Peer Effects Heterogeneity and Social Networks in Education

Livia Shkoza (University of Konstanz, GSDS)
Derya Uysal (University of Munich, CESifo)
Winfried Pohlmeier (University of Konstanz, CASCB, ICEA)

Discussion Paper No. 423

September 06, 2023

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Georg Weizsäcker, Ph.D., Humboldt University Berlin, 10117 Berlin, Germany
info@rationality-and-competition.de

Peer effects heterogeneity and social networks in education*

Livia Shkoza[a]
University of Konstanz, GSDS

S. Derya Uysal[b]
University of Munich, CESifo

Winfried Pohlmeier[c]
University of Konstanz, CASCB, ICEA

August 16, 2023

Abstract

This study focuses on the role of heterogeneity in network peer effects by accounting for network-specific factors and different driving mechanisms of peer behavior. We propose a novel Multivariate Instrumental Variable (MVIV) estimator which is consistent for a large number of networks keeping the individual network size bounded. We apply this approach to estimate peer effects on school achievement exploiting the network structure of friendships within classrooms. The empirical evidence presented is based on a unique network dataset from German upper secondary schools. We show that accounting for heterogeneity is not only crucial from a statistical perspective, but also yields new structural insights into how class size and gender composition affect school achievement through peer behavior.

*Financial support by the Graduate School of Decision Sciences (GSDS), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2117 - 422037984 and the IAAE is gratefully acknowledged. Derya Uysal gratefully acknowledges financial support by DFG through CRC TRR 190. The usual disclaimer applies. The authors report there are no competing interests to declare.

[a]Department of Economics, Universitätsstraße 1, D-78462 Konstanz, Germany. Phone: +49-7531-88-4214, email: Livia.Shkoza@uni-konstanz.de.

[b]Department of Economics, Ludwigstr. 33, D-80539 Munich, Germany. Phone: +49 89-2180-2224, email: Derya.Uysal@econ.lmu.de.

[c]Department of Economics, Universitätsstraße 1, D-78462 Konstanz, Germany. Phone: +49-7531-88-2660, email: Winfried.Pohlmeier@uni-konstanz.de.

1 Introduction

In the social sciences, it is an uncontested hypothesis that social interactions shape an individual's behavior and the behavior of groups of individuals as a whole. A popular framework of representing these interactions is peer effects models, which have a straightforward interpretation but also suffer from their stylized nature. This holds especially true in education where the identification and estimation of peer effects are crucial issues. Closely related to the identification issue is the question of to what extend the strength of peer effects is driven by observable network-specific factors. For instance, in the context of individual educational attainment, obvious candidates for such factors are class size and gender composition, two factors frequently considered to be major determinants of individual performance at school.

This paper takes a closer look at the role of heterogeneity within network peer effects models by augmenting the linear-in-means peer effects model in various dimensions. In particular, we focus on heterogeneous peer effects by accounting for network-specific factors - class size and gender composition - and allowing for different driving mechanisms of peer behavior. Since there is no suitable tool to estimate and test for these dimensions of heterogeneity in peer effects, we propose a novel Multivariate Instrumental Variable (MVIV) approach. Contrary to the IV-GMM type of estimators applied to cross-sectional data which typically assume that the network size approaches infinity, the MVIV estimator introduced here exploits the information from a set of networks and is consistent for a large number of networks keeping the individual network sizes bounded. Moreover, we show that the approach is equivalent to a two-step IV-Minimum Distance (IVMD) approach, which offers several advantages in applied work.

Using our proposed estimation strategy, we provide empirical evidence on the existence and importance of peer effects heterogeneity in education. This evidence is drawn from a previously unexploited dataset for German secondary schools, which contains unique information on friendship networks within classes. We show that neglecting heterogeneity due to network specific features leads to insignificant and potentially misleading findings.

In his seminal work, Manski (1993) explains the dependence of an individual's behavior on the behavior of others in a socially interactive environment through three possible effects: (i) endogenous peer effects (the individual is influenced by the peers' behavior), (ii) exogenous peer effects (the individual is affected by the peers' characteristics), and (iii) correlated effects (individuals' outcomes are similar due to similar environments or common unobserved shocks). Separate identification of these effects is far from trivial and several strategies, primarily driven by data availability issues, have been proposed (see Bramoullé et al. 2020, for a comprehensive survey on the existing literature). Bramoullé et al. (2009) show that exogenous information on the second (and higher) order peers can serve as valid instruments to identify the endogenous peer effect, when network data is observed. This idea of exploiting the network structure by using exogenous variation in the covariates of the second-order peers ('peers of the peers') as valid instruments is similar to the IV/GMM estimation approaches for spatial lag models (e.g., Kelejian & Prucha 1998, 1999). It is important to note that the identification via instrumental variables as suggested by Bramoullé et al. (2009) and several other papers following them is only valid if the networks are exogenous conditional on observed characteristics and network fixed effects, and if all networks are fully observed without measurement errors.

Following Bramoullé et al. (2009), several extensions have been proposed to estimate the endogenous peer effect using network information. As the baseline model we use the (homogeneous) composite peer effects specification proposed by Liu et al. (2014) as such an extension. First, the network-specific peer effects can be given a structural interpretation, as the model relies on a microeconomic foundation, specifically utility-maximizing behavior in a Nash equilibrium (similar to Blume et al. 2015, Calvó-Armengol et al. 2009, among others). Second, the model allows for two channels through which peer effects operate. One hypothesis is that individuals align their behavior with the norm of their peer group represented by the mean behavior of the peers, because deviating from the norm may inflict utility losses. This idea is reflected in the composite model by the local-average part, which uses the mean behavior of the peer group as a predictor for the individual's behavior (for example, Boucher et al. 2014). An alternative hypothesis is that an individual's return to their own effort increases with the aggregate quality of the peers. This is reflected in the local-aggregate part of the composite model. Thus, the composite model allows the data to decide in which way and how strongly peer behavior affects an individual's behavior. In this study, we account for heterogeneity modelling the two peer effects parameters –the local-average and the local-aggregate coefficients of the composite model– as functions of network specific features.

Irrespective of the identification strategies applied, the implicit assumption made in most of the peer effect studies is that the endogenous peer effects are homogeneous across networks. Thus, the empirical evidence on heterogeneous peer effects in education exploiting network information is rather limited. A notable exception is Calvó-Armengol et al. (2009), who investigate the relationship between peer effects and the network topology. They only provide graphical evidence that the strength of the network effect varies with certain structural network measures, such as density, asymmetry, and redundancy. Contrary to Calvó-Armengol et al. (2009) we study the link between network characteristics and the endogenous peer effect in a structural way within a regression approach. In the study on peer effects heterogeneity in education by Masten (2018) the endogenous peer effect is pair-specific, purely random and contrary to our approach, not driven by observable factors. Using a laboratory experiment on individual performance Beugnot et al. (2019) study the role of gender heterogeneity in social networks based on the model by Arduini et al. (2020). From a technical point of view the approach of Arduini et al. (2020) is similar to the composite model as both approaches allow for different peer effects resulting from the presence of several networks. However, in the study conducted by Beugnot et al. (2019) the focus of interest is to detect differences in peer effects between males and females, while our study focuses on the question whether networks operate differently depending on the gender composition.

The empirical evidence on the role of peer effects in education is rather scarce due to the limited availability of appropriate network data in educational context. To the best of our knowledge, most of the existing research papers on network peer effects in education use the National Longitudinal Study of Adolescent to Adult Health (Add Health) (see Bramoullé et al. 2009, Boucher et al. 2014). In our paper, we study heterogeneous peer effects on school grades using unique network data from 85 school classes of secondary schools in Germany. Besides using network data which has not been used before, our study also contributes to the empirical literature on the determinants of school performance by providing deeper insights on how certain network features may affect individual performance through peer behavior. In particular, our study provides a better understanding of how gender composition and class size affect an individual's school grades by enhancing peer behavior.

One strand of the literature on gender effects in school performance concentrates on the difference of outcomes for girls in single-sex and coeducational classes (see for a review Mael et al. 2005, Morse 1999). The results based on observational studies are somewhat mixed: some studies provide evidence for positive effects of single-sex schools, whereas others suggest no difference. The other strand of the literature identifies the gender peer effect using exogenous variation in gender due to experimental or quasi-experimental research design. Hoxby (2002) and Lavy & Schlosser (2011) find that the proportion of female students has positive effects on students' cognitive achievements. They do not find a differential effect on boys and girls. However, none of these studies explicitly considers the classrooms' network structure. The common point in these studies is that gender (or the gender ratio) enters the reduced form equations as a regressor. In contrast in our structural approach, the gender ratio affects the outcome of academic success through the endogenous peer effect. This indirect effect has a clear structural interpretation in the sense that observed differences in academic success between classes with different gender compositions have their roots in different collaborative patterns captured by the peer effects.

Our study also contributes to the long-lasting debate on the effect of class size on academic success. The empirical evidence on this issue is by no means unambiguous. For example, Hanushek (1996) and Hoxby (2000) find no effect of class size reduction on achievement. The results of Dobbelsteen et al. (2002) suggest that students in smaller classes do not have better academic performance (and even sometimes worse) than students in larger classes. On the other hand, Angrist & Lavy (1999), Krueger (1999), and more recent studies, Heinesen (2010), Fredriksson et al. (2012), report a substantial positive effect of reducing class size on academic achievement. Like the studies on gender effects, the vast majority of the empirical studies concentrate on direct effects of class size on school success within reduced form settings. Our study shows that peer effects decrease with class size.

The outline of this paper is as follows. In Section 2 we introduce the composite network model and elaborate on its identification conditions. Moreover, we introduce the new MVIV approach for the modeling and estimation of heterogeneous peer effects and its Minimum Distance representation. In Section 3, we describe our network data and discuss further implementation issues. Section 4 contains the major empirical findings. Section 5 concludes and gives an outlook for future research.

2 The Network Model and Estimation

As a baseline model, we rely in the following on the composite peer effects model by Liu et al. (2014). We assume there is a set of L independent networks, each consisting of n_l

agents ('the network size') and an overall number of observations $N = \sum_{l=1}^{L} n_l$. The social connections for network l are indicated by the adjacency matrix $A_l = [a_{ij,l}]$, where $a_{ij,l} = 1$ if agent i in network l is connected with agent j, and $a_{ij,l} = 0$ otherwise. The diagonal elements $a_{ii,l}$ are set to zero. The reference group of agent i in network l is the set of their peers, and the size of the reference group is the (out)degree $a_{i,l} = \sum_{j=1}^{n_l} a_{ij,l}$. Let $G_l = [g_{ij,l}]$ be the row-normalized adjacency matrix of network l, with elements $g_{ij,l} = a_{ij,l}/a_{i,l}$, where by construction $0 \leq g_{ij,l} \leq 1$ and $\sum_{j=1}^{n_l} g_{ij,l} = 1$.

The econometric specification for the composite peer effects model is of the form

$$y_{i,l} = \eta_l + x'_{i,l}\gamma + \sum_{j=1}^{n_l} g_{ij,l} x'_{j,l}\gamma_g + \beta_{a,l} \sum_{j=1}^{n_l} a_{ij,l} y_{j,l} + \beta_{g,l} \sum_{j=1}^{n_l} g_{ij,l} y_{j,l} + \varepsilon_{i,l}, \tag{1}$$

for $i = 1, \ldots, n_l$ and $l = 1, \ldots, L$ and $E[\varepsilon_l | x_l, A_l, \eta_l] = 0$. This exogeneity assumption is rather common but strong. It requires that there are no unobserved factors which affect both the link formation and the outcome variable. If this assumption fails to hold, then our approach will yield inconsistent estimates as the approaches by Bramoullé et al. (2009) and Liu et al. (2014).[1] Note that the parameters $\beta_{a,l}$ and $\beta_{g,l}$ on the local-aggregate and the local-average endogenous peer effects are heterogeneous, i.e., they are network-specific. There are k_x direct exogenous factors and k_x average exogenous peer group factors with the corresponding parameter vectors γ and γ_g. Finally, the correlated effect is given by η_l. Liu et al. (2014) show that the econometric model (1) can be derived from a utility maximizing network game with a unique Nash equilibrium.

Under homogeneity of the peer effects, i.e., $\beta_{a,l} = \beta_a$ and $\beta_{g,l} = \beta_g$ our model reduces to the model of Liu et al. (2014). The composite model nests two specifications, the local-aggregate ($\beta_{g,l} = 0$) and the local-average model ($\beta_{a,l} = 0$).

[1]Alternatively, one could rely on a model with endogenous link formation. See for example Auerbach (2022) and Johnsson & Moon (2021) for endogenous link formation approach. However, in such models it is not straightforward how to incorporate the heterogeneity of peer effects into the model. We leave this for future research and focus on the heterogeneity aspect of peer effects.

In matrix notation, the general specification (1) for network l takes the form

$$y_l = \eta_l \iota_{n_l} + X_l\gamma + X_l^g\gamma_g + \beta_{a,l}y_l^a + \beta_{g,l}y_l^g + \varepsilon_l, \tag{2}$$

where $y_l = (y_{1,l}, \ldots, y_{n_l,l})'$, $y_l^a = A_l y_l$, $y_l^g = G_l y_l$, $X_l = \left[x_{1,l}\ x_{2,l} \cdots\ x_{n_l,l}\right]'$ and $X_l^g = G_l X_l$, while $\varepsilon_l = (\varepsilon_{1,l}, \ldots, \varepsilon_{n_l,l})'$ and ι_{n_l} is an $n_l \times 1$ vector of ones.

We partial out the network-specific level effects via transformation by multiplying (2) by $J_l = I_{n_l} - \frac{1}{n_l}\iota_{n_l}\iota'_{n_l}$ from left. Because $J_l\iota_{n_l} = 0$, the transformed model is

$$\tilde{y}_l = \tilde{X}_l\gamma + \tilde{X}_l^g\gamma_g + \beta_{a,l}\tilde{y}_l^a + \beta_{g,l}\tilde{y}_l^g + \tilde{\varepsilon}_l\,, \tag{3}$$

with $\tilde{y}_l = J_l y_l$, $\tilde{X}_l = J_l X_l$, $\tilde{X}_l^g = J_l X_l^g$ and $\tilde{\varepsilon}_l = J_l\varepsilon_l$. The regressor matrix $\mathbf{W}_l = \left[\tilde{X}_l\ \tilde{X}_l^g\ \tilde{y}_l^a\ \tilde{y}_l^g\right]$ with dimension $n_l \times k_w$, where $k_w = 2(1+k_x)$, has column full rank. Provided that the identification conditions hold, as outlined in Proposition 3 in Liu et al. (2014), the parameters in (3) can be estimated consistently by instrumental variables for a single network l assuming that the size of the network approaches infinity.[2] Instead, we assume in the following that the size of each network is random and bounded, i.e., $n_{min} \leq n_l \leq n_{max}$. We believe that this is a more realistic assumption when the networks consist of friendship links in school classes as in our empirical study.

In order to introduce network heterogeneity we assume that the parameters $\beta_{a,l}$ and $\beta_{g,l}$ for network-specific peer effects can be explained by a set of network-specific observable factors

$$\beta_{j,l} = \beta_{0,j} + v_l'\beta_j \text{ for } \qquad j = a, g\,, \tag{4}$$

where v_l is a $k_v \times 1$ vector of network-specific characteristics, and $\beta_{0,a}$ and $\beta_{0,g}$ denote the

[2]The identification conditions for the local-average model and local-aggregate model are discussed in Proposition 5 in Bramoullé et al. (2009) and in Proposition 2 in Liu et al. (2014), respectively.

intercept terms for $j = a, g$, respectively. Inserting (4) into (3) gives in matrix notation

$$\tilde{y}_l = \mathbf{X}_l \, \theta + \tilde{\varepsilon}_l \,, \tag{5}$$

with:

$$\begin{aligned} \mathbf{X}_l &= \begin{bmatrix} \tilde{X}_l & \tilde{X}_l^g & \tilde{y}_l^a & \tilde{y}_l^g & \tilde{y}_l^a v_l' & \tilde{y}_l^g v_l' \end{bmatrix} \\ \theta &= \left(\gamma', \gamma_g', \beta_{0,a}, \beta_{0,g}, \beta_a', \beta_g'\right)' \end{aligned}$$

It is important to note that the $n_l \times k$ regressor matrix $\mathbf{X}_l$ with $k = 2(1 + k_x + k_v)$ has no full column rank because the interaction terms with v_l imply perfect multicollinearity. Consequently, IV-estimation of θ for a single network becomes infeasible (see Appendix A.1 for a more detailed exposition).

Multivariate IV Regression

In order to estimate the parameter vector θ of the heterogeneous peer effects network model we propose a multivariate IV approach. By stacking the single network representations (5) into a hyper-system, we obtain:

$$\mathbf{Y} = \mathbf{X} \, \theta + \boldsymbol{\varepsilon} \tag{6}$$

where:

$$\mathbf{Y} = \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_L \end{pmatrix} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \\ \vdots \\ \tilde{\varepsilon}_L \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_L \end{pmatrix} .$$

Moreover, define $\mathbf{Z} = \mathrm{diag}[\mathbf{Z}_l]$ as the $\left((\sum_l n_l) \times qL\right)$-dimensional block-diagonal instrument matrix with $\mathbf{Z}_l$ as the $n_l \times q \geq k$-dimensional network-specific instrument matrix.

Then IV estimation of (6) with instrument matrix $\mathbf{Z}$ yields the well-known form:

$$\begin{aligned}\hat{\theta}_{MVIV} &= \left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \\ &= \left(\sum_l \mathbf{X}_l'\mathbf{Z}_l(\mathbf{Z}_l'\mathbf{Z}_l)^{-1}\mathbf{Z}_l'\mathbf{X}_l\right)^{-1}\sum_l \mathbf{X}_l'\mathbf{Z}_l(\mathbf{Z}_l'\mathbf{Z}_l)^{-1}\mathbf{Z}_l'\tilde{y}_l\,,\end{aligned} \tag{7}$$

where the second equality arises from the block-diagonality of the projection matrix $\mathbf{P} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$, such that the IV estimator is given by sums of single network terms. For iid random draws of networks $\{y_l, X_l, Z_l, A_l\}_{l=1}^{L}$ and further regularity conditions given in Assumption A.1 we show in Proposition A.1 of Appendix A the consistency of the MVIV estimator for large L:

$$\operatorname*{plim}_{L\to\infty} \hat{\theta}_{MVIV} = \theta.$$

For conditionally heteroskedastic errors at the network level, the asymptotic distribution given by:

$$\begin{aligned}&\sqrt{L}\left(\hat{\theta}_{MVIV} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \Lambda^{-1}\Phi\Lambda^{-1}\right)\,, \\ &\text{with} \qquad \Phi = \mathrm{E}\left[\mathbf{Z}_l'\tilde{\varepsilon}_l\tilde{\varepsilon}_l'\mathbf{Z}_l\right] \\ &\qquad\qquad\quad \Lambda = \mathrm{E}\left[\mathbf{X}_l'\mathbf{Z}_l\right]\mathrm{E}\left[\mathbf{Z}_l'\mathbf{Z}_l\right]^{-1}\mathrm{E}\left[\mathbf{Z}_l'\mathbf{X}_l\right]\,,\end{aligned}$$

as outlined in Proposition A.2 and proven in Appendix A.

Note that the large L setting circumvents the problem of weak identification because there are repeated observations of the correlation between instruments for a given sample size. Therefore, the question of a correlation between instruments and endogenous variables, which asymptotically vanishes with increasing n_l, does not arise.

Instrumental Variable Minimum Distance Estimation

From a practical point of view, it is interesting to note that the MVIV estimator also has a Minimum Distance representation, which can easily be implemented in applied research. This is based on the first-stage IV regression:

$$\tilde{y}_l \quad = \quad \mathbf{W}_l \, \pi_l + \tilde{\varepsilon}_l \, , \qquad (l = 1, \ldots, L) \tag{8}$$

where $\pi_l = \left(\gamma_l', \gamma_{g,l}', \beta_{a,l}, \beta_{g,l}\right)'$ is the first-stage ("reduced form") parameter vector. This equations differs from (3) only by the parameters $\gamma_l, \gamma_{g,l}$. As stated earlier, the $n_l \times k_w$ dimensional regressor matrix $\mathbf{W}_l = \left[\tilde{X}_l \; \tilde{X}_l^g \; \tilde{y}_l^a \; \tilde{y}_l^g\right]$ has full column rank.

The restriction between the first-stage reduced form parameter vector π_l and the structural form parameter vector θ is given by

$$\pi_l(\theta) = M_l \, \theta \, , \tag{9}$$

with

$$M_l = \begin{bmatrix} I_{k_x} & 0 & 0 & 0 & 0 & 0 \\ 0 & I_{k_x} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & v_l' & 0 \\ 0 & 0 & 0 & 1 & 0 & v_l' \end{bmatrix} .$$

Stacking the restrictions between the L reduced form parameter vectors π_l and the structural form parameter θ given by (9) into a $(k_w L \times 1)$-vector containing the complete set of restrictions gives:

$$\pi(\theta) = M\theta \, , \tag{10}$$

with $\pi(\theta) = (\pi_1(\theta)', \pi_2(\theta)', \ldots, \pi_L(\theta)')'$ and $M = (M_1', M_2', \ldots, M_L')'$.

In the second step we estimate θ by minimizing the weighted quadratic distance between the estimated reduced form parameter vector $\hat{\pi}(\theta) = \left(\hat{\pi}_1', \hat{\pi}_2', \ldots, \hat{\pi}_L'\right)'$ and $M\theta$ with respect to the structural parameter vector θ. Using the inverse of any consistently estimated variance of $\hat{\pi}(\theta)$, the minimization problem becomes:

$$\hat{\theta}_{MD} \equiv \arg\min_{\theta} \left[\hat{\pi} - M\theta\right]' \hat{\Omega}^{-1} \left[\hat{\pi} - M\theta\right],$$

where $\hat{\Omega}$ is a consistent estimator of $\mathrm{V}\left[\hat{\pi}(\theta)\right]$. Because of the independence of the networks, the estimated weighting matrix $\hat{\Omega}^{-1} = \hat{\mathrm{V}}\left[\hat{\pi}\right]^{-1} = \mathrm{diag}[\hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1}]$ takes the form of a block diagonal matrix with the inverses of the estimated variance-covariance matrices from the first-stage estimates building the blocks. Moreover, due to the linearity of the restriction between π and θ, the IVMD can be represented as a generalized least square estimator of a regression of $\hat{\pi}$ on M:

$$\hat{\theta}_{IVMD} = \left(M'\hat{\Omega}^{-1}M\right)^{-1} M'\hat{\Omega}^{-1}\hat{\pi} = \left(\sum_{l=1}^{L} M_l' \hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1} M_l\right)^{-1} \left(\sum_{l=1}^{L} M_l' \hat{\mathrm{V}}\left[\hat{\pi}_l\right]^{-1} \hat{\pi}_l\right). \qquad (11)$$

The minimum distance approach presented here differs from classical MD settings in the way asymptotic convergence is achieved. In conventional settings, consistency and asymptotic normality results from the first-stage estimates by assuming $n_l \to \infty \quad \forall\, l = 1, \ldots, L$. In contrast, for the MDE approach described above, the asymptotic results are derived by exploiting the independence across networks and the information on the link between $\hat{\pi}_l$ and θ as the number of networks L tends to infinity.

In applied work the MDE approach is attractive since the first-stage estimates can obtained by standard software packages. The IV estimation of π_l in the first-stage also allows for a straightforward implementation of different sets of instruments for each network. This is desirable in the case of isolated individuals when additional moments from the isolated individuals are included in the first-stage estimates. Importantly, this variation in instruments for different networks does not impact the second-stage estimation. Moreover, the MDE approach can provide additional insights into each individual network as a byproduct. In other words, the minimum distance approach can yield valuable information about the specific characteristics of each network under study.

3 Data

Our empirical study is based on the data of the *Gymnasiasten-Studie* (CAESR 2007), a longitudinal survey of 3,385 10th grade students attending upper secondary school (*Gymnasium*) in the German federal state of North Rhine-Westphalia (NRW) in the school year 1969/1970.[3] Initial data contains 121 classes from 68 randomly selected upper secondary schools. The initial survey of the students provides information on their previous school grades as well as individual characteristics such as gender and age. Besides the student survey, a standard psychometric Intelligence Structure Test (IST) was administered in the classroom during the initial data collection period. About ten years after the original survey, the students' grades were collected from the school archives. Central to our study is the network information collected in the Sociometric Test of the *Gymnasiasten Studie* at the time of the initial survey. To construct the adjacency matrices A_l and G_l for each class, we use information about every student's assessment of whom he or she liked in the class based on the question:[4]

[3]Even though it is a longitudinal survey, the classroom network is only measured at the initial survey. There were three follow-up surveys, each approximately 10 years appart. Unfortunately, the follow-up surveys suffer from substantial panel attrition.

[4]The original question in German is "*In jeder Klasse gibt es Mitschüler, die man sympathisch findet und die man mehr als andere in der Klasse gut leiden kann. Einige findet man sicher recht unsympathisch, und das ist auch ganz normal so. Würden sie nun zunächst einmal die Schüler nennen, die Sie persönlich gut leiden können.*"

> "In every class there are fellow students who one likes more than others in the class. Some others one finds quite unpleasant, and that is quite normal. Kindly first list the students who you personally like a lot."

Most of the empirical papers studying network peer effects in education use the National Longitudinal Study of Adolescent to Adult Health (ADD Health) data. Our unique network data differs from the Add Health data in several ways. One difference is that the students in our dataset nominate their friends within the classroom, while in Add health data nominations are made within the schools. At the time the survey, 10th graders attending German upper secondary schools were taught at the class unit with a rather stable class composition over several years. Therefore, using nominations at the class level seems to be more appropriate. More important, in the Add Health survey, the respondents were asked to name up to ten (five female and five male) best friends. This might raise a truncation problem that does not occur with our dataset. In fact, Griffith (2021) shows that peer effects with censored data are under-estimated.

We construct our sample by merging information from three different sources: student surveys (Hummell et al. 1970), administrative data from school archives (Meulemann et al. 1986), and the sociometric test (Hummell et al. 2018). Among the 68 schools that initially participated in the survey, 6 did not agree to data collection due to privacy concerns. The remaining 62 schools accounted for 91% of the initial sample of students (3,010 students). Due to incomplete information in the school archives, it was only possible to collect the information on the grades of approximately 2,700 students. Lastly, we delete an observation when any of the variables used in the empirical model were missing, and we also delete the isolated individuals, i.e., those who did not name anyone as a friend. We lose around 300 observations after this step. This leaves a sample of 2,385 students and 101 classes.

In network data, removing individuals due to missing information might have different consequences then in a cross-sectional data. Specifically, it might lead to mismeasurement in the network if a friendship link is affected by this. Following the identification approach by Bramoullé et al. (2009), an implicit assumption in the literature, which we also adopt here, is that the links among individuals are perfectly observed. It is still not much known about the econometric implications of measurement errors in networks, however, Lewbel et al. (2021) show in a recent paper that relatively small amounts of measurement error in the network can be safely ignored in estimation. In particular, they show for the instrumental variable estimators like Bramoullé et al. (2009) and their standard errors, remain consistent and valid, as long as the number and size of measurement errors in an observed adjacency matrix is relatively small. Although they do not consider a hybrid model with local-aggregate and loca-average effect together, they show for both cases separately that the usual asymptotics provide a good approximation for inference. We believe that in our case the measurement errors created in the adjacency matrix by deleting individuals due to missing observations are small enough to be ignored safely as they suggested. After cleaning the data, we excluded classes with fewer than 18 students, because the first-stage estimates of a small network suffer from low degrees of freedom. Excluding small classes results in 2,165 students in 85 classes.[5] Table 1 contains the summary statistics of the variables used in our empirical analysis and some network statistics. In the left panel of the table, we present the variables before applying the class size restriction. One can see that the sample means are not substantially affected by the class size restrictions.

Academic performance is measured by the average of the final grades (GPA) for all compulsory and elective courses at the end of the school year 1969/70. As mentioned before, we use the administrative data collected from the school archives to construct the GPA. At the time of the survey, the choice of different courses within a class was very limited, i.e., all students of the same class had to take the same courses. Selection of certain specializations (e.g., languages, mathematics and sciences, humanistic secondary

[5] We estimated the model using different thresholds as a robustness check, but in general the results did not change qualitatively.

Table 1: Summary Statistics

	Entire Sample		Estimation Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Outcome Variables				
GPA[a]	3.20	0.48	3.19	0.48
German	3.46	0.76	3.45	0.76
Math	3.51	0.96	3.51	0.95
Individual characteristics				
IQ	40.20	9.11	40.03	9.14
Previous GPA[a]	3.19	0.49	3.19	0.49
Age	15.38	0.87	15.37	0.87
Network measures				
Class size	27.82	6.01	29.19	4.92
Relative Class size[b]	0.71	0.15	0.75	0.13
Sample Class size[c]	23.61	5.98	25.47	4.23
Female share	0.45	0.43	0.49	0.43
Sample female share[c]	0.45	0.43	0.49	0.44
Density	0.24	0.06	0.22	0.04
Clustering	0.03	0.02	0.02	0.01
N	2,385		2,165	
L	101		85	

Note: Own calculations. We exclude classes with fewer than 18 students in the estimation sample. a: Better grades are represented by lower values. b: relative class size represents the class size divided by the size of the largest class. c: Sample class size refers to the number of remaining students in the classroom after deleting the individuals with missing information either in the survey data, administrative data, or in the sociometric test. The same principle applies to the calculation of the female share measure.

school) took place with the choice of the specific secondary school. Therefore, the GPA within a given network is based on mostly the same subjects. The grades are measured in terms of the German grading system: with 1 ('very good') being the best grade and 6 ('insufficient') as the worst grade. In addition to the overall GPA, we also closely examine scores in Mathematics and German to detect potential differences in peer behavior across subjects.

Individual heterogeneity in our model is captured by the student's IQ score, the GPA from the previous school year, and the student's age. The IQ is constructed from the correctly solved questions of the IST. We control for the GPA from previous year as a proxy for the overall school performance at the beginning of the survey year. In order to account for network heterogeneity in the peer effects of the local-average and local-aggregate models, we allow the two peer effect parameters $\beta_{a,l}$ and $\beta_{g,l}$ in Equation (2) to depend on class-specific factors. As such factors we use the relative class size, i.e., the size of the network relative to the largest class size, and the fraction of girls in the class. As discussed in detail in Section 1, the literature on the effects of the class size and gender (ratio) on school outcomes is very rich. In general, the main consideration is the direct causal link from class size to the outcome. It seems, however, reasonable to look for a potential indirect link through heterogeneous peer effects. In a study, Lin (2014) estimates peer effects separately for large and small classes using the Add Health dataset and finds significant differences between the peer effects in the two groups. She also conducted similar analyses of various network attributes, including the gender ratio. Surprisingly, she does not find a significant difference between the peer effects of the two subsamples by gender proportion. However, experimental evidence about gender diversity and performance shows that team collaboration is greatly improved by the presence of women in the group (see, for example, Bear & Woolley 2011, and references therein).

The network density is defined as the ratio of all connections in a network to the number of potential connections. Thus, the denser a network is, the closer its density is to unity. In our sample, the density of the networks varies between 0.14 and 0.37, with 45 classes having a lower density than the mean. Clustering, on the other hand, measures all transitive triads relative to the total number of triads. It is a measure of the probability that two peers of an individual nominate each other. In our dataset this measure varies between 0.004 and 0.08, indicating rather sparse friendship networks at the class level.

To obtain a better understanding of the network structure and its potential role for peer effects, we take a look at the summary statistics of the friends' nominations. The average number of friends a student named (outdegree) is 5.48, indicating that students give considerable thought to nominating their friends. Figure 1 depicts the distribution of friends. The distribution of outdegrees and indegrees is in the line with the fact that the networks are sparse. Most of the students name around five classmates that they like, and very few name more than ten peers.

Our motivation for considering potential heterogeneity in peer effects results from the observation that the network structures and characteristics vary substantially across networks. We hypothesize that network differences might affect how peer effects operate. Using network graphs, we illustrate how school classes (networks) vary in terms of individual performance, class size, gender, and network structure. The size of a node is proportional to the outdegree, its color indicates the GPA score (lighter colors represent better performance), and the shape of the node indicates the gender. Since plotting all networks together for visual inspection would give too small a picture to be detected, we concentrate on four specific classes.

First, we plot the largest and smallest classrooms in Figure 2. Second, Figure 3 depicts the two networks with the highest and lowest densities in the sample. Without the intention of stressing the following argument too much, a comparison of the largest with the smallest network in Figure 2 illustrates that the performance of students might depend on their degree of connectivity and the class size. For both networks, the better performing students are slightly more central (being named as friends more often), while particularly in the larger network, the less well-connected students are also associated with lower performance.

Figure 3 provides further exploratory evidence that the network structure varies across classrooms. The network on the right, with the least density, reveals two major clusters. In contrast, the network on the left, with higher density, is centered around a single cluster of students. In the densest network we find several very popular students, who have average grades. The least dense classroom corresponds to a boys class, where there are no students with good grades and the most central ones are again average students. Comparing the network graphs for different subjects we hardly find any differences. The four classroom networks depicted appear very similar if the color of the nodes is based not on the GPA scores but on the grades in Math and German.

4 Empirical Results

The primary specification in our study is the heterogeneous composite model given by (2), which incorporates the local-aggregate and the local-average peer effects model as nested specifications. Although most of the empirical studies focus on peer effects resulting from norm behavior, and therefore favor the local-average model, ex-ante, both hypotheses on how peers affect individual educational achievement are reasonable. In fact, the two effects may complement or even counteract each other. As mentioned above, our main outcome variable of interest is the GPA. However, since peer effects may operate differently depending on the subject, we also study the peer effects for Math and German (see Tables A1 and A2 in the Appendix). As predetermined or exogenous explanatory variables we use the GPA of the previous year, IQ, and age, as well as their counterparts for the student's peer group. As network-specific characteristics we use the relative class size and the fraction of girls.

As it has been pointed out in Section 2 the exogeneity assumption on the friendship links conditional on observed characteristics and network fixed effects might fail if there are unobserved factors affecting both the link formation and the outcome variable. We believe that by including previous GPA, IQ and age we control for the most important factors (or proxies for such factors) which might affect the current grades, as well as forming friendship links. In particular, we believe that previous GPA captures overall school performance at the beginning of 10th grade, IQ captures cognitive skills, while age proxies certain age specific skills.

Table 2 summarizes the estimation results for the composite, local-aggregate, and local-average models with heterogeneous peer effects based on the IVMD approach with globally differenced variables. To facilitate the interpretation of the estimation results, we centered the network-specific characteristics around their means. As a result, the two intercept terms in (4) reflect the aggregate and average peer effects for a class with mean characteristics.

First and most importantly, our estimation results reveal that taking into account heterogeneity in peer effects turns out to be absolutely crucial. We find clear evidence that both peer effects significantly differ by class size and gender composition. Several studies that concentrate on homogeneous peer effects fail to provide sufficient statistical evidence for the existence of peer effects (e.g., Boucher et al. 2014, Liu et al. 2014). Like these studies, we also find insignificant estimates for our sample when network heterogeneity is ignored (see Table A3 in the Appendix).

Second, the way peers affect a student's performance also matters, as both mechanisms, the local-aggregate and the local-average one, have a positive impact on a student's educational attainment in a representative classroom with average relative size and gender composition. Comparing the Wald statistics in the last row of Table 2, we observe that the heterogeneous local models are rejected in favor of the heterogeneous composite model.

In the composite model both intercepts turn out to be positive at the 1% significance level. This means that if the peers perform better individually or on average, then so does the individual. It is important to note that the coefficients of the two local models are not directly comparable in magnitude. The peer effects due to local-aggregate behavior are proportional to the number of peers (outdegree of student), i.e., the larger the student's peer group, the stronger that student's performance is affected by their peers. Therefore, we compute the impact of a one-unit change in the GPA of the peers on the individual's GPA, for both models. For a class with average characteristics the local-aggregate effect would exceed the local-average effect if the student has more than 52 peers. Noting that the median outdegree in our sample is 5 (see Figure 1) we can conclude that the local-average peer effect generally dominates the local-aggregate peer effect.

Interestingly, the coefficients on the female share variable in local-aggregate and local-average part operate in different directions. The gender effect is large and positive for the local-average part but small and negative for the local-aggregate part. By interpreting the local-average effect as a proxy for norm behavior, we can conclude from our estimates that norm behavior is significantly more prevalent in girls-only classes compared to boys-only classes. To illustrate the sizes of the effects of the two components, consider two classes of average relative size – one being a girls-only class while the other class is a boys-only class. In this case, a student with 5 friends in a girls-only class experiences an overall peer effect of 0.1841 ($= 5 \times (0.0021 - 0.0027 \times (1 - 0.49)) + (0.1099 + 0.1386 \times (1 - 0.49))$), while the corresponding effect for a student with the same number of friends in a boys-only class is 0.0590.

The effect of class size is negative and significant for both peer effect mechanisms, implying that for larger classrooms, the enhancing contribution of peer behavior diminishes. Assuming an average female share, the peer effect of the smallest class size is 0.2223, while for the largest class size the overall peer effect is 0.0142. It is important to emphasize that this class size effect is novel in the literature, as it operates through peer behavior. It operates in addition to a potential direct effect of class size on a student's performance, which is traditionally considered in studies on the determinants of educational achievement.

Because of the partialling out network-specific level effects the effect of class size on peer behavior can be seen as a second channel for the impact of class size on educational achievement. Unlike the conventional direct effect of class size obtained from reduced form specifications, the effect of class size through peer behavior indicates the role of social interactions in a classroom, which in turn partially determines the individual performance. In this sense, our approach also offers a specific structural explanation for the variations in

Table 2: IVMD Estimation Results for GPA

	Heterogeneous Peer Effects Model		
	Composite	Local-aggregate	Local-average
Local-aggregate peer effect			
Intercept	0.0021***	0.0012***	
	(0.0003)	(0.0003)	
Relative Class Size	-0.0094***	-0.0109***	
	(0.0025)	(0.0025)	
Female Share	-0.0027***	-0.0010	
	(0.0008)	(0.0008)	
Local-average peer effect			
Intercept	0.1099***		0.1526***
	(0.0262)		(0.0293)
Relative Class Size	-0.3768***		-0.3252***
	(0.1096)		(0.1229)
Female Share	0.1386***		0.1432***
	(0.0326)		(0.0353)
Own characteristics			
IQ	-0.0033***	-0.0031***	-0.0027***
	(0.0004)	(0.0004)	(0.0004)e
Previous GPA	0.7344***	0.7401***	0.7317***
	(0.0071)	(0.0071)	(0.0076)
Age	-0.0049	-0.0036	-0.0023
	(0.0038)	(0.0039)	(0.0043)
Peers' characteristics			
IQ	0.0002	-0.0003	0.0001
	(0.0008)	(0.0008)	(0.0008)
Previous GPA	-0.0249	0.084***	-0.0465*
	(0.0246)	(0.0144)	(0.0274)
Age	-0.0203***	-0.0148*	-0.023**
	(0.0076)	(0.0081)	(0.0089)
Wald statistics (d.f.)		66.32 (3)	50.83 (3)

Estimates of the three model variants obtained by IVMD estimation. The first column corresponds to the composite model, the second column corresponds to the local-aggregate model and the third column corresponds to the local-average model. Following Liu et al. (2014), we use the average characteristics of the second order peers as instruments for the local-average part, and the aggregate characteristics of the peers for the local-aggregate part. The Wald statistics test the composite model against the local models. Robust standard errors are reported in parentheses. First-stage errors are assumed to be heteroskedastic, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$, N=2165, L=85.

class performance.

Figure 4 depicts the size of the combined peer effect over different class sizes and gender compositions. For larger classes with a low fraction of female students, the peer effect is negative. Classes with a mean female composition have a negative overall peer effect if the class size is larger than 34. All in all, the combined peer effect ranges from -0.12 to 0.23.

The coefficients on own IQ and own previous GPA have the expected signs. Not very surprisingly, the GPA of the previous year is a strong predictor of current performance. Students with a higher IQ also perform better. Our results do not suggest a significant impact of age. For the exogenous peer effects, we observe that having smarter or less smart peers does not have an impact on the individual outcome. Neither does the previous GPA of the peers. The results show that having older peers helps to have better grades.

Columns 2 and 3 in Table 2 summarize the results from the heterogeneous local models. The impact of the female share in the local-aggregate model is no longer significant but has the same sign as in the composite model and is similar in magnitude. In the local-aggregate model, we see that having peers with better grades has a negative influence on the individual outcome. Other estimates are similar to those for the composite model.

The IVMD estimates of the heterogeneous local and composite models with grades in Math and German as dependent variables are given in Tables A1 and A2 in the Appendix. For these two dependent variables we also find heterogeneity in peer effects in terms of the class characteristics and the transmission mechanism. With a few exceptions, these findings for the two subjects are consistent with the findings for the overall GPA score. However, a notable exception is the significant and negative coefficient on female share for the local-average effect for Math, which indicates that the role of gender composition in educational outcomes must be discussed in the light of the field of study. Similarly, the positive coefficient of class size for Math and German suggests that the role of class size on peer behavior also depends on the field of study.

Finally, we also use the Wald test to test the heterogeneous composite model against its nested specifications. Comparing the Wald statistics in the last row of Table A3 between the nested homogeneous specifications and their counterparts in the heterogeneous specifications reveals that the null hypothesis of homogeneous peer effects has to be rejected against the heterogeneous peer effects models for all three outcome variables. These findings hold for the heterogeneous composite model as well as for the heterogeneous local models.

5 Conclusions

This paper contributes to the growing literature on the empirical analysis of social networks. In particular, we focus on the role of heterogeneity in network peer effects by accounting for network-specific factors and different driving mechanisms of peer behavior. For our empirical study of the role of network peer effects on educational attainment, we use a unique network dataset of 85 school classes of secondary schools in Germany, which allows us to exploit exogenous variation in second degree friends to identify the endogenous peer effects.

For the estimation of peer effects in a set of networks we introduce a Multivariate Instrumental Variable approach and its easily implementable Minimum Distance representation which proves to be a valuable tool to study parametrically rich network environments. The asymptotic properties of the estimator are based on a large number of networks while keeping the size of the individual networks bounded.

As network-specific factors, we find that the size of the network (i.e., of a school class) and gender composition are important determinants of the peer effects, while conventional model specifications with homogeneous peer effects turn out to be too crude and lead to insignificant findings. In addition to the network-specific factors, heterogeneity in terms of the underlying behavioral assumptions matters. In particular, we can show that a student's educational attainment is affected by both the pure size of their peer group, as reflected by the local-aggregate model, and the norm behavior captured by the local-average model.

Our study contributes to the voluminous empirical literature on the determinants of educational attainment. We show that an increase in the class size alone reduces peer behavior and may even lead to negative peer effects in very large classes. Unlike many empirical studies in this field, which are largely based on reduced form approaches, our approach gives rise to a structural interpretation of why class size and gender composition matter and why these factors differ across subjects. For instance, our study sheds light on peer behavior as a specific channel through which class size affects educational attainment.

We regard our study as a promising starting point for more realistic modeling of heterogeneous network behavior and for a deeper understanding of how networks operate. Future work should be devoted to more elaborate specifications of network heterogeneity (e.g., nonlinear or nonparametric peer effects) as well as to the analysis of the relationship between network structures (e.g., properties of the adjacency matrices) and the identification of network peer effects. In particular, a promising extension of our approach would be to treat the heterogeneous peer effects within the correlated random coefficient framework.

References

Angrist, J. & Lavy, V. (1999), 'Using Maimonides' rule to estimate the effect of class size on scholastic achievement', *Quarterly Journal of Economics* 114(2), 533–575.

Arduini, T., Patacchini, E. & Rainone, E. (2020), 'Treatment effects with heterogeneous externalities', *Journal of Business & Economic Statistics* 38(4), 826–838.

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

Bear, J. B. & Woolley, A. W. (2011), 'The role of gender in team collaboration and performance', *Interdisciplinary Science Reviews* 36(2), 146–153.

Beugnot, J., Fortin, B., Lacroix, G. & Villeval, M. C. (2019), 'Gender and peer effects on performance in social networks', *European Economic Review* 113, 207–224.

Blume, L. E., Brock, W. A., Durlauf, S. N. & Jayaraman, R. (2015), 'Linear social interactions models', *Journal of Political Economy* 123(2), 444–496.

Boucher, V., Bramoullé, Y., Djebbari, H. & Fortin, B. (2014), 'Do peers affect student achievement? Evidence from Canada using group size variation', *Journal of Applied Econometrics* 29(1), 91–109.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* 150(1), 41–55.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2020), 'Peer effects in networks: A survey', *Annual Review of Economics* 12(1), 603–629.

CAESR (2007), *Dataset Gymnasiastenstudie*, Central Archive for Empirical Social Research, Cologne, Germany.

Calvó-Armengol, A., Patacchini, E. & Zenou, Y. (2009), 'Peer effects and social networks in education', *The Review of Economic Studies* 76(4), 1239–1267.

Dobbelsteen, S., Levin, J. & Oosterbeek, H. (2002), 'The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition', *Oxford Bulletin of Economics and Statistics* 64(1), 17–38.

Fredriksson, P., Öckert, B. & Oosterbeek, H. (2012), ' Long-term effects of class size', *The Quarterly Journal of Economics* 128(1), 249–285.

Griffith, A. (2021), 'Name your friends, but only five? the importance of censoring in peer effects estimates using social network data', *Journal of Labor Economics* 40(4), 779–805.

Hanushek, E. A. (1996), 'Measuring investment in education', *Journal of Economic Perspectives* 10(4), 9–30.

Heinesen, E. (2010), 'Estimating class-size effects using within-school variation in subject-specific classes', *The Economic Journal* 120(545), 737–760.

Hoxby, C. M. (2000), 'The effects of class size on student achievement: New evidence from population variation', *The Quarterly Journal of Economics* 115(4), 1239–1285.

Hoxby, C. M. (2002), 'The power of peers: How does the makeup of a classroom influence achievement', *Education Next* Summer 2(2.2), 57–63.

Hummell, H. J., Klein, M., Wieken-Mayser, M. & Ziegler, R. (1970), 'Structure analysis of the school (schoolchildren survey)', GESIS Data Archive, Cologne. ZA0600 Data file Version 1.0.0, https://doi.org/10.4232/1.0600.

Hummell, H. J., Klein, M., Wieken-Mayser, M. & Ziegler, R. (2018), 'Structure analysis of the school (schoolchildren survey - sociometric test)', GESIS Data Archive, Cologne. ZA0942 Data file Version 2.0.0, https://doi.org/10.4232/1.13036.

Johnsson, I. & Moon, H. R. (2021), 'Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach', *The Review of Economics and Statistics* 103(2), 328–345.

Kelejian, H. H. & Prucha, I. R. (1998), 'A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *The Journal of Real Estate Finance and Economics* 17(1), 99–121.

Kelejian, H. H. & Prucha, I. R. (1999), 'A generalized moments estimator for the autoregressive parameter in a spatial model', *International Economic Review* 40(2), 509–533.

Krueger, A. B. (1999), 'Experimental estimates of education production functions', *The Quarterly Journal of Economics* 114(2), 497–532.

Lavy, V. & Schlosser, A. (2011), 'Mechanisms and impacts of gender peer effects at school', *American Economic Journal: Applied Economics* 3(2), 1–33.

Lewbel, A., Qu, X. & Tang, X. (2021), Social networks with mismeasured links, Boston College Working Papers in Economics 1031, Boston College Department of Economics.

Lin, X. (2014), 'Network attributes and peer effects', *Economics Bulletin* 34(3), 2060–2079.

Liu, X., Patacchini, E. & Zenou, Y. (2014), 'Endogenous peer effects: Local aggregate or local average?', *Journal of Economic Behavior and Organization* 103, 39–59.

Mael, F., Alonso, A., Gibson, D., Rogers, K. & Smith, M. (2005), Single-sex versus coeducational schooling: A systematic review. doc# 2005-01., Technical report.

Manski, C. F. (1993), 'Identification of endogenous social effects: The reflection problem', *The Review of Economic Studies* 60(3), 531–542.

Masten, M. A. (2018), 'Random coefficients on endogenous variables in simultaneous equations models', *The Review of Economic Studies* 85(2), 1193–1250.

Meulemann, H., Wieken-Mayser, M. & Wiese, W. (1986), 'Social origins and school career of high school students', GESIS Data Archive, Cologne. ZA1440 Data file Version 1.0.0, https://doi.org/10.4232/1.1440.

Morse, S. (1999), 'Separated by sex: A critical look at single-sex education for girls', *Journal of Chemical Education* 76, 615.

A Asymptotic Results

For ease of exposition the results presented in this appendix are based on the local-average model with heterogeneous peer effects ($\beta_{a,l} = 0, \beta_{g,l} = \beta_l$). To simplify the notation, we assume that the correlated effect η_l is zero or has been partialled out. Moreover, we concentrate on one exogenous factor ($k_x = 1$) and one factor driving the peer effects parameter $k_v = 1$. Thus, the local-average model in vector notation is given by:

$$y_l = \gamma x_l + \gamma_g x_l^g + \beta_l \, y_l^g + \varepsilon_l \, , \qquad (l = 1, \ldots, L) \tag{12}$$

where y_l is the $n_l \times 1$ vector of outcome variables, G_l the row normalized adjacency matrix, x_l the exogenous factor, $x_l^g = G_l x_l$ the exogenous factor of the peers. The local-average outcome of the peers is $y_l^g = G_l y_l$. For the corresponding reduced form expression we have:

$$y_l = \big(I_{n_l} - \beta_l G_l\big)^{-1} \big(\gamma I_{n_l} + \gamma_g G_l\big) x_l + \varepsilon_l \, .$$

If $\gamma_g + \gamma\beta_l \neq 0$ and I_{n_l}, G_l, G_l^2 and G_l^3 are linearly independent, the parameters are identified (see Bramoullé et al. 2009).

A.1 (Non-) Identification of a single network l

Assuming that there are no isolated individuals we get for the expected outcome of the peers:

$$\mathrm{E}[y_l^g | \, x_l, G_l] = \gamma G_l x_l + (\gamma_g + \gamma\beta_l)(G_l^2 + \beta_l G_l^3 + \ldots)$$

The IV estimator of parameters in (12) is consistent and asymptotically normal for $n_l \to \infty$. Under this assumption Minimum Distance estimation of the heterogeneous peers effects model satisfies the classical assumptions of MDE even if the number of networks is fixed.

Assume now that the heterogeneous peer effects parameter β_l depends linearly on the observable factor v_l: $\beta_l = \beta_0 + \beta_1 v_l$. Even without the presence of endogeneity the $k \times 1$-dimensional overall parameter vector $\theta = (\gamma, \gamma_g, \beta_0, \beta_1)'$ of the structural equation

$$y_l = \gamma x_l + \gamma_g x_l^g + \beta_0 y_l^p + \beta_1 v_l y_l^p + \varepsilon_l = \mathbf{X}_l\, \theta + \varepsilon_l \tag{13}$$

are not identified on the network level since the y_l^p and $v_l y_l^p$ are proportional such that the $n_l \times k$ dimensional regressor matrix $\mathbf{X}_l = \begin{bmatrix} x_l & x_l^g & y_l^g & v_l y_l^g \end{bmatrix}$ has no full column rank.

A.2 Consistency

Assumption A.1 (System of Peer Effects Networks)

Consider the model for the l-th network as defined in (13) and assume that the following conditions hold:

(i) $\{y_l, \mathbf{X}_l, \mathbf{Z}_l\}_{l=1}^{L}$ is a sequence of iid random matrices with finite moments, where

- $\mathbf{Z}_l = \begin{bmatrix} \mathbf{z}_{1,l} & \mathbf{z}_{2,l} \ldots \mathbf{z}_{n_l,l} \end{bmatrix}'$ with $\mathbf{z}_{i,l}$ a $q \times 1$ vector of instrumental variables
- $\mathbf{X}_l = \begin{bmatrix} \mathbf{x}_{1,l} & \mathbf{x}_{2,l} \ldots \mathbf{x}_{n_l,l} \end{bmatrix}'$ with $\mathbf{x}_{i,l}$ a $k \times 1$ vector of explanatory variables

for network l with $q \geq k$ and

(ii) $\mathrm{E}[\varepsilon_{i,l} | \, \mathbf{z}_{i,l}] = 0$

(iii) $\mathrm{V}\left[\varepsilon_{i,l} | \, \mathbf{z}_{i,l}\right] = \sigma^2$

(iv) $\mathrm{E}\left[\mathbf{z}_{i,l}\, \mathbf{x}_{i,l}'\right]$ is a finite $q \times k$ - matrix

(v) $\mathrm{E}\left[\mathbf{z}_{i,l}\, \mathbf{z}_{i,l}'\right]$ is a finite, non-singular $q \times q$ - matrix

(vi) $\mathrm{E}\left[\mathbf{x}_{i,l}\, \mathbf{x}_{i,l}'\right]$ is a finite $k \times k$ - matrix

(vii) The size of network l is a bounded random variable with $0 < n_{min} \leq n_l \leq n_{max}$ and independent of the network variables

Proposition A.1 (Large L consistency)

Given the assumptions A.1 hold, the multivariate IV estimator for heterogeneous networks $\hat{\theta}_{MVIV}$ given by (7) is consistent for large L and random network sizes:

$$\operatorname*{plim}_{L\to\infty} \hat{\theta}_{MVIV} = \theta$$

Proof A.1 (Large L consistency)

Decompose the MVIV estimator as follows:

$$\hat{\theta}_{MVIV} = \theta + \big(\frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\mathbf{X}\big)^{-1} \times \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\boldsymbol{\varepsilon}$$

and taking the plim on both sides gives:

$$\begin{aligned}
\operatorname*{plim}_{L\to\infty} \hat{\theta}_{MVIV} &= \theta + \big(\operatorname*{plim}_{L\to\infty} \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\mathbf{X}\big)^{-1} \times \operatorname*{plim}_{L\to\infty} \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\boldsymbol{\varepsilon} \\
&= \theta + \big(\operatorname*{plim}_{L\to\infty} \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\mathbf{X}\big)^{-1} \times 0 \\
&= \theta\,,
\end{aligned}$$

where we made use of Lemma A.1 and Lemma A.2 below to show the convergence of the first plim to a finite invertible matrix and the second plim to a zero vector. □

Lemma A.1 (Weak Convergence of $\frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\mathbf{X}$)

Given the assumptions stated in Assumption A.1 hold, then

$$\operatorname*{plim}_{L\to\infty} \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\mathbf{X} = \mathrm{E}\,[n_l]\,\mathrm{E}\left[\mathbf{x}_{i,l}\,\mathbf{z}'_{i,l}\right]\,\mathrm{E}\left[\mathbf{z}_{i,l}\,\mathbf{z}'_{i,l}\right]^{-1}\mathrm{E}\left[\mathbf{z}_{i,l}\,\mathbf{x}'_{i,l}\right]\ .$$

Proof A.2 (Lemma A.1)

Decompose the probability limit into the three parts:

$$
\begin{aligned}
\operatorname*{plim}_{L\to\infty} \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\mathbf{X} =& \operatorname{plim} \frac{\mathbf{X}'\,\mathbf{Z}}{L} \big(\operatorname{plim} \frac{\mathbf{Z}'\,\mathbf{Z}}{L}\big)^{-1} \operatorname{plim} \frac{\mathbf{Z}'\,\mathbf{X}}{L} \\
=& \operatorname{E}[n_l]\operatorname{E}\left[\mathbf{x}_{i,l}\,\mathbf{z}'_{i,l}\right]\operatorname{E}\left[\mathbf{z}_{i,l}\,\mathbf{z}'_{i,l}\right]^{-1}\operatorname{E}\left[\mathbf{z}_{i,l}\,\mathbf{x}'_{i,l}\right] \ ,
\end{aligned}
$$

where the first and the last (transposed) term arise from:

$$
\begin{aligned}
\operatorname*{plim}_{L\to\infty} \frac{1}{L}\,\mathbf{X}'\,\mathbf{Z} =& \operatorname*{plim}_{L\to\infty} \frac{1}{L} \sum_{l=1}^{L} \mathbf{X}'_l\,\mathbf{Z}_l \\
=& \operatorname{E}[\mathbf{X}'_l\,\mathbf{Z}_l] \\
=& \operatorname{E}\left[\sum_{i=1}^{n_l} \mathbf{x}_{i,l}\,\mathbf{z}'_{i,l}\right] \\
=& \operatorname{E}_{n_l}\operatorname{E}\left[\sum_{i=1}^{n_l} \mathbf{x}_{i,l}\,\mathbf{z}'_{i,l} \,\middle|\, n_l\right] \\
=& \operatorname{E}\left[n_l\operatorname{E}\left[\mathbf{x}_{i,l}\,\mathbf{z}'_{i,l}\right]\right] \\
=& \operatorname{E}[n_l]\operatorname{E}\left[\mathbf{x}_{i,l}\,\mathbf{z}'_{i,l}\right] \ .
\end{aligned}
$$

The second equality arises from weak law of large numbers based on assumption (i), while the fourth and the fifth equality arise from the law of iterated expectations and the independence assumption (vii), respectively.

For the second term the following property holds:

$$
\begin{aligned}
\operatorname*{plim}_{L\to\infty} \frac{1}{L}\,\mathbf{Z}'\,\mathbf{Z} =& \operatorname*{plim}_{L\to\infty} \frac{1}{L}\sum_{l=1}^{L} \mathbf{Z}_l'\,\mathbf{Z}_l \\
=& \,\mathrm{E}\left[\mathbf{Z}_l'\,\mathbf{Z}_l\right] \\
=& \,\mathrm{E}\left[\sum_{i=1}^{n_l} \mathbf{x}_{i,l}\,\mathbf{z}_{i,l}'\right] \\
=& \,\mathrm{E}\left[n_l\right]\mathrm{E}\left[\mathbf{z}_{i,l}\,\mathbf{z}_{i,l}'\right]\ .
\end{aligned}
$$

Invertability of $\mathrm{E}\left[\mathbf{z}_{i,l}\,\mathbf{z}_{i,l}'\right]$ is assumed by (v). □

Lemma A.2 (Weak Convergence of $\frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\boldsymbol{\varepsilon}$)

Given the assumptions stated in Assumption A.1 hold, then

$$
\operatorname*{plim}_{L\to\infty} \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\boldsymbol{\varepsilon} = 0\,.
$$

Proof A.3 (Lemma A.2)

Decompose the probability limit into the three parts:

$$
\begin{aligned}
\operatorname*{plim}_{L\to\infty} \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\boldsymbol{\varepsilon} =& \operatorname{plim} \frac{\mathbf{X}'\,\mathbf{Z}}{L}\big(\operatorname{plim} \frac{\mathbf{Z}'\,\mathbf{Z}}{L}\big)^{-1}\operatorname{plim} \frac{\mathbf{Z}'\,\boldsymbol{\varepsilon}}{L} \\
=& \,\mathrm{E}\left[n_l\right]\mathrm{E}\left[\mathbf{x}_{i,l}\,\mathbf{z}_{i,l}'\right]\mathrm{E}\left[\mathbf{z}_{i,l}\,\mathbf{z}_{i,l}'\right]^{-1} \times 0 \\
=& 0.
\end{aligned}
$$

The convergence proofs of the first two terms are given by Lemma A.1. For the last term we have:

$$
\begin{aligned}
\underset{L\to\infty}{\text{plim}}\,\frac{1}{L}\,\mathbf{Z}'\,\boldsymbol{\varepsilon} =& \underset{L\to\infty}{\text{plim}}\,\frac{1}{L}\sum_{l=1}^{L}\mathbf{Z}_l'\,\varepsilon_l \\
=& \,\mathrm{E}\left[\mathbf{Z}_l'\,\varepsilon_l\right] \\
=& \,\mathrm{E}\left[\sum_{i=1}^{n_l}\mathbf{z}_{i,l}\,\varepsilon_{i,l}\right] \\
=& \mathrm{E}_{n_l}\left[\,\mathrm{E}\left[\sum_{i=1}^{n_l}\mathbf{z}_{i,l}\,\varepsilon_{i,l}\middle|\,n_l\right]\,\right] \\
=& \,\mathrm{E}\left[n_l\,\mathrm{E}\left[\mathbf{z}_{i,l}\,\varepsilon_{i,l}\right]\right] \\
=& 0,
\end{aligned}
$$

where we made use again from the iid assumption across the networks and the validity of the instruments (ii). □

A.3 Asymptotic Normality

Proposition A.2 (Asymptotic Normality)

Assume the Assumptions A.1 hold, but replace Assumption (iii) by the Assumption (iii)' $\mathrm{V}\left[\mathbf{z}_{i,l}\,\varepsilon_{i,l}\right] = \mathrm{E}\left[\varepsilon_{i,l}^2\,\mathbf{z}_{i,l}\,\mathbf{z}_{i,l}'\right]$ is a finite non-singular matrix, then the asymptotic distribution of the multivariate IV estimator under heteroskedasticity is given by:

$$
\begin{aligned}
&\sqrt{L}\left(\hat{\theta}_{MVIV} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \Lambda^{-1}\Phi\Lambda^{-1}\right)\ , \qquad (14) \\
&\text{with} \qquad \Phi = \mathrm{E}\left[\mathbf{Z}_l'\,\varepsilon_l\varepsilon_l'\,\mathbf{Z}_l\right] \\
&\qquad\qquad\;\; \Lambda = \mathrm{E}\left[\mathbf{X}_l'\,\mathbf{Z}_l\right]\mathrm{E}\left[\mathbf{Z}_l'\,\mathbf{Z}_l\right]^{-1}\mathrm{E}\left[\mathbf{Z}_l'\,\mathbf{X}_l\right]\ ,
\end{aligned}
$$

Proof A.4 (Asymptotic Normality)

Scale the stabilizing transformation by L and $\sqrt{L}$ as follows:

$$\sqrt{L}\big(\hat{\theta}_{MVIV} - \theta\big) = \big(\frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\mathbf{X}\,\big)^{-1} \times \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\boldsymbol{\varepsilon} = A_L^{-1} b_L$$

where weak convergence of $A_L \equiv \frac{1}{L}\,\mathbf{X}'\,\mathbf{P}\,\mathbf{X}$ to a finite matrix has already been shown by Lemma A.1. Moreover, for the random vector b_L can be decomposed as:

$$b_L \equiv \big(\frac{1}{L}\,\mathbf{X}'\,\mathbf{Z}\,\big)\big(\frac{1}{L}\,\mathbf{Z}'\,\mathbf{Z}\,\big)^{-1}\frac{1}{\sqrt{L}}\,\mathbf{Z}'\,\boldsymbol{\varepsilon}$$

As the weak law of large number and the central limit theorem hold the last term converges to a normally distributed vector:

$$\frac{1}{\sqrt{L}}\,\mathbf{Z}'\,\boldsymbol{\varepsilon} \xrightarrow{d} \mathcal{N}(0, \Phi)\ ,$$

with $\Phi = \mathrm{E}\,[\mathbf{Z}_l'\,\varepsilon_l \varepsilon_l'\,\mathbf{Z}_l]$. Consequently, we obtain for the limiting distribution of b_L

$$b_L \xrightarrow{d} \mathcal{N}\Big(0, \mathrm{E}\,[\mathbf{X}_l'\,\mathbf{Z}_l]\,\mathrm{E}\,[\mathbf{Z}_l'\,\mathbf{Z}_l]^{-1}\,\Phi\,\mathrm{E}\,[\mathbf{Z}_l'\,\mathbf{Z}_l]^{-1}\,\mathrm{E}\,[\mathbf{Z}_l'\,\mathbf{X}_l]\Big)\ ,$$

The limiting distribution of $A_L^{-1} b_L$ yields (14). □

A.4 Equivalence of MVIV and IVMD estimation

Proposition A.3 (Equivalence of Estimators)

For a homoskedastic design with $\mathrm{V}\,[\varepsilon_{i,l}] = \sigma^2 \quad \forall i, l$ and independent networks the Minimum Distance estimator $\hat{\theta}_{MD}$ given by (11) and the MVIV estimator $\hat{\theta}_{MVIV}$ given by (7) are numerically equivalent.

Proof A.5 (Equivalence of Estimators)

Under homoskedasticity all true optimal weighting matrices Ω_l^{-1} are scaled by the same factor $1/\sigma^2$, so that it does not affect the MD estimator. Thus ignoring the scaling factor the optimal weighting matrix is $\hat{\Omega}_l^{-1} = \mathbf{W}_l' \mathbf{P}_l \mathbf{W}_l$ with $\mathbf{W}_l = \left[X_l \, X_l^g \, y_l^g\right]$.

Define the MD estimator (11) as the product of the two matrices A^{-1} and B such that $\hat{\theta}_{MD}(\hat{\Omega}^{-1}) = A^{-1}B$. For the first term we get:

$$
\begin{aligned}
A \equiv M'\hat{\Omega}^{-1}M = \sum_l M_l'\hat{\Omega}_l^{-1}M_l &= \sum_l M_l' \, \mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{W}_l \, M_l \\
&= \sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{X}_l \, ,
\end{aligned}
$$

where we used $\mathbf{W}_l \, M_l = \mathbf{X}_l$ to obtain the last equality. The second term is given by:

$$
B \equiv M'\hat{\Omega}^{-1}\hat{\pi} = \sum_l M_l'\hat{\Omega}_l^{-1}\hat{\pi}_l = \sum_l M_l' \, \mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{W}_l \, \hat{\pi}_l
$$

Decomposing the first-stage IV estimate $\hat{\pi}_l$ into the true parameter vector $\pi_l = M_l\theta$ and the error term part yields:

$$
\begin{aligned}
B &= \sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{W}_l \left(M_l\theta + \left(\mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{W}_l\right)^{-1} \mathbf{W}_l' \, \mathbf{P}_l \, \varepsilon_l\right) \\
&= \sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{X}_l \, \theta + \sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{W}_l \left(\mathbf{W}_l' \, \mathbf{P}_l \, \mathbf{W}_l\right)^{-1} \mathbf{W}_l' \, \mathbf{P}_l \, \varepsilon_l \\
&= \sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{X}_l \, \theta + \sum_l M_l' \, \mathbf{W}_l' \, \mathbf{P}_l \, \varepsilon_l \\
&= \sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{X}_l \, \theta + \sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \varepsilon_l
\end{aligned}
$$

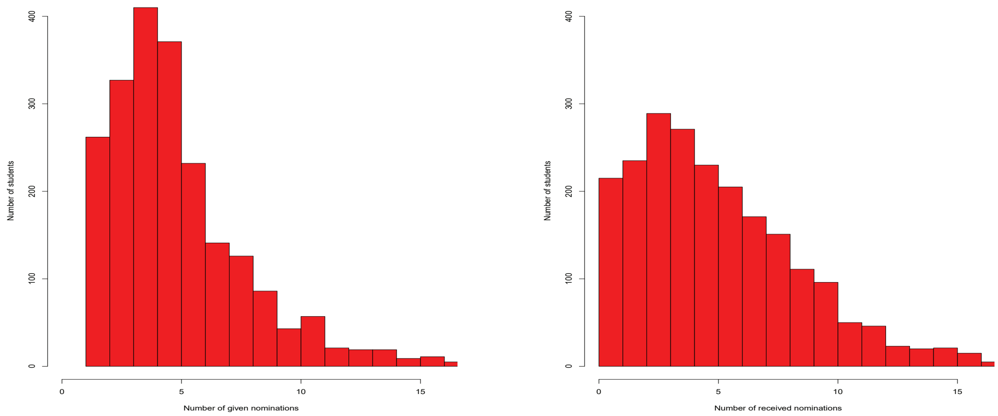
Combining both terms:

$$
\begin{aligned}
\hat{\theta}_{IVMD} = A^{-1}B &= \left(\sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{X}_l\right)^{-1} \left(\sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{X}_l \, \theta + \sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \varepsilon_l\right) \\
&= \theta + \left(\sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \mathbf{X}_l\right)^{-1} \left(\sum_l \mathbf{X}_l' \, \mathbf{P}_l \, \varepsilon_l\right) = \hat{\theta}_{IV}
\end{aligned}
$$

□

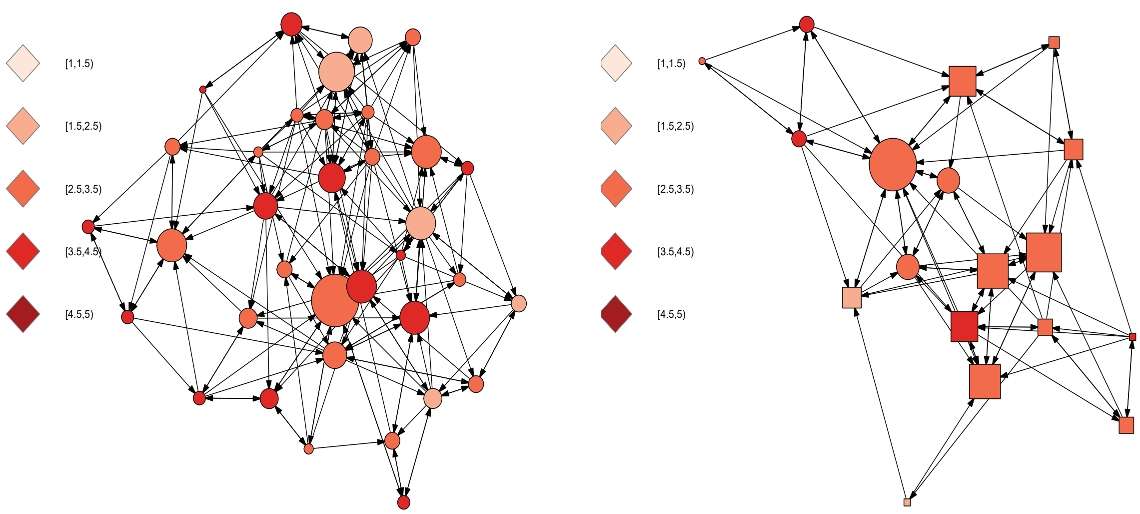
B Figures

Figure 1: Distribution of naming friends



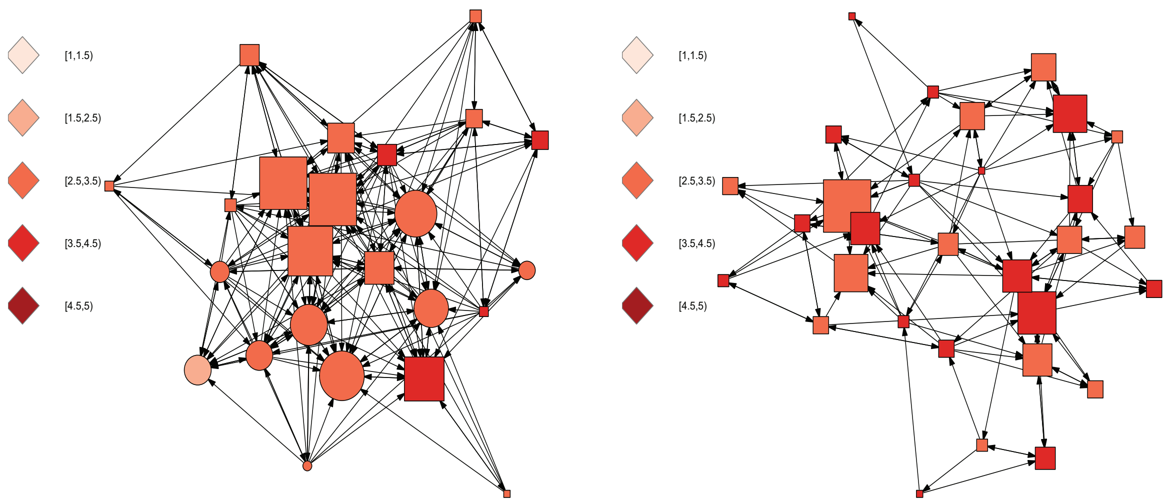
Histogram of the names given as friends (outdegrees) and individuals named as friends (indegrees). The median (mean) for the outdegrees and as well as for the indegrees is 5 (5.8). Source: *NRW Gymnasiasten-Studie*.

Figure 2: The largest and the smallest classroom networks



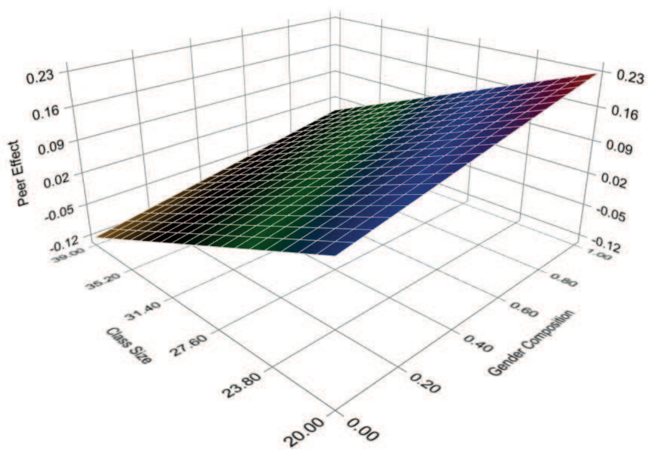
The size of a node is proportional to its outdegree, its color indicates the GPA score (lighter colors representing better performance), and the shape of the node indicates the gender, i.e., circles represent female students and squares represent male students. Left: Largest classroom network, $n = 35$, density $= 0.15$, clustering $= 0.007$, girls class. Right: smallest classroom, $n = 18$, density $= 0.26$, clustering $= 0.03$, female ratio $= 0.33$. Source: *NRW Gynasiasten-Studie*.

Figure 3: Classroom networks with the highest and the lowest density



The size of a node is proportional to its outdegree, its color indicates the GPA score (lighter colors representing better performances), and the shape of the node indicates the gender, i.e., circles represent female students and squares represent male students. Left: Densest classroom network, $n = 24$, density = 0.37, clustering = 0.07, female ratio = 0.37. Right: least dense classroom network, $n = 30$, density = 0.14, clustering = 0.004, boys class. Source: *NRW Gynasiasten-Studie*.

Figure 4: Peer Effects by Class Size and Gender Composition



Surface of peer effect by gender composition and class size for an outdegree of 5 based on the parameter estimates of the composite model given in Table 2. The peer effect denotes the change in a student's GPA score due to a one-unit change in the GPA of all peers assuming a median outdegree of 5.

C Tables

Table A1: IV-MD Estimation Results: German

	Heterogeneous Peer Effects Model		
	Composite	Local-aggregate	Local-average
Local-aggregate peer effect			
Intercept	0.0024***	0.0016**	
	(0.0007)	(0.0007)	
Relative Class Size	0.0049	0.0053	
	(0.0055)	(0.0056)	
Female Share	-0.005***	-0.0029*	
	(0.0016)	(0.0016)	
Local-average peer effect			
Intercept	0.0338		0.1442***
	(0.0267)		(0.0311)
Relative Class Size	0.8538***		0.576***
	(0.1805)		(0.2017)
Female Share	0.2476***		0.2099***
	(0.0525)		(0.0580)
Own characteristics			
IQ	-0.0046***	-0.0058***	-0.0053***
	(0.0009)	(0.0009)	(0.0010)
Previous GPA	0.8238***	0.8722***	0.8458***
	(0.0160)	(0.0165)	(0.0172)
Age	-0.0508***	-0.0496***	-0.0653***
	(0.0088)	(0.0092)	(0.0097)
Peers' characteristics			
IQ	0.0025	-0.0013	0.0059***
	(0.0017)	(0.0018)	(0.0018)
Previous GPA	0.0107	0.0191	0.0202
	(0.0390)	(0.0329)	(0.0440)
Age	0.0023	-0.0315*	0.0343**
	(0.0165)	(0.0181)	(0.0172)
Wald statistics (d.f.)		24.48 (3)	46.38 (3)

Estimates of the three model variants obtained by IVMD estimation. The first column corresponds to the composite model, the second column corresponds to the local-aggregate model and the third column corresponds to the local-average model estimation results. Following Liu et al. (2014), we use the average characteristics of the second order peers as instruments for the local-average part, and the aggregate characteristics of the peers for the local-aggregate part. The Wald statistics test the composite model against the local models. Robust standard errors are reported in parentheses. First-stage errors are assumed to be heteroskedastic, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$, N=2165, L=85.

Table A2: IV-MD Estimation Results: Math

	Heterogeneous Peer Effects Model		
	Composite	Local-aggregate	Local-average
Local-aggregate peer effect			
Intercept	0.0004	-0.0002	
	(0.0007)	(0.0008)	
Relative Class Size	0.0107*	0.0037	
	(0.0064)	(0.0066)	
Female Share	-0.0047***	-0.0036*	
	(0.0018)	(0.0019)	
Local-average peer effect			
Intercept	-0.0146		-0.0001
	(0.0285)		(0.0338)
Relative Class Size	1.2058***		0.4928**
	(0.1946)		(0.2335)
Female Share	-0.1942***		-0.1405**
	(0.0551)		(0.0614)
Own characteristics			
IQ	-0.0162***	-0.0172***	-0.0171***
	(0.0011)	(0.0012)	(0.0012)
Previous GPA	0.8370***	0.8178***	0.8678***
	(0.0197)	(0.0207)	(0.0216)
Age	0.0695***	0.0615***	0.0626***
	(0.0106)	(0.0111)	(0.0122)
Peers' characteristics			
IQ	0.0185***	0.0143***	0.0109***
	(0.0020)	(0.0022)	(0.0024)
Previous GPA	-0.0938**	0.025	-0.0929**
	(0.0416)	(0.0386)	(0.0460)
Age	0.0036	0.0583***	0.0577**
	(0.0216)	(0.0212)	(0.0237)
Wals statistics (d.f.)		8.61 (3)	56.23 (3)

Estimates of the three model variants obtained by IVMD estimation. The first column corresponds to the composite model, the second column corresponds to the local-aggregate model and the third column corresponds to the local-average model estimation results. Following Liu et al. (2014), we use the average characteristics of the second order peers as instruments for the local-average part, and the aggregate characteristics of the peers for the local-aggregate part. The Wald statistics test the composite model against the local models. Robust standard errors are reported in parentheses. First-stage errors are assumed to be heteroskedastic, $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$, N=2165, L=85.

Table A3: Homogeneous Model Estimates

	GPA			**German**			**Math**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
Aggregate Peer Effect	0.0007	0.0008		0.0030**	0.0030**		0.0014	0.0012	
	(0.0006)	(0.0006)		(0.0012)	(0.0012)		(0.0015)	(0.0014)	
Average Peer Effect	0.8025		0.4181	-0.0977		-0.0277	0.2464		0.1863
	(0.6213)		(0.6822)	(0.6231)		(0.6381)	(0.6165)		(0.6191)
Own characteristics									
IQ	-0.0032***	-0.0032***	-0.0032***	-0.0061***	-0.0062***	-0.0063***	-0.0197***	-0.0195***	-0.0197***
	(0.0008)	(0.0008)	(0.0008)	(0.0016)	(0.0016)	(0.0016)	(0.0022)	(0.0021)	(0.0022)
Previous GPA	0.7295***	0.7351***	0.7322***	0.7971***	0.7966***	0.7969***	0.8093***	0.8059***	0.8083***
	(0.0164)	(0.0150)	(0.0159)	(0.0309)	(0.0308)	(0.0310)	(0.0377)	(0.0371)	(0.0376)
Age	-0.0095	-0.0159**	-0.0130	-0.0616***	-0.0608***	-0.0629***	0.0297	0.0318	0.0294
	(0.0102)	(0.0079)	(0.0102)	(0.0178)	(0.0166)	(0.0178)	(0.0219)	(0.0213)	(0.0219)
Peers' characteristics									
IQ	0.0034	0.0014	0.0024	0.0029	0.0032	0.0029	0.0125	0.0078*	0.0112
	(0.0024)	(0.0016)	(0.0024)	(0.0039)	(0.0035)	(0.0040)	(0.0128)	(0.0044)	(0.0128)
Previous GPA	-0.5522	0.0607**	-0.2548	0.1640	0.0766	0.1191	-0.2161	-0.0112	-0.1582
	(0.4798)	(0.0297)	(0.5247)	(0.5613)	(0.0613)	(0.5756)	(0.5190)	(0.0730)	(0.5181)
Age	-0.0122	-0.0268	-0.0192	-0.0567	-0.0516	-0.0535	0.0203	0.0276	0.0223
	(0.0217)	(0.0174)	(0.0219)	(0.0497)	(0.0340)	(0.0506)	(0.0472)	(0.0436)	(0.0471)
Wald statistics (d.f.)	63.84 (4)	26.30 (2)	20.39 (2)	53.85 (4)	3.36(2)	23.28 (2)	68.90 (4)	3.51 (2)	9.04 (2)

IV estimates for the three models. For each outcome variable, the first column presents the results for the composite model, the second column presents the results for the local-aggregate model, and the third column presents the results for the local-average model estimation results. Following Liu et al. (2014), we use the average characteristics of the second order peers as instruments for the local-average part, and the aggregate characteristics of the peers for the local-aggregate part. The Wald statistics test the heterogeneous models against the homogeneous models. Robust standard errors are in parentheses. Robust standard errors are in parentheses. First-stage errors are assumed to be heteroskedastic. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$, N=2165, L=85.