}$ and the distribution of $\{\varepsilon_i\}_i$.

For any realization of $\{\epsilon_i\}_{i=1}^n$ and the corresponding ordering π, there is a reduced-form linear relationship between $\{y_i\}_{i=1}^n$ and $\{x_i\}_{i=1}^n$: construct a $n \times n$ matrix $\mathbb{B}(\pi)$ such that

$$\mathbb{B}(\pi) = \Big(\beta_{ij}(\pi)\Big)_{i,j}$$

$$\beta_{ij}(\pi) = \begin{cases} 0 & \text{if } A_{ij} = 0 \\ \beta_{kd_i} \text{ for some } k \text{ s.t. } \sum_{j'=1}^{\pi^{-1}(j)} A_{i\pi(j')} = k & \text{if } A_{ij} = 1 \end{cases}$$

$\mathbb{B}(\pi)$ takes the ordering π as fixed and finds the corresponding rank-dependent coefficient $\beta_{k,d}$ for each of the peer outcomes. Find that $(I_n - \mathbb{B}(\pi))$ is invertible for any π: suppose that there is some $x \neq \mathbf{0}$ such that $x = \mathbb{B}(\pi)x$; for some i such that $|x_i| = \|x\|_\infty$,

$$|x_i| = \left| \sum_{j=1}^{n} \beta_{ij}(\pi) x_j \right| \leq \sum_{j=1}^{j} |\beta_{ij}(\pi)| \cdot |x_j| \leq \bar{\beta}|x_i| < |x_i|,$$

leading to a contradiction; $I - \mathbb{B}(\pi)$ is full rank. Then,

$$\begin{aligned}\mathbb{Y} &= \mathbb{B}(\pi)\mathbb{Y} + \mathbb{X}\gamma + \mathbb{E} \\ \mathbb{Y} &= \left(I - \mathbb{B}(\pi)\right)^{-1}\mathbb{X}\gamma + \left(I_n - \mathbb{B}(\pi)\right)^{-1}\mathbb{E}.\end{aligned}$$

We have a reduced-form linear relationship between $\{y_i\}_{i=1}^n$ and $\{x_i\}_{i=1}^n$.

Since the reduced-form relationship holds for every realization of $\{\epsilon_i\}_{i=1}^n$ such that the ordering π stays the same, we can consider a conditional expectation of the linear relationship given the event that $\{\varepsilon_i\}_{i=1}^n$ induces the ordering π. Let Π denote the set of all possible orderings π, then

$$\begin{aligned}\mathbf{E}_n\left[Y_i\right] &= \sum_{\pi\in\Pi}\sum_{j=1}^n \theta_{ij}(\pi)\big(x_j{}^\intercal\gamma + \mathbf{E}_n\left[\varepsilon_j|\pi\right]\big)\cdot \mathrm{Pr}_n\left\{\pi\right\} \\ &= \sum_{j=1}^n \theta_{ij}x_j{}^\intercal\gamma + \eta_i\end{aligned}$$

where $\Big(\theta_{i1}(\pi), \cdots, \theta_{in}(\pi)\Big)$ is the i-th row of the $n\times n$ matrix $\left(I - \mathbb{B}(\pi)\right)^{-1}$ and

$$\begin{aligned}\theta_{ij} &= \sum_{\pi\in\Pi}\theta_{ij}(\pi)\mathrm{Pr}_n\left\{\pi\right\} \\ \eta_i &= \sum_{\pi\in\Pi}\sum_{j=1}^n \theta_{ij}(\pi)\mathbf{E}_n\left[\varepsilon_j|\pi\right]\cdot \mathrm{Pr}_n\left\{\pi\right\}.\end{aligned}$$

Likewise, by taking a different student while summing over the ordering π, we get

$$\begin{aligned}\mathbf{E}_n\left[\tilde{Y}_{i,k}\right] &= \sum_{\pi\in\Pi}\sum_{j=1}^n \tilde{\theta}_{i,k,j}(\pi)\big(x_j{}^\intercal\gamma + \mathbf{E}_n\left[\varepsilon_j|\pi\right]\big)\cdot \mathrm{Pr}_n\left\{\pi\right\} \\ &= \sum_{j=1}^n \tilde{\theta}_{i,k,j}x_j{}^\intercal\gamma + \tilde{\eta}_{i,k}\end{aligned}$$

where $\left(\tilde{\theta}_{i,k,1}(\pi), \cdots, \tilde{\theta}_{i,k,n}(\pi)\right)$ is one row of the $n \times n$ matrix $(I_n - \mathbb{B}(\pi))^{-1}$, which corresponds to the k-th lowest performing peer of student i, and

$$\tilde{\theta}_{i,k,j} = \sum_{\pi \in \Pi} \tilde{\theta}_{i,k,j}(\pi) \mathrm{Pr}_n \{\pi\}$$

$$\tilde{\eta}_{i,k} = \sum_{\pi \in \Pi} \sum_{j=1}^{n} \tilde{\theta}_{i,k,j}(\pi) \mathbf{E}_n [\varepsilon_j | \pi] \cdot \mathrm{Pr}_n \{\pi\}.$$

□

B.3 Proof for Theorem 1

Construct a $\frac{\bar{d}(\bar{d}+1)}{2} \times 1$ vector $\tilde{Y}_i$ such that

$$\tilde{Y}_i = \begin{pmatrix} 0 & \cdots & Y_{i1}^p & \cdots & Y_{id_i}^p & \cdots & \mathbf{0}_{\bar{d}}{}^\intercal \end{pmatrix}.$$

$\tilde{Y}_i$ stacks up (hypothetical) vectors of peer outcomes with varying peer numbers $d = 1, \cdots, \bar{d}$; the peer outcome vector of peer number d is nonzero only when $d = d_i$. Let $W_i = \left(\tilde{Y}_i^\intercal, x_i{}^\intercal\right)^\intercal$. Then, the rank-dependent peer effect model can be written as follows:

$$Y_i = \tilde{Y}_i^\intercal \beta + x_i{}^\intercal \gamma + \varepsilon_i = W_i{}^\intercal \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \varepsilon_i.$$

Likewise, construct a vector of instrument z_i such that z_i stacks up vectors of instruments with varying peer numbers $d = 1, \cdots, \bar{d}$.

$$z_i = \begin{pmatrix} z_{i1}{}^\intercal & \cdots & z_{i\bar{d}}{}^\intercal & x_i{}^\intercal \end{pmatrix}^\intercal.$$

z_{id}, the instrument vector of peer number d, is nonzero only when $d = d_i$.

Consider a stacked up matrix of the peer effect model (4):

$$\mathbb{Y} = \mathbb{W}\begin{pmatrix}\beta \\ \gamma\end{pmatrix} + \mathbb{E}$$

$$\mathbb{Z}^{\intercal}\mathbb{Y} = \mathbb{Z}^{\intercal}\mathbb{W}\begin{pmatrix}\beta \\ \gamma\end{pmatrix} + \mathbb{Z}^{\intercal}\mathbb{E}.$$

From Assumption 3, by taking expectation conditional upon $\{x_i, A_{ij}\}_{i,j}$,

$$\mathbf{E}_n\left[\mathbb{Z}^{\intercal}\mathbb{Y}\right] = \mathbf{E}_n\left[\mathbb{Z}^{\intercal}\mathbb{W}\right]\begin{pmatrix}\beta \\ \gamma\end{pmatrix}.$$

The parameter of interest β and γ are identified from the conditional distribution of $\{Y_i\}_{i=1}^n$ given $\{x_i, A_{ij}\}_{1\leq i,j\leq n}$ when $\mathbf{E}_n\left[\mathbb{Z}^{\intercal}\mathbb{W}\right]$ is full rank.

From Corollary 1, we have the tools to write out the matrix $\mathbf{E}_n\left[\mathbb{Z}^{\intercal}\mathbb{W}\right]$:

$$\mathbf{E}_n\left[\mathbb{Z}^{\intercal}\mathbb{W}\right] = \begin{pmatrix} Q_{ZY^p,1} & \mathbf{O}_{l\times 2} & \cdots & \mathbf{O}_{l\times \bar{d}} & Q_{ZX,1} \\ \mathbf{O}_{2l\times 1} & Q_{ZY^p,2} & \cdots & \mathbf{O}_{2l\times \bar{d}} & Q_{ZX,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{O}_{\bar{d}l\times 1} & \mathbf{O}_{\bar{d}l\times 2} & \cdots & Q_{ZY^p,\bar{d}} & Q_{ZX,\bar{d}} \\ Q_{XY^p,1} & Q_{XY^p,2} & \cdots & Q_{XY^p,\bar{d}} & Q_{XX} \end{pmatrix}$$

where

$$
\begin{aligned}
Q_{Z\tilde{Y},d} &= \sum_{i=1}^{n} z_{id} \left(\mathbf{E}_n \left[\tilde{Y}_{i,1} \right] \quad \cdots \quad \mathbf{E}_n \left[\tilde{Y}_{i,d_i} \right] \right) \mathbf{1}\{d_i = d\}, \\
Q_{ZX,d} &= \sum_{i=1}^{n} z_{id} x_i{}^{\intercal} \mathbf{1}\{d_i = d\}, \\
Q_{X\tilde{Y},d} &= \sum_{i=1}^{n} x_i \left(\mathbf{E}_n \left[\tilde{Y}_{i,1} \right] \quad \cdots \quad \mathbf{E}_n \left[\tilde{Y}_{i,d_i} \right] \right) \mathbf{1}\{d_i = d\}, \\
Q_{XX} &= \sum_{i=1}^{n} x_i x_i{}^{\intercal}.
\end{aligned}
$$

A slight abuse of notation is applied here: the row vectors in the summation of $Q_{Z\tilde{Y},d}$ and $Q_{X\tilde{Y},d}$ do not have the same dimension, but the indicator $\mathbf{1}\{d_i = d\}$ only selects the ones with d components.

Thanks to the diagonal structure of $Q_{Z\tilde{Y},1}, \cdots, Q_{Z\tilde{Y},\bar{d}}$, each submatrix

$$
Q_d := \begin{pmatrix} Q_{Z\tilde{Y},d} & Q_{ZX,d} \\ Q_{X\tilde{Y},d} & Q_{XX} \end{pmatrix}
$$

being full rank for $d = 1, \cdots, \bar{d}$ implies that $\mathbb{Z}^{\intercal} \mathbf{E}_n [\mathbb{W}]$ is full rank. From the reduced-form representation of $\mathbf{E}_n \left[\tilde{Y}_{i,k} \right]$ with θ and η, we get

$$
Q_d = \sum_{i=1}^{n} \begin{pmatrix} z_{i,d} \\ x_i \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n} \tilde{\theta}_{i,1,j} x_j{}^{\intercal} \gamma + \tilde{\eta}_{i,1} \\ \vdots \\ \sum_{j=1}^{n} \tilde{\theta}_{i,d_i,j} x_j{}^{\intercal} \gamma + \tilde{\eta}_{i,d_i} \\ x_i \end{pmatrix}^{\intercal} \mathbf{1}\{d_i = d\}.
$$

From Assumption 3, Q_d has full rank.

□

B.4 Proof for Proposition 2

The LIM model is:

$$Y_i = \beta^{\text{LIM}} \bar{Y}_i + x_i{}^\intercal \gamma + \varepsilon_i,$$

With some z_i that is a function of $\{x_i, A_{i,j}\}_{i,j}$ is used as instruments, the moment condition used in the estimation of the LIM estimator is:

$$\begin{aligned}
\mathbf{E}_n\left[\mathbb{Y}\right] &= \beta^{\text{LIM}} G \mathbf{E}_n\left[\mathbb{Y}\right] + \mathbb{X}\gamma \\
\mathbb{M}_X \mathbf{E}_n\left[\mathbb{Y}\right] &= \beta^{\text{LIM}} \mathbb{M}_X G \mathbf{E}_n\left[\mathbb{Y}\right] + \mathbb{M}_X \mathbb{X}\gamma \\
&= \beta^{\text{LIM}} \mathbb{M}_X G \mathbf{E}_n\left[\mathbb{Y}\right] \\
\mathbb{Z}^\intercal \mathbb{M}_X \mathbf{E}_n\left[\mathbb{Y}\right] &= \beta^{\text{LIM}} \mathbb{Z}^\intercal \mathbb{M}_X G \mathbf{E}_n\left[\mathbb{Y}\right] \\
\beta^{\text{LIM}} &= \left(\mathbb{Z}^\intercal \mathbb{M}_X G \mathbf{E}_n\left[\mathbb{Y}\right]\right)^{-1} \mathbb{Z}^\intercal \mathbb{M}_X \mathbf{E}_n\left[\mathbb{Y}\right]
\end{aligned}$$

where G is the row-normalized network adjacency matrix and

$$\mathbb{M}_X = I_n - \mathbb{X}\left(\mathbb{X}^\intercal \mathbb{X}\right)^{-1} \mathbb{X}^\intercal.$$

We know that

$$\mathbf{E}_n\left[Y_i | \{x_i, A_{ij}\}_{i,j}\right] = \sum_{k=1}^{d_i} \beta_{k,d} \mathbf{E}_n\left[\tilde{Y}_{i,k}\right] + x_i{}^\intercal \gamma$$

from Assumption 3. Then, the misspecified parameter β^{LIM} becomes

$$\beta^{\mathrm{LIM}} = \left(\mathbb{Z}^{\intercal}\mathbb{M}_X G \mathbf{E}_n\left[\mathbb{Y}\right]\right)^{-1} \mathbb{Z}^{\intercal}\mathbb{M}_X \begin{pmatrix} \sum_{k=1}^{d_1} \beta_{k,d}\mathbf{E}_n\left[\tilde{Y}_{1,k}\right] \\ \vdots \\ \sum_{k=1}^{d_n} \beta_{k,d}\mathbf{E}_n\left[\tilde{Y}_{n,k}\right] \end{pmatrix}$$

By rewriting the matrix multiplications as a summation with

$$\hat{\mathcal{A}} = \sum_{i=1}^{n} z_i {x_i}^{\intercal} \left(\sum_{i=1}^{n} x_i {x_i}^{\intercal}\right)^{-1},$$

we get

$$\begin{aligned}\beta^{\mathrm{LIM}} &= \frac{\sum_{i=1}^{n}\left(z_i - \hat{\mathcal{A}}x_i\right)\sum_{k=1}^{d_i}\mathbf{E}_n\left[\tilde{Y}_{i,k}\right]\beta_{k,d}}{\sum_{i=1}^{n}\left(z_i - \hat{\mathcal{A}}x_i\right)\frac{1}{d_i}\sum_{k=1}^{d_i}\mathbf{E}_n\left[\tilde{Y}_{i,k}\right]} \\ &= \frac{\sum_{d=1}^{\bar{d}}\sum_{k=1}^{d}\sum_{i=1}^{n}\mathbf{E}_n\left[\left(z_i - \hat{\mathcal{A}}x_i\right)\tilde{Y}_{i,k}\mathbf{1}\{d_i = d\}\right]\beta_{k,d}}{\sum_{d=1}^{\bar{d}}\sum_{i=1}^{n}\frac{1}{d}\sum_{k=1}^{d}\mathbf{E}_n\left[\left(z_i - \hat{\mathcal{A}}x_i\right)\tilde{Y}_{i,k}\mathbf{1}\{d_i = d\}\right]}.\end{aligned}$$

By letting

$$w_{k,d}^{\mathrm{LIM}} = \frac{\sum_{i=1}^{n}\frac{1}{d}\mathbf{E}_n\left[\left(z_i - \hat{\mathcal{A}}x_i\right)\tilde{Y}_{i,k}\mathbf{1}\{d_i = d\}\right]}{\sum_{d=1}^{\bar{d}}\sum_{k=1}^{d}\sum_{i=1}^{n}\frac{1}{d}\mathbf{E}_n\left[\left(z_i - \hat{\mathcal{A}}x_i\right)\tilde{Y}_{i,k}\mathbf{1}\{d_i = d\}\right]},$$

β^{LIM} becomes

$$\beta^{\mathrm{LIM}} = \sum_{d=1}^{\bar{d}}\sum_{k=1}^{d} w_{k,d}^{\mathrm{LIM}} d\beta_{k,d}.$$

The weights satisfy the sum-to-one constraints: $\sum_{d=1}^{\bar{d}}\sum_{k=1}^{d} w_{k,d}^{\mathrm{LIM}} = 1$.

Likewise, for the LIS peer effect parameter, we get

$$\begin{aligned}\beta^{\text{LIM}} &= \frac{\sum_{d=1}^{\bar{d}} \sum_{k=1}^{d} \sum_{i=1}^{n} \mathbf{E}_n \left[\left(z_i - \hat{\mathcal{A}} x_i \right) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right] \beta_{k,d}}{\sum_{d=1}^{\bar{d}} \sum_{i=1}^{n} \mathbf{E}_n \left[\left(z_i - \hat{\mathcal{A}} x_i \right) \sum_{k=1}^{d} \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right]} \\ &= \sum_{d=1}^{\bar{d}} \sum_{k=1}^{d} w_{k,d}^{LIS} \beta_{k,d}.\end{aligned}$$

where

$$w_{k,d}^{\text{LIS}} = \frac{\sum_{i=1}^{n} \mathbf{E}_n \left[\left(z_i - \hat{\mathcal{A}} x_i \right) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right]}{\sum_{d=1}^{\bar{d}} \sum_{k=1}^{d} \sum_{i=1}^{n} \mathbf{E}_n \left[\left(z_i - \hat{\mathcal{A}} x_i \right) \tilde{Y}_{i,k} \mathbf{1}\{d_i = d\} \right]}.$$

Again, the weights sum to one: $\sum_{d=1}^{\bar{d}} \sum_{k=1}^{d} w_{k,d}^{\text{LIS}} = 1$.

When z_i is not a scalar, we can modify the argument above and get similar expressions, by replacing $z_i - \hat{\mathcal{A}} x_i$ with $c^\intercal \left(z_i - \hat{\mathcal{A}} x_i \right)$ with some weighting c. In the canonical TSLS case, the weighting vector is

$$c^\intercal = \mathbf{E}_n \left[\mathbb{Y}^\intercal G^\intercal \mathbb{M}_X \mathbb{Z} \left(\mathbb{Z}^\intercal \mathbb{Z} \right)^{-1} \right];$$

misspecification weights depend on the weighting matrix when the LIM or LIS is overidentified.

□