Heterogeneity in Network Peer Effects*

Livia Shkoza[a]
University of Konstanz, GSDS

Derya Uysal[b]
University of Munich, CESifo

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

September 17, 2020

Abstract

We study the peer effects on school achievement exploiting the network structure of friendships within a classroom. 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 this purpose we propose a novel Instrumental Variable–Minimum Distance (IV-MD) estimation approach. Our empirical findings are based on a unique network dataset from the 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.

Keywords: Social network, Peer effects, Network heterogeneity, Instrumental Variables, Minimum Distance estimation, School achievement
JEL classification: D85, L14, C3, I21,

---

* 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. For helpful comments we would like to thank Michael König, Guido Schwerdt, Tobias Kärner and Stephan Schumann. The usual disclaimer applies.

[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. This is especially valid and essential in education and hence the identification and estimation of peer effects in education has been an important research area. We contribute to this area by analyzing heterogeneity in peer effects using a previously unexploited data set from Germany. We rely on identification results which are generalizations of Bramoullé et al. (2009) to estimate the peer effects on school grades exploiting the friendship network structure within classrooms. Unlike the existing literature, where the peer effect is assumed to be the same for each network, independent of the characteristics of the network, we allow there to be heterogeneity in peer effects across networks. Only a few studies take into account heterogeneity of peer effects by individual characteristics (see, for example, Hoxby & Weingarth, 2005; Beugnot et al., 2019, among others), while heterogeneity of network effects at the network level has only been investigated by Lin (2014) and Calvó-Armengol et al. (2009). They, however, do not incorporate heterogeneity into the estimation framework. In the present paper we propose a novel Instrumental Variable-Minimum Distance (IV-MD) approach that is particularly suitable to estimate and to test for various dimensions of heterogeneity in peer effects.

In his seminal work, Manski (1993) explains the dependence of an individual's behavior on the behavior of others in a socially interactive environment via 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). When it comes to the identification of these effects, researchers have proposed several strategies. Bramoullé et al. (2020) provide a comprehensive survey of the existing literature. They divide the literature into two strands depending on how correlated effects are dealt with. One strand of the literature assumes that the formation of links in the network may be endogenous and that unobserved correlated factors may affect the outcome, and develops strategies for the identification of peer effects.[1] The second strand of the literature rests on the assumption that the network of interactions and observed characteristics are exogenous, which circumvents the problem of correlated errors. This strand of the literature also includes the cases where the correlated effects are controlled for in the form of network fixed effects. Once the problem of correlated effects is solved, the reflection problem (Manski, 1993), i.e., the challenge of the separate identification of the endogenous peer effect from the contextual peer effect, remains to be tackled. As Boucher et al. (2014) state, most applications, in particular those which lack network information, are incapable of offering an explicit solution to this problem. For example, Graham (2008) proposes a new method for identifying social interactions using conditional variance restrictions. However, he does not distinguish between these two types of social interactions. One notable exception is Lee (2007), where the identification relies on variation in group sizes.

Observing the network structure, however, opens up more avenues for identification of the endogenous peer effects. Bramoullé et al. (2009) show that the endogenous peer effects are identified separately using instruments which can be derived naturally from the network structure if there are no correlated effects or if the correlated effects are addressed as network fixed effects. The intuition behind their identification results is that by exploiting the network structure along the lines of the econometric literature on spatial lags, exogenous variation in the covariates of the second-order peers ('peers of the peers') serve as valid instruments to identify endogenous peer effects provided the network is sufficiently sparse.

In the spirit of Bramoullé et al. (2009), Liu & Lee (2010) and Liu et al. (2014) also

[1] Since we follow the second strand of the literature, we refer to Bramoullé et al. (2020) for the strategies used for studying endogenous link formation.

discuss identification based on instruments in different models. In the present paper, we follow the identification strategy proposed by Liu et al. (2014) because it has several inherent advantages. First, the network-specific peer effects can be given a structural interpretation, as the model relies on a microeconomic foundation, i.e., utility-maximizing behavior in a Nash equilibrium (similar to Blume et al., 2015; Calvó-Armengol et al., 2009, among others). Secondly, the model of Liu et al. (2014) is flexible in how peers affect an individual's behavior. A widespread assumption is that peer effects reflect norms in the sense that the individuals align their behavior with the standards of their peer group represented by the mean behavior of the peer group, since behavior that deviates from that of one's peer group may inflict a loss in terms of the individual's utility. This idea is reflected in the local-average model specification, which uses the mean behavior of the peer group as a predictor for the individual's behavior (for example, Boucher et al., 2014). Alternatively, one can argue that the individual profits from the pure size of their network and the quality of the peers in that network. For instance, if the transmission mechanism between an individual and their network is simply the quality of the information, the local-aggregate model specification seems to be a reasonable behavioral hypothesis (for example, Calvó-Armengol et al., 2009). Both specifications can be incorporated in a single composite model so that it is open to the data to decide in which way and how strongly peer behavior affects an individual's behavior.

Regardless of the identification strategies used to tackle the above–mentioned challenges, the implicit assumption made in the majority of studies is that the endogenous peer effects are homogeneous across networks. A notable exception is offered by Calvó-Armengol et al. (2009), who investigate the relationship between peer effects and the network topology. They provide graphic evidence that the strength of the network effect varies with certain structural network measures, such as density, asymmetry, and redundancy. However, the structural relation between the characteristics of the network and the endogenous effect, and thus the indirect effect of the characteristics of the network on the outcome variable, has not been investigated yet.

To fill this gap in the literature, we take a closer look at peer effects in heterogeneous networks. Starting from a model which incorporates two different types of endogenous peer effects, along the lines of Liu et al. (2014), we introduce network-specific heterogeneity. This allows us to study how exogenous network-specific factors determine the overall peer behavior. Since the model can be derived from utility-maximizing behavior, the network-specific peer effects have a structural interpretation and provide a different perspective on how heterogeneity affects peer behavior and, lastly, the outcome variable under investigation. We propose an easy to implement 'Instrumental Variables–Minimum Distance' estimator (IV-MD), which is motivated by the heterogeneity of the independent networks. In the modeling strategy we propose in this paper, the peer effects are driven by observable network-specific factors. In this way, we take into account heterogeneity and explain the impact of these network variables on the structure of the social interactions.

In the last couple of decades, numerous studies have tried to generate empirical evidence about the role of peer effects for the explanation of individual behavior. Sacerdote (2011) provides a comprehensive review of the existing empirical evidence, however, leaving aside any empirical evidence based on network data. The empirical evidence on the identification of peer effects exploiting network information is rather scarce due to the limited availability of appropriate network data. To the best of our knowledge, all the existing research papers on network peer effects use the National Longitudinal Study of Adolescent to Adult Health (Add Health). The empirical evidence on the existence and economic relevance of peer effects in networks using Add Health is rather diverse and involves various outcome variables and estimation approaches. Using the Generalized 2SLS strategy proposed by Kelejian & Prucha (1998), Bramoullé et al. (2009) estimate peer effects on recreational activities for the local-average model. With the same dataset, Liu et al. (2014) study peer effects on effort and sports activities within the composite model framework. They find that the IV estimates of the local-aggregate and local-average peer effects on study effort are small and positive but insignificant. However, for the GMM estimates, they find significant local-average peer effects. Concerning sports activities, the estimated local-aggregate effect also turns out to be small but is statistically significant.

In our paper, we study heterogeneous peer effects on school grades using unique network data from 85 school classes of secondary schools in Germany. The contribution of our study to the existing empirical literature is varied. Besides using a network dataset which has not been used before, our study reveals that ignoring network heterogeneity generally leads to insignificant estimates of peer effects, whereas taking it into account yields novel insights on how the peer effects operate. 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. Therefore our study also contributes to the literature investigating how students' gender affects peer outcomes. One strand of the literature on gender effects 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 the 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. Lu & Anderson (2015) investigate the same question looking at subclassroom groups and find differential effects among boys and girls, i.e., they find that being surrounded by girls has a positive effect on girls' test scores but no effect on boys' test scores. In a more recent study, Hill (2015) finds that a student's share of opposite gender school friends negatively affects high school GPA. The common point in these studies is that gender or gender ratio enters the reduced form equations as a regressor. In contrast to the reduced form approaches, 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 of the effect of class size on student achievement is even more varied than the evidence of the effect of gender (ratio). For example, Hanushek (1996) and Hoxby (2000) find no effect of class size reduction on achievement. Dobbelsteen et al. (2002) find 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), find a substantial positive effect of reducing class size on academic achievement. Similar to 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. In the following, we present a more structurally motivated approach, which relies on identifying another pathway for how class size affects school success indirectly via peer behavior.

The outline of this paper is as follows. In Section 2 we introduce the composite network model and work out its identification condition. In that section, we also introduce the new combined instrumental variables–minimum distance approach for the estimation of heterogeneous peer effects. In Section 3, we describe our network data and discuss further implementation issues. Section 4 contains the major empirical findings, and then Section 5 concludes and gives an outlook for future research.

2 The Network Model and Estimation

Theoretical Foundation

In our setup we assume there is a finite set of N agents, partitioned into L independent networks, and write n_l for the number of agents in the lth network ($l = 1, \ldots, L$). The social connections for network l are indicated in 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 $a_{i,l} = \sum_{j=1}^{n_l} a_{ij,l}$.

Write $G_l = [g_{ij,l}]$ for 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 $\sum_{j=1}^{n_l} g_{ij,l} = 1$. Each agent i exerts time or effort $y_{i,l}$ in some activity and $Y_l = (y_{1,l}, \ldots, y_{n_l,l})'$ is the vector of effort for network l. The utility an agent gets from exerting a specific level of effort depends on the return to that effort as well as on the cost realized due to exerting that effort. We assume the following utility function which captures both the individual and the social aspects of the costs of and returns to the ith activity:

$$u_{i,l} = u_{i,l}(y_{i,l}; Y_l, A_l) = \left(\kappa_{i,l}^* + \lambda_{1l} \sum_{j=1}^{n_l} a_{ij,l} y_{j,l} \right) y_{i,l} - \frac{1}{2} \left[y_{i,l}^2 + \lambda_{2l} \left(y_{i,l} - \sum_{j=1}^{n_l} g_{ij,l} y_{j,l} \right)^2 \right], \tag{1}$$

where $\lambda_{1l} \geq 0$ and $\lambda_{2l} \geq 0$. The benefit component is linear in own effort with a return equal to $\kappa_{i,l}^* + \lambda_{1l} \sum_{j=1}^{n_l} a_{ij,l} y_{j,l}$. It is assumed that the aggregate effort of the peers scaled by λ_{1l}, the social multiplier coefficient, together with the ex-ante individual heterogeneity $\kappa_{i,l}^*$ determine the return to effort. The cost component has two parts, too. The first part, $y_{i,l}^2$, is the direct cost of one's own effort. The second part represents the cost due to the deviations from the social norm, i.e., from the average level of effort of one's peers, which is scaled by a social conformity coefficient, λ_{2l}. Our theoretical model is based on Liu et al.

(2014), with one notable difference: unlike Liu et al., we allow the social multiplier coefficient as well as the social conformity coefficient to change with l, i.e., to be network-specific.

Given the utility function (1), the best response function of individual i is given by

$$y_{i,l} = \beta_{1l} \sum_{j=1}^{n_l} a_{ij,l} y_{j,l} + \beta_{2l} \sum_{j=1}^{n_l} g_{ij,l} y_{j,l} + \kappa_{i,l}, \tag{2}$$

with $\beta_{1l} = \lambda_{1l}/(1+\lambda_{2l})$ and $\beta_{2l} = \lambda_{2l}/(1+\lambda_{2l})$, and $\kappa_{i,l} = \kappa^*_{i,l}/(1+\lambda_{2l})$.[2] The coefficients β_{1l} and β_{2l} capture the local-aggregate endogenous peer effect and local-average peer effects, respectively.

Econometric Model

A general econometric network model can be formulated based on the best response function (2). We define the individual heterogeneity $\kappa_{i,l}$ as a function of individual characteristics, average characteristics of the peers, and some network-specific parameters, as follows:

$$\kappa_{i,l} = x'_{i,l}\delta_l + \sum_{j=1}^{n_l} g_{ij,l} x'_{j,l}\gamma_l + \eta_l + \epsilon_{i,l}. \tag{3}$$

Here, $x_{i,l}$ is a k_x-dimensional vector of exogenous variables for agent i in network l, δ_l, γ_l, η_l are the corresponding parameters, and $\epsilon_{i,l}$ is the error term of the model. Inserting Equation (3) into the best response function (2), we obtain our general econometric model

$$y_{i,l} = \beta_{1l} \sum_{j=1}^{n_l} a_{ij,l} y_{j,l} + \beta_{2l} \sum_{j=1}^{n_l} g_{ij,l} y_{j,l} + \sum_{j=1}^{n_l} g_{ij,l} x'_{j,l}\gamma_l + x'_{i,l}\delta_l + \eta_l + \epsilon_{i,l}, \tag{4}$$

for $i = 1, \ldots, n_l$ and $l = 1, \ldots, L$. The coefficients of our model are directly linked to the three effects defined by Manski (1993) which we introduced in Section 1. The coefficients β_{1l} and β_{2l} jointly represent the endogenous effect. The contextual effect is captured by

[2] The proposition on the uniqueness of Nash equilibrium, as well as the discussion on the existence of equilibrium in Liu et al. (2014) apply to our model, too. Obviously, the condition has to be specified in terms of network-specific coefficients λ_{1l} and λ_{2l}. Thus, the condition reads $a_l^{\max}\beta_{1l} + \beta_{2l} < 1$ where $a_l^{\max}$ is the highest outdegree in $l-$th network.

γ_l. Finally, the correlated effect is given by η_l. Under homogeneity of the peer effects, i.e., $\beta_{1l} = \beta_1$, $\beta_{2l} = \beta_2$, $\gamma_l = \gamma$, $\delta_l = \delta$ for $l = 1, \ldots, L$, our model reduces to the model of Liu et al. (2014). Two other special cases occur for both homogeneous and heterogeneous network models when $\lambda_1 = 0$ ($\lambda_{1l} = 0$ for the heterogeneous model) or $\lambda_2 = 0$ ($\lambda_{2l} = 0$ for the heterogeneous model). If $\lambda_1 = 0$ (or $\lambda_{1l} = 0$), the best response depends only on the average effort of the peers. Symmetrically, if $\lambda_2 = 0$ (or $\lambda_{2l} = 0$), the best response depends only on the aggregate effort of the peers. The models in these special cases are called the local-average and local-aggregate network models, respectively. Provided that the network adjacency matrix A is exogenous conditional on the control variables x_i and the network fixed effects η_l, identification can be achieved for the homogeneous local-average and local-aggregate network models, as well as the composite model under intransitivity when some peers of peers of an agent are not their peers. The results are directly applicable to our heterogeneous network model. The only practical difference is that the identification conditions must be satisfied for each network rather than for the whole network data.

In matrix notation, the general econometric model (4) takes the form

$$Y_l = \beta_{1l} A_l Y_l + \beta_{2l} G_l Y_l + G_l X_l \gamma_l + X_l \delta_l + \eta_l \iota_{n_l} + \epsilon_l, \qquad l = 1, \ldots, L\,, \tag{5}$$

where $Y_l = (y_{1,l}, \ldots, y_{n_l,l})'$, $X_l = (x_{1,l}, \ldots, x_{n_l,l})'$, while $\epsilon_l = (\epsilon_{1,l}, \ldots, \epsilon_{n_l,l})'$ and ι_{n_l} is an $n_l \times 1$ vector of ones. In a quasi-panel data fashion, we can partial out the network-specific effects by a within transformation by multiplying (5) 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

$$J_l Y_l = \beta_{1l} J_l A_l Y_l + \beta_{2l} J_l G_l Y_l + J_l X_l \delta_l + J_l G_l X_l \gamma_l + J_l \epsilon_l, \qquad l = 1, \ldots, L. \tag{6}$$

For simplicity of exposition, we rewrite the differenced model as follows:

$$\tilde{Y}_l = W_l \pi_l + \tilde{\epsilon}_l, \qquad l = 1, \ldots, L, \tag{7}$$

where $\tilde{Y}_l = J_l Y_l$ is the transformed vector of dependent variables, $W_l = J_l \begin{bmatrix} A_l Y_l & G_l Y_l & X_l & G_l X_l \end{bmatrix}$ is the regressor matrix of dimension $n_l \times k_w$ with $k_w = 2(1 + k_x)$ and $\pi_l = \left(\beta_{1l}, \beta_{2l}, \delta_l', \gamma_l'\right)'$ is the parameter vector of dimension $k_w \times 1$.

In what follows, we assume that the parameters β_{1l} and β_{2l} for network-specific peer effects can be explained by a set of network-specific observable factors:[3]

$$\beta_{jl} = m_l' \beta_j\,, \qquad j = 1, 2, \tag{8}$$

where m_l is a $k_m \times 1$ vector of network-specific characteristics including an intercept. For the remaining parameters, we assume that they are the same across networks, i.e., $\gamma_l = \gamma$ and $\delta_l = \delta$. The restriction between the first stage reduced form parameter vector π_l and the structural form parameter vector $\theta = \left(\beta_1', \beta_2', \gamma', \delta'\right)'$ is given by

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

with

$$M_l = \begin{bmatrix} m_l' & 0 & 0 & 0 \\ 0 & m_l' & 0 & 0 \\ 0 & 0 & I_{k_x} & 0 \\ 0 & 0 & 0 & I_{k_x} \end{bmatrix}_{2(1+k_x)\times 2(k_m+k_x)}.$$

Stacking the restrictions between the L reduced form parameter vectors π_l and the structural form parameter θ given by (9) into a hyper-system yields

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

[3] This assumption can be generalized by assuming that β_{jl} can be replaced by a linear predictor representation along the lines of Chamberlain (1984) for panel data models with correlated random effects.

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

Estimation

We propose estimating θ by Minimum Distance (MD) employing two estimation stages: First, the reduced form parameters π_l are estimated Instrumental Variables (IV) for each l and stacked together in $\hat{\pi}$. In the second stage, the structural form parameter θ is obtained by minimizing the distance between $\hat{\pi}$ and $\pi(\theta)$ based on the (estimated) asymptotically optimal weighting matrix.

Since the networks are assumed to be independent, a systems regression approach does not yield any gains in efficiency over a simple single equation estimation approach for the first estimation stage. As instruments we use the exogenous variables, their counterparts for the peers, aggregate characteristics of the peers, and the covariates of the second order peers as over-identifying instruments, i.e., $Z_l = J_l \begin{bmatrix} X_l & G_l X_l & A_l X_l & G_l^2 X_l \end{bmatrix}$. It is intuitive that our instruments are valid under intransitivity, since the characteristics of the second order friends have only an effect on the outcome through their effect on the outcomes of the first order friends. Thus, the instrumental variable estimator for each network is

$$\hat{\pi}_l = \left[W_l' Z_l \left(Z_l' Z_l \right)^{-1} Z_l' W_l \right]^{-1} W_l' Z_l \left(Z_l' Z_l \right)^{-1} Z_l' Y_l, \qquad l = 1, \ldots, L \tag{11}$$

If heteroskedasticity is assumed, then for each network the asymptotic distribution of $\hat{\pi}_l$ is

$$\sqrt{n_l} \left(\hat{\pi}_l - \pi_l \right) \xrightarrow{d} \mathcal{N} \left(0, \Omega_l \right)$$

where $\Omega_l = \Delta_l V_l \Delta_l'$ with

$$\Delta_l = \left(\mathrm{E}\left[W_l' Z_l\right]' \mathrm{E}\left[Z_l' Z_l\right]^{-1} \mathrm{E}\left[W_l' Z_l\right]\right)^{-1} \mathrm{E}\left[W_l' Z_l\right]' \mathrm{E}\left[Z_l' Z_l\right]^{-1},$$
$$V_l = \mathrm{E}\left[Z_l' \tilde{\epsilon}_l \tilde{\epsilon}_l' Z_l\right].$$

For the homoskedastic case with $\mathrm{E}\left[\tilde{\epsilon}_l \tilde{\epsilon}_l' \middle| Z_l\right] = \sigma_l^2 I_{n_l}$ and $V_l = \sigma_l^2\, \mathrm{E}\left[Z_l' Z_l\right]$, the asymptotic variance of $\hat{\pi}_l$ reduces to

$$\Omega_l = \sigma_l^2 \left(\mathrm{E}\left[W_l' Z_l\right]' \mathrm{E}\left[Z_l' Z_l\right]^{-1} \mathrm{E}\left[W_l' Z_l\right]\right)^{-1}.$$

The variance of $\hat{\pi}_l$ can be estimated using the sample counterpart of the asymptotic variance where expectations are replaced with sample means and the unknown population parameters with their estimates. For homoskedastic errors, the estimator for the variance covariance matrix is

$$\widehat{V}\left(\hat{\pi}_l\right) = \hat{\sigma}_l^2 \left(W_l' Z_l{}' (Z_l' Z_l)^{-1} W_l' Z_l\right)^{-1},$$

where $\hat{\sigma}_l^2$ is a consistent estimator of σ_l^2. For example, $\hat{\sigma}_l^2 = \frac{1}{n_l - k_w} \hat{\tilde{\epsilon}}_l' \hat{\tilde{\epsilon}}_l$ is a consistent estimator for σ_l^2, where $\hat{\tilde{\epsilon}}_l$ is the residual of the differenced model in Equation 7, i.e., $\hat{\tilde{\epsilon}}_{i,l} = \tilde{Y}_l - W_l \hat{\pi}_l$. If the error term is assumed to be heteroskedastic, $\mathrm{E}\left[Z_l' \tilde{\epsilon}_l \tilde{\epsilon}_l' Z_l\right]$ can be estimated by $Z_l' D Z_l / n_l$, where D is an $n_l \times n_l$ diagonal matrix with entries $\widehat{\tilde{\epsilon}}_{i,l}^2$. In that case, the variance can be estimated as follows:

$$\widehat{V}\left(\hat{\pi}_l\right) = \left[W_l' Z_l (Z_l' Z_l)^{-1} Z_l' W_l\right]^{-1} \left[W_l' Z_l (Z_l' Z_l)^{-1} Z_l' D Z_l (Z_l' Z_l)^{-1} Z_l' W_l\right] \left[W_l' Z_l (Z_l' Z_l)^{-1} Z_l' W_l\right]^{-1}.$$

For the stacked vector of reduced form parameters π, we get $\sqrt{N}\left(\hat{\pi} - \pi\right) \xrightarrow{d} \mathcal{N}\big(0, \Omega\big)$. Because of the independence of the networks, Ω is a block-diagonal matrix with diagonal elements equal to Ω_l, i.e., $\Omega = diag[\Omega_l]$.

In the second step, we estimate the structural form parameter by minimizing the weighted quadratic distance between the estimated reduced form parameter vector $\hat{\pi}(\theta) = \big(\hat{\pi}_1', \hat{\pi}_2', \ldots, \hat{\pi}_L'\big)'$ and $M\theta$ with respect to the structural parameter vector θ. Using the efficient weighting matrix, the inverse of any consistent estimator of Ω, the minimization problem can be written formally as follows:

$$\hat{\theta}_{MD} \equiv \arg\min_{\theta} \big[\hat{\pi} - M\theta\big]'\hat{\Omega}^{-1}\big[\hat{\pi} - M\theta\big]. \tag{12}$$

Since the restriction between π and θ is linear, the MD estimation simply reduces to a generalized least squares regression of $\hat{\pi}$ on M:

$$\hat{\theta}_{MD} = \big(M'\hat{\Omega}^{-1}M\big)^{-1}M'\hat{\Omega}^{-1}\hat{\pi},$$

with $\sqrt{N}\left(\hat{\theta}_{MD} - \theta\right) \xrightarrow{d} \mathcal{N}\big(0, (M'\Omega^{-1}M)^{-1}\big)$. Since the optimal weighting matrix is block diagonal in our case, the minimum distance estimator can be simply computed as $\hat{\theta}_{MD} = \big(\sum_{l=1}^{L} M_l'\hat{\Omega}_l^{-1}M_l\big)^{-1}\big(\sum_{l=1}^{L} M_l'\hat{\Omega}_l^{-1}\hat{\pi}_l\big)$ even for a large number of networks and a large number of covariates. The standard errors can be estimated based on an asymptotic variance formula by replacing the unknown Ω with its consistent estimator.

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 North Rhine-Westphalia (NRW) in the years 1969 and 1970. The students were sampled from 121 classes at 68 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 this initial survey, a standard psychometric Intelligence Structure Test (IST) was administered in the classroom during the 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*. In order 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]

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

As mentioned in the Introduction, the vast majority of the empirical papers studying network peer effects use the Add Health data. Our unique network data differs from the Add Health data in several ways. One difference is that unlike the Add Health data, all the students in the sample were asked about their relationships of friendship, so that we can observe the entire network within each classroom. Another difference is that in the Add Health data, no information was collected at the class level. Lastly, 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.

We constructed our dataset by merging information from three different sources: student surveys, administrative data from school archives, and the sociometric test. We dropped an observation if any of the variables used in the empirical model were missing. Similarly, we dropped isolated individuals, namely, one who did not name anyone as a friend. This leaves a sample of 2,385 students and 101 classes. After cleaning the data, in our empirical study we excluded classes with fewer than 18 students, because the first-stage estimates of a small networks suffer from low degrees of freedom.[5] Excluding small classes leads to 2,165 students in 85 classes. Table 1 contains the summary statistics of the variables

[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.*"

[5] We estimated the model using several thresholds, but in general the results did not change qualitatively.

used in our empirical analysis. 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.

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	3.19	0.49	3.19	0.49
Age	15.38	0.87	15.37	0.87
Network measures				
Original Class size	27.82	6.01	29.19	4.92
Effective Class size[b]	23.61	5.98	25.47	4.23
Original gender composition	0.45	0.43	0.49	0.43
Effective gender composition[b]	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: Effective class size (gender composition) refers to the number of remaining students in the classroom after dropping the individuals with missing information either in the survey data, administrative data, or in the sociometric test. The same holds for the gender composition measures.

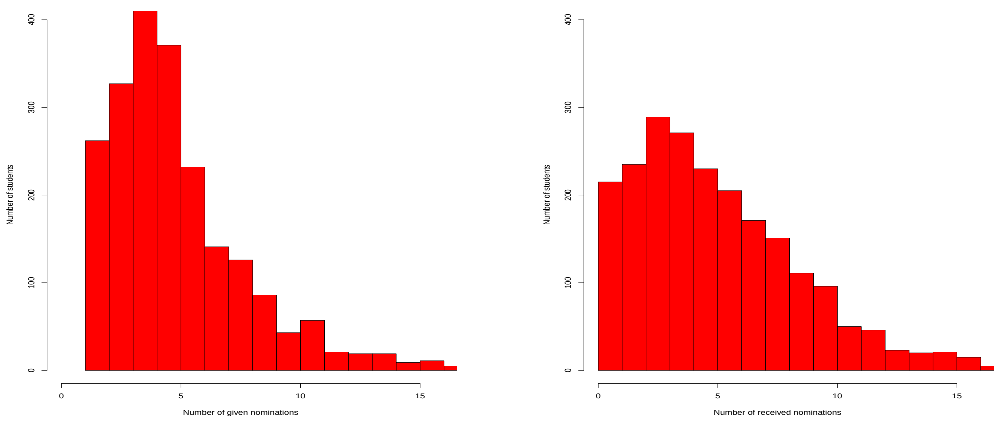
We measure the academic performance using the average of the final grades (GPA) for all compulsory and elective courses at the end of the school year 1969/70. We use the administrative data collected from the school archives to construct the GPA. At the time of the survey the choice of different courses within a class was very limited, i.e., all students of the same class basically had to take the same courses. Selection to certain specializations (e.g., languages, mathematics and sciences, humanistic secondary 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. Besides the overall GPA, we will take a closer look at the scores in Mathematics and German in order to detect potential differences in peer behavior across subjects.

The individual heterogeneity in our model is captured by the student's IQ score, the overall GPA from the previous school year, and the student's age. The IQ is constructed from the correctly solved questions of the IST. 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 β_{1l} and β_{2l} to depend on class-specific factors. As such factors we use the class size, i.e., the size of the network, and the fraction of girls in the class. As we already discussed in detail in the Introduction, 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 fact, Lin (2014) estimates the peer effects for large, i.e., larger than the median, and small classes separately, using the Add Health dataset, and finds that the peer effects are considerable different for 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 the 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 of i's peers name each other. In our dataset this measure varies between 0.004 and 0.08, supporting the argument that peers of peers provide new information for the model.

To obtain a better understanding of the network structure and its potential role for peer effects, a look at the summary statistics of the naming of friends may be helpful. The average number of friends a student named (outdegree) is 5.48, which indicates that the students take the selection of friends seriously. Isolated students are not part of the estimation sample, so every student has named at least one peer and 51 students are not named as a friend by any other classmate. Figure 1 depicts the distribution of friends. The distribution of outdegrees and indegrees indicates that the networks are sparse which is essential for the identification of network-specific endogenous peer effects. Most of the students name around five people that they like, and very few name more than ten peers.

Figure 1: Distribution of naming friends

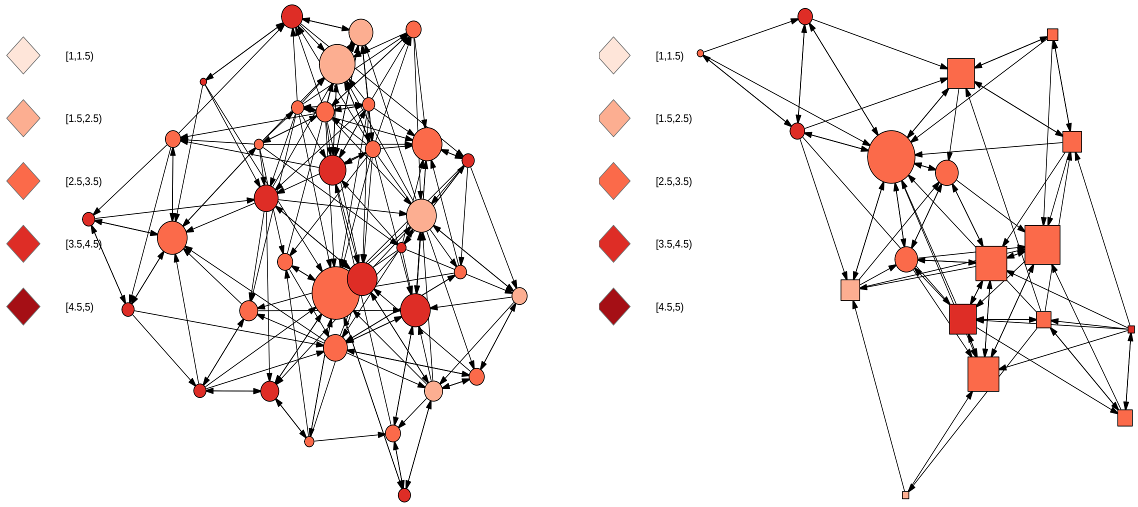


Histogram of the names given as friends (outdegrees) and individuals named as friends (indegrees). The median (mean) for the outdegrees and as well as for the indegrees is 5 (5.8). Source: *NRW Gymnasiasten-Studie*.

Our motivation for considering potential heterogeneity in peer effects results from the observation that the network structures and characteristics vary substantially across networks. We claim that the differences in networks might affect how peer effects operate. With the help of some network graphs, we illustrate the variation of the school classes (networks) 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 a small picture to be detected, we concentrate on four classes.

First, we plot the largest and smallest classrooms in Figure 2. Second, Figure 3 depicts the two networks with the highest and the 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 2: The largest and the smallest classroom networks

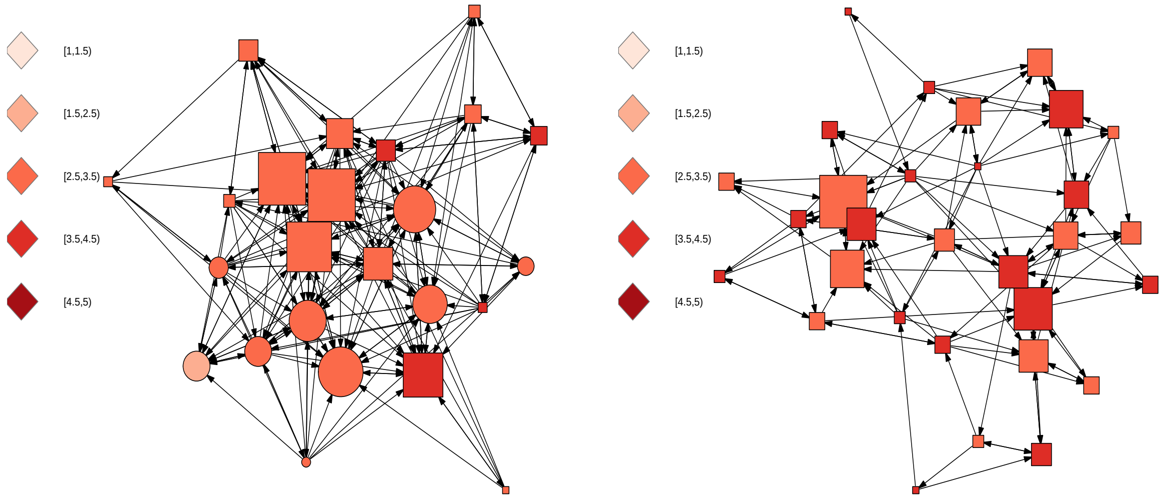


Note: The size of a node is proportional to its outdegree, its color indicates the GPA score (lighter colors representing better performance), and the shape of the node indicates the gender, i.e., circles represent female students and squares represent male students. Left: Largest classroom network, $n = 35$, 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 provides further exploratory evidence that the network structure varies across classrooms. The least dense network on the right reveals two major clusters, while the dense network on the left 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 good students 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.[6]

[6] The corresponding network graphs for the two other outcome variables can be obtained from the authors on request.

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



Note: 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*.

4 Empirical Results

The primary specification in our study is the heterogeneous composite model given by (5). Although most of the empirical studies focus on peer effects as a result of 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 taught, 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 standardized GPA of the previous year, standardized IQ, and age, as well as their counterparts for the student's peer group.

Table 2 summarizes the estimation results for the composite, local-aggregate, and local-average models with heterogeneous peer effects based on the IV-MDE approach with globally differenced variables. In order to ease the interpretation of the estimation results, we centered the network-specific characteristics around their means, so that the two intercept terms in (8) reflect the aggregate and the 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 peer effects significantly differ by class size and gender decomposition. This finding stands in clear contrast to many studies focusing on homogeneous peer effects, which rarely find sufficient statistical evidence for their presence (e.g. Boucher et al., 2014; Liu et al., 2014). This is supported by our estimation results for the composite model and the two submodels for the GPA and the scores in German and Math assuming homogeneity in the peer effects given in Table A3 in the Appendix. Similar to the results of previous studies, we also find for our sample insignificant estimates when peer effects are assumed to homogeneous.

Secondly, the way peers affect a student's performance also matters, as both mechanisms, the local-aggregate behavior and the local-average behavior, both have a positive impact on a student's educational attainment in a representative classroom with average size and gender composition. The comparison of MD statistics in the last row of Table 2 shows that the heterogeneous submodels have to be rejected in favor of the heterogeneous composite model.

In the composite model both intercepts turn out to be positive at the 1% and 10% significance levels, respectively. This means that if the peers perform better individually or on average, then so does the individual. It is important to note that the size of the coefficients from the two submodels are not directly comparable. Because the adjacency matrix for the local-aggregate effect is not normalized 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. In order to have a comparable measure of the size of the two peer effects, we measure the strength of a peer effect by the change in the GPA score due to a one unit increase of the GPA score of the peers. For a class with average characteristics the local-aggregate effect would exceed the local-average effect if the student has more than 22 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.

Table 2: IV-MD Estimation Results: GPA

	Heterogeneous Peer Effects Model		
	Composite	Local-aggregate	Local-average
Local-aggregate peer effect			
Intercept	0.0020***	0.0012***	
	(0.0003)	(0.0003)	
Class Size	-0.0002***	-0.0003***	
	(0.0001)	(0.0001)	
Gender Composition	-0.0023***	-0.0011	
	(0.0008)	(0.0008)	
Local-average peer effect			
Intercept	0.0459*		0.1526***
	(0.0260)		(0.0293)
Class Size	-0.0114***		-0.0083***
	(0.0028)		(0.0032)
Gender Composition	0.1207***		0.1432***
	(0.0324)		(0.0353)
Own characteristics			
IQ	-0.0299***	-0.0284***	-0.0243***
	(0.0038)	(0.0038)	(0.0040)
Previous GPA	0.3630***	0.3660***	0.3621***
	(0.0035)	(0.0035)	(0.0038)
Age	-0.0056	-0.0036	-0.0023
	(0.0038)	(0.0039)	(0.0043)
Exogenous peer effects			
IQ	-0.0024	-0.0031	0.0010
	(0.0069)	(0.0074)	(0.0077)
Previous GPA	0.0223 *	0.0421***	-0.0230*
	(0.0120)	(0.0071)	(0.0135)
Age	-0.0254***	-0.0145 *	-0.0230**
	(0.0076)	(0.0081)	(0.0089)
MD-statistics (d.f.)	3168.16 (668)	3202.36 (671)	3225.67 (671)

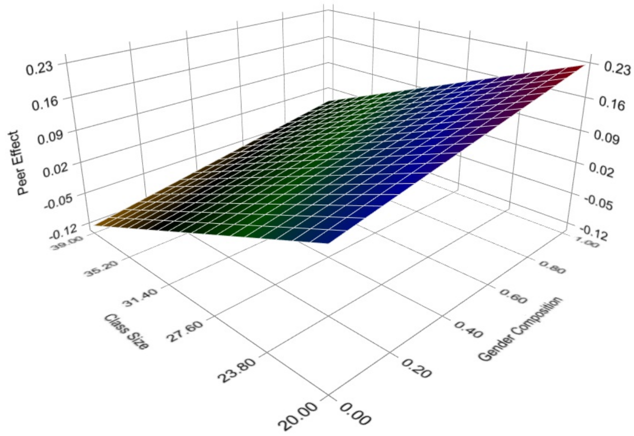
Estimates of the three model variants obtained by IV-MD estimation. The first column corresponds to the composite model, the second column corresponds to the local-aggregate model and the third column corresponds to the local-average model estimation results. The IV-MD estimates of the two submodels are based on first stage IV-estimates of the submodels. 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.

Interestingly, the two coefficients on the gender composition variable operate in different directions. The gender effect is large and positive for the local-average mechanism but small and negative for the local-aggregate mechanism. Interpreting the local-average effect as a proxy for norm behavior, we conclude that our estimates indicate that in girls-only classes, norm behavior is much more present than in boys-only classes. In order to illustrate the sizes of the effects of the two components, consider two classes of average size, one being a girls-only class while the other class is a boys-only class. Moreover, assume that the median outdegree is 5. In this case, the overall peer effect of a female student is 0.1115 (= 0.0041 + 0.1074) compared to 0.0023 (= 0.0156 - 0.0132) for a student in a boys-only class.

The effect of class size is negative and significant for both mechanisms, meaning that for larger classrooms, the enhancing contribution of peer behavior diminishes and can become negative for large classes. Assuming an average gender decomposition, the peer effect of the smallest class size is 0.1698 (= 0.0192 + 0.1506), while for the largest class size we find a small but negative overall peer effect of -0.0657 (= -0.0659 + 0.0002). 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 under consideration in studies on the determinants of educational achievement. Our approach takes into accounts the conventional direct effect by global differencing, so that the effect of class size on peer behavior is 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 has a unique structural interpretation. It indicates the role of social interactions in a classroom, which then partly determines the individual performance. Therefore this approach also offers a specific explanation of why certain classes have certain performances.

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 gender composition have a negative overall peer effect if the class size is larger than 34 students. All in all, the combined peer effect ranges from -0.12 to 0.23.

Figure 4: Peer Effects by Class Size and Gender Decomposition



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.

The coefficients on own IQ and own previous GPA have the expected signs. Not very surprisingly, the GPA of the previous year is a very good 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. The previous GPA of the peers does significantly impact the individual outcomes at the 10% level: if the peers have better grades, the individual does as well. 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 gender composition 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-average 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 IV-MD estimates of the heterogeneous local and composite models with scores in Math and German as dependent variables are given in Tables A1 and A2 in the Appendix. There is also a heterogeneity by subject 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 large and significant positive coefficient on gender composition for the local-average effects for German and Math, which indicates that the role of gender composition in educational attainment has to be discussed in the light of the specific subject or field of study. In the same spirit, the role of class size on peer behavior needs to be discussed, since the subject, as related to the class size, has a significant positive effect on the local-aggregate peer effect.

Robustness Checks

For identification, the first stage IV estimates require that certain rank conditions on the adjacency matrices are satisfied (Liu et al., 2014). For the local-average model to be identified, I, G, G^2 and G^3 have to be linearly independent. For the local-aggregate model, when there are different outdegrees, identification is achieved if I, A, G and AG are linearly independent. We follow the procedure proposed by Bramoullé et al. (2009) to check the linear independence of these matrices. For both types of models, the identification conditions in our application are satisfied.

In the literature on network peer effects, the instruments are naturally derived from the reduced form of the model, but their strength depends on the network topology. Startz &

Wood-Doughty (2017) show that the strength of the instruments in the local-average model is closely related to the density of the network. The results of their Monte Carlo simulation suggest weak instruments for densities larger than 0.05. Revisiting the summary statistics in Table 1, we see that the classrooms are denser on average, compared to the 0.05 cut-off in Startz & Wood-Doughty (2017). Checking the first stage F-statistics for the local and composite models, we find that the instruments are strong for the endogenous local-aggregate peer effect but weaker for the local-average peer effect. This is reflected in an instability of the local-average peer effect across different outcomes or model specifications.

Our IV-MD approach for the heterogeneous composite model allows us to test against a number of nested specifications using the Minimum χ^2-statistics defined as the difference between the MD-statistics of the nested model against the unrestricted alternative with degrees of freedom equal to the number of restrictions. A comparison of the MD-statistics presented in the last row of Table A3 for the nested homogeneous specifications with their counterparts for the heterogeneous specifications reveal that for all three outcome variables, the H_0 of homogeneous peer effects has to be rejected against the heterogeneous peer effects models. These findings hold for the heterogeneous composite model as well as for the heterogeneous local models.

Besides gender composition and class size, we further included network measures such as density and clustering as additional explanatory variables in the peer effects specification, assuming that these measures might provide additional information on how peer behavior differs across networks. Our results suggest that including one network measure alone produces rather mixed and unstable results. Including both network variables leads to rather unreliable estimates due to the multicollinearity between the two measures. We interpret the findings from this robustness check as evidence that the network information resulting from the adjacency matrices already contains sufficient information to explain peer behavior.

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 have used a unique network dataset of 85 school classes of secondary schools in Germany, which has allowed us to exploit exogenous variation in second degree friends to identify the endogenous peer effects.

As network-specific factors, we find that the size of the network (i.e., of the 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 matter. In particular, we have shown 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 have shown that the pure size of the class reduces peer behavior and may even lead to negative peer effects in very large classes. Unlike the vast majority of 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 may differ by the subject being taught. For instance, our study sheds light on peer behavior as a specific channel through which class size effects 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.

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.

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. (2020). Peer effects in networks: A survey. *Annual Review of Economics*, *12* (1), 603–629.

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

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.

Chamberlain, G. (1984). Panel data. In Z. Griliches, & M. D. Intriligator (Eds.) *Handbook of Econometrics*, (pp. 1247–1318). Elsevier.

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.

Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica*, *76*(3), 643–660.

Hanushek, E. A. (1996). Measuring investment in education. *The 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.

Hill, A. J. (2015). The girl next door: The effect of opposite gender friends on high school achievement. *American Economic Journal: Applied Economics*, *7*(3), 147–77.

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.

Hoxby, C. M., & Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Mimeograph.

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.

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.

Lee, L. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, *140*(2), 333–374.

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

Liu, X., & Lee, L. (2010). GMM estimation of social interaction models with centrality. *Journal of Econometrics*, *159*(1), 99–115.

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

Lu, F., & Anderson, M. L. (2015). Peer effects in microenvironments: The benefits of homogeneous classroom groups. *Journal of Labor Economics*, *33*(1), 91–122.

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

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

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

Sacerdote, B. (2011). Peer Effects in Education: How might they work, how big are they and how much do we know thus far? In E. Hanushek, S. Machin, & L. Woessmann (Eds.) *Handbook of the Economics of Education*, (pp. 249–277). Elsevier.

Startz, R., & Wood-Doughty, A. (2017). *Improved estimation of peer effects using network data*. Dissertation, Department of Economics, University of California, Santa Barbara.

A Appendix: Tables

Table A1: IV-MD Estimates: German

	Heterogeneous Peer Effects Model		
	Composite	Local-aggregate	Local-average
Local-aggregate peer effect			
Intercept	0.0023***	0.0017**	
	(0.0007)	(0.0007)	
Class Size	0.0001	0.0001	
	(0.0001)	(0.0001)	
Gender Composition	-0.0042***	-0.0030*	
	(0.0016)	(0.0016)	
Local-average peer effect			
Intercept	0.0420		0.1442***
	(0.0259)		(0.0311)
Class Size	0.0229***		0.0148***
	(0.0047)		(0.0052)
Gender Composition	0.2024***		0.2099***
	(0.0521)		(0.0580)
Own characteristics			
IQ	-0.0369***	-0.0531***	-0.0486***
	(0.0082)	(0.0083)	(0.0087)
Previous GPA	0.4203***	0.4319***	0.4185***
	(0.0080)	(0.0082)	(0.0085)
Age	-0.0466***	-0.0498***	-0.0653***
	(0.0088)	(0.0092)	(0.0097)
Exogenous peer effects			
IQ	0.0434***	-0.0103	0.0537***
	(0.0162)	(0.0167)	(0.0169)
Previous GPA	0.0200	0.0111	0.0100
	(0.0189)	(0.0163)	(0.0218)
Age	-0.0061	-0.0330*	0.0343**
	(0.0163)	(0.0181)	(0.0172)
MD statistics (d.f.)	2755.44 (668)	2796.96 (671)	2777.90 (671)

IV-MD estimates for the composite model (first column), local-aggregate model (second column), and local-average model (third column). The IV-MD estimates of the two submodels are based on first stage IV estimates of the submodels. 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.

Table A2: IV-MD Estimates: Math

	Heterogeneous Peer Effects Model		
	Composite	Local-aggregate	Local-average
Local-aggregate peer effect			
Intercept	0.0003	0.0001	
	(0.0007)	(0.0008)	
Class Size	0.0003*	0.0000	
	(0.0002)	(0.0002)	
Gender Composition	-0.0050***	-0.0037*	
	(0.0018)	(0.0019)	
Local-average peer effect			
Intercept	-0.0523*		-0.0001
	(0.0277)		(0.0338)
Class Size	0.0308***		0.0126**
	(0.0050)		(0.0060)
Gender Composition	-0.1302**		-0.1405**
	(0.0548)		(0.0614)
Own characteristics			
IQ	-0.1517***	-0.1568***	-0.1564***
	(0.0100)	(0.0105)	(0.0112)
Previous GPA	0.4160***	0.4047***	0.4294***
	(0.0097)	(0.0102)	(0.0107)
Age	0.0676***	0.0616***	0.0626***
	(0.0105)	(0.0111)	(0.0122)
Exogenous peer effects			
IQ	0.1381***	0.1309***	0.0995***
	(0.0184)	(0.0200)	(0.0220)
Previous GPA	-0.0197	0.0136	-0.0460**
	(0.0202)	(0.0191)	(0.0228)
Age	0.0033	0.0538**	0.0577**
	(0.0215)	(0.0211)	(0.0237)
MD statistics (d.f.)	3188.25 (668)	3246.58 (671)	3197.29 (671)

IV-MD estimates for the composite model (first column), local-aggregate model (second column), and local-average model (third column). The IV-MD estimates of the two submodels are based on first stage IV estimates of the submodels. 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.

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**		0.0015	0.0012	
	(0.0006)	(0.0006)		(0.0012)	(0.0012)		(0.0015)	(0.0014)	
Average Peer Effect	0.8532		0.4181	-0.0946		-0.0277	0.3139		0.1863
	(0.6248)		(0.6822)	(0.6242)		(0.6381)	(0.5991)		(0.6191)
Own characteristics									
IQ	-0.0288***	-0.0292***	-0.0293***	-0.0561***	-0.0563***	-0.0573***	-0.1804***	-0.1779***	-0.1799***
	(0.0072)	(0.0069)	(0.0069)	(0.0147)	(0.0147)	(0.0147)	(0.0199)	(0.0189)	(0.0200)
Previous GPA	0.3608***	0.3637***	0.3623***	0.3944***	0.3942***	0.3943***	0.4009***	0.3988***	0.4000***
	(0.0082)	(0.0074)	(0.0079)	(0.0153)	(0.0153)	(0.0153)	(0.0187)	(0.0184)	(0.0186)
Age	-0.0090	-0.0159**	-0.0130	-0.0616***	-0.0608***	-0.0629***	0.0292	0.0318	0.0294
	(0.0103)	(0.0079)	(0.0102)	(0.0178)	(0.0166)	(0.0178)	(0.0220)	(0.0213)	(0.0219)
Exogenous peer effects									
IQ	0.0321	0.0125	0.0218	0.0268	0.0295	0.0269	0.1264	0.0714*	0.1026
	(0.0224)	(0.0149)	(0.0222)	(0.0359)	(0.0318)	(0.0362)	(0.1145)	(0.0404)	(0.1169)
Previous GPA	-0.2924	0.0300**	-0.1261	0.0798	0.0379	0.0589	-0.1347	-0.0055	-0.0783
	(0.2389)	(0.0147)	(0.2596)	(0.2782)	(0.0303)	(0.2848)	(0.2495)	(0.0361)	(0.2564)
Age	-0.0113	-0.0268	-0.0192	-0.0566	-0.0516	-0.0535	0.0183	0.0276	0.0223
	(0.0217)	(0.0174)	(0.0219)	(0.0497)	(0.0340)	(0.0506)	(0.0473)	(0.0436)	(0.0471)
MD statistics	3228.51	3238.43	3257.12	2801.79	2803.69	2816.28	3248.83	3262.43	3249.16
d.f.	672	673	673	672	673	673	672	673	673

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. 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.