The Key Player in Disruptive Behavior: Whom Should We Target To Improve Classroom Learning Environment?*

Julia Boguslaw†

Draft version: August 20, 2017

Abstract

In this paper, I address the question: who is the individual that exerts the greatest negative influence on the classroom learning environment? To answer this I invoke the key player model from network economics and use self-reported friendship data in order to solve the methodological problems associated with identifying and estimating peer effects. I overcome the issue of endogenous group formation by using the control function approach where I simultaneously estimate network formation and outcomes. The results show that the typical key player scores well on language and cognitive ability tests and is not more likely to be a boy than a girl. I also find evidence that removing the key player has a significantly larger effect on aggregate disruptiveness in a network than removing the most disruptive individual implying policy aimed at the most active individual could be inadequate. Finally, I find that the average model fits the data best suggesting group-based policies should be more effective than policies aimed at specific individuals.

Keywords: key player, spatial autoregressive model, education, disruptiveness

*I am grateful to Eskil Wadensjö, Matthew Lindquist, Yves Zenou, Valerio Leone Sciabolazza and Eirini Tatsi for their advise and support. This paper has benefited greatly from comments and suggestions from Per Engzell and Niklas Kaunitz. I further thank Anna Sandberg Trolle-Lindgren and Kristian Koerselman for their constructive comments. Finally, I thank the seminar participants at The Swedish Institute for Social Research (SOFI), the audience at the EALE conference 2016 and the Social Networks Group at Stockholm University for valuable comments and suggestions. This paper uses the Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU, Kalter et al. (2013)).

†SOFI, Stockholm University. Contact: `julia.boguslaw@sofi.su.se`

1 Introduction

In this paper, I address the question of how disruptive behavior spreads in a classroom. More specifically, I ask: who is the individual that exerts the greatest negative influence on the classroom learning environment? In a world of competing ends and scarce means, this is a question of potentially great relevance namely, if aggregate outcomes can be improved by focusing existing resources on a small number of disruptive peers.

To answer this I invoke the key player model from network economics (Calvó-Armengol & Zenou 2004, Ballester et al. 2006, 2010). Based on a set of behavioral assumptions, this model predicts how much each individual contributes to disruptive behavior in the classroom not just as a function of their own behavior, but also their location in the network as facilitators or inhibitors of the disruptive behavior of peers. I use the socio-metric information on individuals' localities in the network to investigate the structure of the network and how it affects own disruptive behavior.[1] By combining the key player model with a unique data set on disruptive behavior and student networks among eight graders, I can provide novel evidence on how disruptiveness spreads in the classroom. Moreover, an application of the key player strategy in the school context can yield important insights on how to create effective policy interventions in education, for example on how to alter the grouping of students in order to improve the learning environment for all.

Although the field of peer effects is well-established within economics, the empirical evidence concerning peer effects in school outcomes is not conclusive, which can in part be explained by the econometric problems associated with identifying and estimating causal peer effects (Manski 1993, Sacerdote et al. 2011, Angrist 2014). Previous studies on this topic suffer from a number of inferential obstacles like *selection*, the *reflection problem*, or *common shocks*. In addition, research based on observational data often suffers from endogeneity problems. To circumvent these issues, this paper employs a theoretically informed model of peer influence and tests it using the unique classroom network data from Swedish schools. To address the issue of simultaneity, I use instrumental variables arising from the network structure. I overcome the issue of endogenous group formation by using the control function approach where I simultaneously estimate network formation and outcomes (Heckman et al. 2013).

The study draws on recent sociometric data in the longitudinal cohort sur-

[1]It would have been interesting to look at the actual changes in behavior within networks before and after a student has left a class (some students are missing in wave 2 since they have either changed class or school) and to compare the predictions of the key player model to actual outcomes from changing class composition. Due to the small number of missing students in each school year such an exercise will not be possible in this study.

vey Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) from more than 100 schools across Sweden (n=4,794 students), collected when participating students were in the eighth grade (aged 14-15). The respondents have been asked to name the names of their best friends in the classroom. By using network data on students' friendship links and self-reported problem behavior I am able to identify the most disruptive individual in a peer group (network). I use a composite of different measures for problem behavior indicated by survey self-reports of delinquency (e.g. arguing with teacher(s), getting punished and skipping school).

The empirical analysis encompasses three main steps. First, I estimate two standard models of peer effects, the average and the aggregate, using the estimation methods Two stage least squares (2SLS) and Maximum Likelihood (ML). Borrowing from the literature on identification in social networks (e.g. Goldsmith-Pinkham & Imbens (2013) and Hsieh & Lee (2016)), I use an instrumental variable arising from the network structure to arrive at a causal estimate of peer effects. The idea is to use characteristics of the friends of friends, under the assumption that own friends, but not friends friends, are actively chosen (Bramoullé et al. 2009).[2]

I choose the model that fits the data best and use the estimate of peer effects together with the behavioral model to identify the "key player" in terms of classroom disruptive behavior. I identify the key player in a social network as the individual who once removed generates the largest reduction in aggregate disruptive behavior. In the third and final step I calculate the predicted reduction in aggregate disruptiveness from changing class composition, i.e. when the key player is missing. Following Lindquist & Zenou (2014), I first calculate the change in aggregate delinquency by each network. I then create dummies for different types of players: the key player, the most active player and a random player. I focus on the individual who has the highest self-reported disruptiveness level (most active individual) and the individual who once removed generates the largest reduction in aggregate disruptive behavior (key player). Finally, I regress the change in aggregate disruptiveness on these dummies in each network separately. This procedure allows me to address to what extent the key player strategy outperforms alternative policies such as targeting the most active individual.

The contribution of this paper is threefold, First, I provide a micro-founded behavioral model of the contagion of disruptive behavior in the classroom. Second, I measure the size of network effects in disruptive behavior using field data. Third,

[2]I complement this with alternative instruments, including variation in the number of friendship links Lee et al. (2010), Liu & Lee (2010). Individuals have different number of friends and the idea here is that the more friends one has, the higher is the aggregate disruptiveness in one's friendship network. The instrument can only be used in the estimations of the aggregate model of peer effects.

I nail down the type of mechanism at work and resort to a key player simulation in order to pick optimal candidates for treatment. To the best of my knowledge, this is the first study that applies the key player strategy to social networks in education.

I find that the key player and the most active individual is the same person in 28 out of 329 networks (approximately 8.5 percent). Interestingly, the typical key player scores well on the language and cognitive tests and is not more likely to be a boy than a girl. I find evidence that removing the key player has a significantly larger effect on aggregate disruptiveness in a network than removing the most disruptive individual implying policy aimed at the most noisy individual could be inadequate. Based on these results, I suggest alternative strategies on classroom organization to rectify aggregate disruptive behavior.

The paper unfolds as follows. Previous literature is presented in section 2 followed by a description of the model in section 3. In section 4, I present the data and the definitions. I describe the identification strategy and the identification of structural parameters in section 5. The results from the estimations of the peer effects models and the key player simulation are presented in section 6, followed by a discussion of policy implications in section 7. I conclude in section 8.

2 Related literature

In this section, I give an overview of the related literature and present the theoretical framework of this study.

2.1 Peer effects in education

Previous research shows that peers influence adolescent behaviors (see for example Sacerdote et al. (2011) for an overview of the literature). According to standard models of peer effects, influence can occur both through the composition of the classroom, e.g. the average level of parental education among peers (the so-called contextual effect), or through direct interaction with classmates. For example, one student's decision not to disrupt the class can directly influence the behavior of other students in the classroom. In addition, students may respond differently to different categories of peers.

The literature on peer effects in education suggests several plausible models of peer effects: the *bad apple*, *shining star*, *average* and *aggregate model* among others and the behavior mechanisms of these models point to different policy implications. For example, the average model suggests policies that aim at changing the group norm while the bad apple or the shining star models imply individual-based rules targeting students in the extreme parts of the ability distribution. In this paper, I compare the average and the aggregate model of peer effects.[3] This comparison is informative since it tells us whether it is the sum of friends disruptive behavior or the norm, i.e. the average disruptiveness among friends, that best describes peer effects in disruptive behavior and to what extent policy should be aimed at targeting the most active or the most central individual. Below, I describe the average and the aggregate model in more detail.

According to the *average model*, or the so-called standard linear-in-means model, individual outcome may be affected by the mean outcome of the peer group (endogenous), individual characteristics (exogenous), the mean characteristics of the peer group (exogenous) and unobservable correlated effects at the group level. Peers can set norms of conduct and exert social pressures for or against misbehavior and this model incorporates a cost from deviating from the social norm; individuals may be penalized if they deviate from the average activity of the reference group (see e.g. Liu et al. (2014) for a discussion on the social conformity effect). If students tend to conform to the social norm, then policy should be aimed at the majority in the classroom to promote desirable behavior.

One of the underlying assumptions of this model is that the peer effect is the same for all members of a given peer group. However, this assumption may be erroneous as the spillover effects may be larger for some categories of students than for others. In addition, the effects of peers may operate non-linearly or through moments other than the mean. A number of papers have recently addressed this issue by trying to estimate different types of heterogeneous peer effect models. Overall, the findings are mixed; while some studies reject the linear-in-means model (see in particular Hoxby & Weingarth (2005)), others provide evidence in favor of the model when compared to individual-based models such as the bad apple or the shining star model (Liu et al. 2014, Tatsi 2015). Hoxby & Weingarth (2005) find that students seem to benefit from interacting with classmates at the top of the ability distribution while Tatsi (2015) finds support for the linear-in-means model, implying that students tend to conform to the classroom norm.

The *aggregate model* suggests that it is the peers sum outcome that matters for individual outcome. Furthermore, this effect may be multiplied by the number of students engaging in disruptive behavior. For example, one students decision not to disrupt the class can directly influence the behavior of other students in the classroom. This mechanism is the so-called social multiplier. Put simply, the model predicts that the more friends an individual has, the higher the sum of friends activity and the higher is individual activity. If it is the complementarities of friends

[3] Liu et al. (2014) also compare the average and the aggregate model but in contrast to this study, they examine the interaction between the variables study effort and sport activity using the National Longitudinal Survey of Adolescent Health (AddHealth) survey.

behavior that affects individual outcome, i.e. if students are more influenced by high-status peers rather than for example the most active individual, then the aggregate model should be more relevant in explaining peer effects in disruptive conduct.

2.2 Disruptive classroom behavior

Although prior work on spillovers in eduction is extensive, literature on student misbehavior and its dynamics remains fairly unexplored. Due to both observed and unobserved heterogeneity across schools and classrooms and the complex nature of social interaction, obtaining credible estimates of peer effects is particularly challenging. The social dynamics of the classroom are complex as defiance of teacher authority can be either overt or covert (McFarland 2001). Moreover, the rules on classroom interaction vary across schools and classrooms.[4] The same applies to teacher sanctions which may vary in form (formal and informal). A rather popular method of dealing with the endogeneity issues in studies on peer effects in the school setting is to exploit year-to-year variation in peer composition in schools in order to identify a causal influence of peers on individual outcome.[5]

The recent study of Kristoffersen et al. (2015) makes use of the variation in peer composition in school-cohorts to estimate the influence of peer quality on individual academic achievement. The researchers exploit the entry of disadvantaged children, or so-called "potentially disruptive peers", to identify the peer effect in reading test scores. Three categories of children are of particular interest: children with divorced parents, children with criminal parents and children with a psychiatric diagnosis. They find significant and robust effects on peers' academic achievement in reading when a new potentially disruptive student is enrolled in a school.[6] A related study of Carrell & Hoekstra (2010) investigates the influence of children from troubled families on peers' test scores in maths and reading and in deviant conduct.[7] The authors exploit the variation within families to arrive at a credible estimate of peers' behavioral externalities. They use childrens' school records matched with domestic violence cases and find a significant effect of being exposed to a child from a troubled home. The effect is mainly driven by boys and children from low-SES families. According to the authors, the results provide evidence in support of the "bad apple"

[4]Group sizes are also important. See for example Lazear (1999), McFarland et al. (2014), Roman (2016) and Frank et al. (2013).

[5]A large strand of the literature (Black et al. 2013, Hoxby 2000, Gould et al. 2009) use idiosyncratic variation in peer characteristics across cohorts.

[6]The authors also find heterogeneous effects. The effect seems to be strongest when the new student is a child with a psychiatric diagnosis.

[7]See also Carrell et al. (2016) who show that there are long-run consequences of being exposed to a disruptive peer. The authors apply the same identification strategy as in Carrell & Hoekstra (2010).

model of peer effects.

Contrary to prior work based on observational data, I approach the issue of disruptive behavior in the classroom by investigating the architecture of classroom networks. By using a networks approach to this topic, I can identify the transmission channels of teenage group pressure, thus generating new insights on how adolescent behaviors spread in the classroom.

Are boys more susceptible to peer pressure in disruptive behavior than girls? It is possible that teachers reorganize their classrooms in a fashion that disconnects networks of misbehaving students, for example groupings of boys where the peer effect or group pressure to rebel against the teacher is strong. A long-established strategy is to place boys in the front row or next to girls based on an alternating gender rule.[8] The purpose of a rule as such is to restrain boys from disruptive conduct which suggests that the baseline disruptiveness and/or the peer contagion effect is stronger among boys than girls. Studies like for example Hoxby (2000) and Lavy & Schlosser (2007) examine the effect of the gender composition of the classroom on school outcomes and their findings suggest that both sexes perform better in school in classrooms with a higher proportion of girls.

In this paper, I study the observable characteristics of the key player and examine the notion that boys are more often facilitators of problematic behavior than girls. I assume that the relevant peer group is the direct friendship network: the decision to disrupt depends on the social values of one's friends rather than a random disruptive individual in the classroom.[9]

2.3 The key player

While network measures of centrality have long been used in the sociological literature (see for example Wasserman & Faust (1994)), the issue of identifying key players in networks was first introduced by Borgatti (2006, 2003). Previous studies on social networks and behavior have mainly applied the key player strategy to networks of juvenile delinquency Liu & Lee (2010) and co-offending networks (Lindquist & Zenou 2014). In the studies of Ballester et al. (2010), Ballester & Zenou (2014), the key player is defined as the individual who once removed generates the greatest reduction in aggregate crime.

The idea behind the key player strategy is to aim interventions at key individuals.

[8] In Swedish this strategy is called varannan flicka och varannan pojke ("everyother girl and everyother boy").

[9] Presumably, it is not the behavior, in this case the level of disruptiveness *per se*, that influence individual choices but the social values and norms held by one's peers (for example unobservable effort). Fruehwirth (2013) and Boucher & Fortin (forthcoing) draw attention to the importance of modeling the proxy and the "true interaction variable" separately.

According to the key player theory, removing the key player can have substantial effects on adolescent behavior because of social multipliers (Zenou 2016). By lowering the disruptive behavior of central individuals with many social connections, the sum of the disruptiveness among their friends is reduced through both a direct and an indirect effect. The direct effect being the individual's own disruptiveness and the indirect effect being the effect of that individual's behavior on other students in the network (the social multiplier effect).

The literature on social networks in education is relatively scarce (important exceptions include Calvó-Armengol et al. (2009), Bifulco et al. (2011), Patacchini et al. (2016) and Hsieh & Lee (2016)). Apart from the studies of Calvó-Armengol et al. (2009) and Hahn et al. (2015), I am not aware of any other paper that tries to identify the key player in a classroom setting. The scarcity of previous research on social networks in this field is partly due to lack of detailed network data on schools and classrooms. This paper picks up where Calvó-Armengol et al. (2009) left off and provides the first illustration of how of the key player strategy could be applied in educational settings.

2.4 Contribution

The first main contribution of this paper is to the literature on peer effects. In contrast to the majority of peer effects studies which base their empirical analysis on observational data, I use self-reported friendship data in order to solve the methodological problems associated with identifying and estimating peer effects. While it is difficult to construct a research design that convincingly captures the causal effect of peer effects, the theoretically informed model of peer influence presented in this paper and the unique network data in CILS4EU enable me to provide credible estimates of peer effects on adolescent misbehavior. The second contribution is to the literature on social networks in education. To my knowledge, this is the first study that applies the key player strategy to social networks in a school setting. In the spirit of Lindquist & Zenou (2014), I identify the key player in educational networks and discuss optimal targets for treatment. The third contribution is to the literature on disruptive behavior. It is the first study that explicitly models disruptive behavior in the classroom.

3 Model

In this section, I present two models of peer effects and derive the respective model equilibrium. Thereafter, I present the key player strategy.

I adopt the network model of peer effects of Calvó-Armengol et al. (2009). Following Lindquist & Zenou (2014), I present two separate utility functions for the aggregate and the average model of disruptiveness.[10] In the aggregate model, each agent chooses his or her level of disruptiveness, y_i, proxied by problem behavior in order to maximize own utility $u_i(\cdot)$, which is an increasing function of the "gains" of disruptiveness $(a_i+\eta+\epsilon_i)$, the disruptiveness of other students in network $\mathbf{y} = (y_1, ..., y_n)'$, the social cost or stigma of being punished by the teacher $-\frac{1}{2}y_i^2$, and the adjacency matrix $\mathbf{G}$ capturing the friendship network. The parameter ϕ captures the strength of the complementarities.

$$u_i(\mathbf{y}, \mathbf{G}) = (a_i + \eta + \epsilon_i)y_i - \frac{1}{2}y_i^2 + \phi \sum_{j=1}^{n} g_{ij} y_i y_j \tag{1}$$

Moreover, each individual has his or her own disruptive ability a_i which depends on one's observable attributes, the average observable characteristics of one's friends, and the total number of friends indicated by g_i.

$$a_i = \mathbf{x_i}\beta_\mathbf{1} + \frac{1}{g_i}\sum_{j=1}^{n} g_{ij}\mathbf{x_j}'\beta_\mathbf{2} \tag{2}$$

The individual characteristics are captured by β_1 while β_2 represents the contextual effects.[11] ϵ_i represents idiosyncratic shocks and η are network fixed effects which capture the environment at the network level.

Analogously, the average model of peer effects is the following:

$$u_i(\mathbf{y}, \mathbf{G}) = (a_i + \eta + \epsilon_i)y_i - \frac{1}{2}y_i^2 - \frac{1}{2}\lambda \left(y_i - \sum_{j=1}^{n} g_{ij}^* y_j \right)^2 \tag{3}$$

where g_{ij}^* is the row-normalized adjacency matrix and the parameter λ captures the strength of social-conformity.

The difference between (1) and (3) is the last term. In the average model individuals are influenced by the social norm. There is a punishment (a cost) for deviating from the social norm which is increasing with the distance from the average activity among one's peers, as indicated by the expression $\left(y_i - \sum_{j=1}^{n} g_{ij}^* y_j \right)^2$. The parameter λ measures the strength of conformism in a network. In the aggregate model, an increase in the total disruptiveness of one's reference group increases individual marginal disruptiveness, represented by the expression $\sum_{j=1}^{n} g_{ij} y_i y_j$.

[10] In this subsection I closely follow Lindquist & Zenou (2014).

[11] See Lindquist and Zenou (2014) for a detailed discussion of the model.

3.1 Model equilibrium

In equilibrium each agent chooses y_i, one's own level of disruptiveness, in order to maximize utility $u_i(\mathbf{y}, \mathbf{G})$. The choices are made simultaneously by all agents. Thus, agent i's best-reply function in the aggregate model is:

$$y_i^* = \phi \sum_{j=1}^{n} g_{ij} y_j + a_i + \eta + \epsilon_i \tag{4}$$

where a_i is defined above. Let $\alpha_i = a_i + \eta + \epsilon_i$ for each agent i and α be a vector (non-negative) keeping track of all α_i. Also, let $\mu(\mathbf{G})$ denote the spectral radius of $\mathbf{G}$.

The best-reply function in the average model is:

$$y_i^* = \frac{\lambda \sum_{j=1}^{n} g_{ij}^* y_j + a_i + \eta + \epsilon_i}{(1+\lambda)} \tag{5}$$

For notational simplicity, let $\phi = \frac{\lambda}{(1+\lambda)}$, the social conformity coefficient in the average network game. Analogously, let $\alpha_i = \frac{a_i+\eta+\epsilon_i}{(1+\lambda)}$ for each agent i and α be a vector (non-negative) keeping track of all α_i.

Proposition 1 Consider a disruptiveness game where the utility function of each agent i is given by (1) with $a_i > 0$ for all i defined by (2). If $\phi\mu(\mathbf{G}) < 1$, then the game has a unique Nash equilibrium in pure strategies given by:

$$\mathbf{y}^* = \mathbf{b}_\alpha(g, \phi) = (\mathbf{I} - \phi\mathbf{G})^{-\mathbf{1}}\alpha. \tag{6}$$

See proof in Calvó-Armengol et al. (2009).

Proposition 1 (Ballester & Zenou 2014) says that at the Nash equilibrium, each agent's disruptiveness is proportional to her weighted Katz-Bonacich centrality. The influence is heterogeneous as a result of the locational differences of individual agents in the network. Both direct and indirect friendship ties matter, but more connected agents are given a higher weight.

3.2 The key player strategy

The key player in a social network is defined as the individual who once removed generates the largest reduction in aggregate disruptive behavior. Hence, the planner solves the following problem:

$$\max y^*(g) - y^*(g^{-i})|i = 1, ..., n \tag{7}$$

where $y^*(g)$ is equal to the aggregate level of disruptiveness in network g and $y^*(g^{-i)}$ the aggregate disruptiveness once individual i is removed. The maximization problem (7), or the so-called key player strategy, involves identifying the individual who contributes most to the aggregate disruptiveness in the network.

The key player strategy is generally applied to the aggregate model. Below, I outline how I define the key player in the aggregate network game. At this point two assumptions are in order. First, I assume that the adjacency matrix $\mathbf{G}$ is fixed. Second, I assume that individual disruptive ability denoted a_i in (2) is unrelated to $\mathbf{G}$.

As a measure of centrality, I use the *Bonacich network centrality* (Katz 1953, Bonacich 1987). To identify key players in networks I use the Bonacich centrality measure and a concept called contextual intercentrality defined as below.

Definition 1 Given a vector $\mathbf{u} \in \mathbb{R}$, and a small enough scalar $\phi \geq 0$, the vector of Bonacich centralities of parameter ϕ in network g is defined as:

$$\mathbf{b_u}(g, \phi) = (\mathbf{I} - \phi\mathbf{G})^{-1}\mathbf{u} = \sum_{k=0}^{\infty} \phi^{\mathbf{k}}\mathbf{G}^{\mathbf{k}}\mathbf{u}$$

The Bonacich centrality of agent i is constructed as the sum of all paths between agent i and all $j \in 1, ..., n$ of length 0 to k. Each path of length k is weighted by ϕ_k. This number is then multiplied by u_i.

Definition 2 For all networks g and for all i, the contextual intercentrality measure (Ballester & Zenou 2014) of agent i is:

$$\begin{aligned} d_i(g, \phi) &= B(g, \phi) - B(g^{-i}, \phi) \\ &= \mathbf{\Gamma_n'}\mathbf{M}\alpha - \mathbf{\Gamma_n'}\mathbf{M}\alpha^{\mathbf{i}} - \mathbf{\Gamma_n'}\mathbf{M^i}\alpha^{\mathbf{i}} \\ &= B(g, \phi) - B(g^i, \phi) + \frac{b_{\alpha[i],i}(g, \phi)\sum_{j=1}^{n} m_{ji}(g, \phi)}{m_{ii}(g, \phi)} \end{aligned} \tag{8}$$

$B(g, \phi)$ corresponds to the total Bonacich intercentrality in network g while $B(g^{-i}, \phi)$ is the total intercentrality once agent i is removed from the network. An agent i^* is the key player that solves the planner's problem in (7) if and only if i^* is the agent with the highest contextual intercentrality $d_i(g, \phi)$.[12]

Given that individuals are *ex ante* homogeneous, network location is irrelevant

[12]See proof in Ballester & Zenou (2014).

in the average model. Lindquist et al. (2015) provide the first study that includes an application of the key player strategy for the average model.[13] When individuals are identical with respect to their observable characteristics, which individual to target will not matter unless her locality in the network has the feature of a bridge, i.e. the removal of this agent will give rise to isolated individuals (Liu et al. 2014).

Although there is no analytical solution to (5) in the average network game if individuals are heterogeneous, as noted by Lindquist et al. (2015), it is still possible to identify key players numerically using the estimated parameters in the best reply function and the first line of equation (8).

An application of the key player player strategy in the average network game is possible in the case outlined in this paper since the friendship networks are incomplete, i.e. individuals are not fully linked with each other. This means that there will be variations in the connectedness and the localities of individual agents as well as individual heterogeneity in disruptiveness which will be captured by the social multiplier.

4 Data and descriptives

In the following section, I describe the data and present some descriptive statistics.

4.1 Sociometric data

The data set I use, Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU, Kalter et al. (2013)), is a new, longitudinal cohort survey conducted in four countries: England[14], Germany, the Netherlands, and Sweden. The sample is designed to be nationally representative in each country and was created using a stratified three-stage design, interviewing students in sampled school classes. Schools were stratified according to the proportion of children of migrant background, thus the sample contains an overweighting of schools with a high number of children with foreign-born parents. Since these schools tend to be located in areas of concentrated economic disadvantage where classroom disruptive behavior is also more widespread, the sample is congenial with my purposes.

CILS4EU data entails several advantages compared to the data used in previous studies. First, it includes detailed information on the survey participants' friendship links and negative nominations in 249 Swedish classrooms (4,794 students in total). CILS4EU includes not only in-school friendship nominations but also outside-school nominations (not sociometric). Second, the best friend questionnaire included in

[13] Liu et al. (2014) gives some examples.

[14] Only England took part in the UK.

CILS4EU contains additional information on the characteristics of friends outside of school (see questionnaire items in Appendix C).[15]

The stratified sample allows detailed analyses of the social integration of immigrant children specifically, a group of great interest given the increased importance of immigration in Western countries. Immigrant children and children with an immigrant background lag behind children of native-born in educational performance. Foreign-born students are, for example, less likely to be eligible to attend upper secondary school than their native-born counterparts, but tend to make more ambitious study choices given attained school grades (see Arai et al. (2000), Jonsson & Rudolphi (2011) and Heath & Brinbaum (2014)).

The first wave of CILS4EU was performed in the school year 2010-2011 when participating students were in the eighth grade (ages 14-15). The number of respondents in the main questionnaire in the school year 2010-2011 was 5,025 and the response rate was about 86 percent. I use the Swedish sociometric classroom data (n=4,794) which was collected in the first wave of CILS4EU. [16] I define friendship on the basis of the question "Who are your best friends in this class?" to which the student could nominate a maximum of five individuals. A link between two students exists if either of them, or both, nominated the other as a "best" friend. Thus, I treat the network as undirected (although an interesting extension in future work may be to allow for directed networks).[17]

Students who were absent on the day of the network questionnaire or who refused to participate were excluded from the class roaster and the set of potential friend nominees. Individuals who did not nominate anyone have been dropped from the friendship network analysis (see Appendix A for more details on data creation procedures). Due to these restrictions the sample is reduced to 4,219 observations.

Figure 1 plots the distribution of the number of links per actor, the so-called degree centrality. The visible drop at 5 on the x axis is explained by the maximum number of possible nominations; those with degree greater than 5 have at least one incoming nomination that is not reciprocal.

[FIGURE 1 HERE.]

[15]To my knowledge, the only comparable data set to CILS4EU in both survey design and size is the AddHealth data set which includes longitudinal sociometric classroom data in the US.

[16]The advantage of using Swedish data compared to data from the other participating countries in CILS4EU is that there is no formal tracking within Swedish compulsory school system (grades 1-9). Hence, one would expect there to be less formal sorting of students according to ability than in for example Germany with relatively early tracking procedures.

[17]Although it has been argued that a non-response rate of more than about 75 percent could risk the reliability of the nomination measure (see for example Hjalmarsson & Mood (2015) and the references therein), I keep all the classrooms in the analysis for efficiency reasons. See Appendix B for robustness checks.

4.2 Network properties

A classroom network is an $n \times n$ matrix g with generic element g_{ij}. The relationship between any two actors (i, j) is mapped by their value of $g_{ij} \in \{0, 1\}$ where $g_{ij} = 1$ if i and j are friends, 0 otherwise. I assume that links are reciprocal, i.e. $g_{ij} = g_{ji}$. Each network is represented by an adjacency matrix $\mathbf{G}$. The friends of friends adjacency matrix $\mathbf{G}^2$ is derived by multiplying $\mathbf{G}$ by itself, $\mathbf{G}^3$ is the adjacency matrix cubed and so on. Hence, $\mathbf{G}^k$ holds the number of walks of length k.

The *degree* of actor i, denoted $\gamma_i(g)$, is defined as the number of friends that i is directly linked to, and is equivalent to the number of 1's in row (column) i of g. A *walk* is a sequence of links or edges. A *path* exists between actors i and j either if they are directly linked, and/or if there exists a chain of individuals h such that $g_{ih_1} = g_{h_1h_2} = \cdots = g_{Hj} = 1$. The *distance* between two actors is the shortest such path that it takes to reach j from i, which I write $\delta_{ij}(g)$.

All these measures have corresponding expectations at the network level. Thus, I define the average degree of a network g as $\gamma(g) = \sum_{i=1}^{n} \gamma_i(g)/n$. Similarly, the average distance of g is taken over all possible pairings as $\delta(g) = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij}(g)/n(n-1)$.

The *degree centrality* measure is defined as the number of links of an actor. Figure 1 shows the distribution of degree centrality in the Swedish classroom data. The *betweenness centrality* is defined as the number of times an actor i lies along the shortest path (*geodesic*) between two other actors (all other pairings). Actors with high betweenness centrality can act as gatekeepers by connecting isolated parts of the network. The *eigenvector centrality* of an actor is defined as:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{j=n} g_{ij} x_j$$

and depends not only on the number of links but also on the quality of those connections. Two actors can have the same number of links but considerably different eigenvector centrality scores. An actor that is connected to more central nodes (i.e. high-quality connections) will be more "important" in the network and thus outrank others. x is an eigenvector of the adjacency matrix and $\lambda > 0$ is the corresponding eigenvalue (a constant). A related centrality measure is the *Bonacich network centrality* (Katz 1953, Bonacich 1987) previously defined in section 3.2 on the key player.

4.3 Descriptive statistics

Table 1 shows descriptive statistics for selected variables in the data set. The underlying questionnaire items are described in greater detail in Appendix C. The analysis sample consists of 4,219 individuals and 374 networks. Half of the sample is male and approximately 68 percent have two native-born parents. The sample includes individuals who have nominated others and have themselves been nominated. Students with no friendship links have been dropped.

[TABLE 1 HERE.]

The disruptiveness measure is created using the question: "How often do you... (Every day, Once or several times a week, Once or several times a month, Less often, Never) *(i)* argue with a teacher, *(ii)* get a punishment in school (for example being kept in detention, being sent out of class, writing lines), *(iii)* skip a lesson, and *(iv)* come late to school?". The response options are coded as 1 (Never) to 5 (Every day). The imputed disruptiveness measure is thus a summed index of the four delinquency behavior dummies presented in table 1.[18] The minimum score on the disruptiveness index is 4 and the maximum is 20. Individuals with missings on all the underlying variables of the imputed disruptiveness measure have been removed (in total 12 students).

An important question is what the actual underlying distribution of disruptiveness is as this is going to matter for the treatment. Is a small change in friends' disruptiveness associated with a large or a small change in individual disruptiveness? Alternative versions of the disruptiveness measure include the first principal component from a factor analysis and the average of the top four delinquency variables.[19]

Figure 2 shows the distribution of disruptiveness. The distribution is skewed to the right and has a mean of 6.4. The distribution of friendship networks in the sampled classrooms is shown in figure 3.

[FIGURE 2 HERE.]

[FIGURE 3 HERE.]

Figure 4 depicts the architecture of a classroom network with 27 students. Table 2 reports selected network characteristics. The largest network consists of 28 students and the smallest of 3.[20] The mean network size is roughly 16. The average number of links (undirected) is roughly 4. The Katz-Bonacich measure ranges from 7.5 to approximately 15.0.

[FIGURE 4 HERE.]

[18] The index is created using the full sample.

[19] I have documented how the effect changes (sign, magnitude and significance) depending on the definition of disruptiveness and the analysis is available upon request.

[20] Networks with less than 3 members have been removed from the key player simulation.

[TABLE 2 HERE.]

I use a dummy variable to indicate the gender of a student (1=male). The variable HISEI is defined as the highest index of occupational status of parents.[21] Throughout the main analysis of this paper, I define children of immigrants as children with both parents born abroad regardless of own birthplace. The immigrant background variable is based on students' questionnaire answers about their parents' region of birth. The reference category consists of students with at least one native-born parent.

The CILS4EU data include individual scores on both a cognitive and a language test. The two tests were performed in the first wave of the survey during the school year 2010-2011.[22] The language test is a test of a child's lexicon of Swedish antonyms. The test includes 30 items with 4 alternatives each (for more information, see the technical report by Kruse & Konstanze (2016)). The cognitive test is "language free" and as such does not require any particular language skills. It is a 7 minute multiple-choice test of graphical puzzles including 27 items with properties similar to Raven's Progressive Matrices (Raven 2003). The maximum score is of this test is 27 and the minimum is 0.

5 Empirical strategy and identification

In this section, I describe the identification strategy along with the identification of structural parameters.

5.1 Econometric model

I employ two econometric network models: the average the aggregate.[23] The econometric equivalent (written in matrix form) of the best reply function for disruptive

[21]ISEI stands for International Socio-Economic Index of Occupational Status. The variable indicates the maximum based on the variables $y1_i seifG$ and $y1_i seimG$ indicating the occupational status of mother and father respectively. Individuals with missing values on the variable indicating HISEI (272 cases) have been given the sample average.

[22]In the analysis these variables are treated as exogenous, however, since they are measured at the same time as the outcome variable individual disruptiveness they could be endogenous. In an ideal setting, these would be constructed with a lag as is common in the network literature.

[23]There are applications of the key player strategy (see for example Lindquist et al. (2015)) that employ a hybrid model of peer effects where both adjacency matrices are included in the same model estimation. A potential issue here is that the two matrices could be collinear, i.e. one is a linear combination of the other. To circumvent this problem, Tatsi (2015) transforms the adjacency matrices. However, even if one of the matrices comes out as more important than the other it is still impossible to rule out collinearity. Hence, in this study I test the models separately.

behavior in the aggregate model specified in (4) is the following:

$$Y_r = \phi G_r Y_r + X_r \beta + G_r^* X_r \gamma + \mu_{n_r} + \epsilon_r \quad (9)$$

Similarly, the econometric specification of the best reply function for disruptive behavior in the average model specified in (5) is:

$$Y_r = \phi G_r^* Y_r + X_r \beta + G_r^* X_r \gamma + \mu_{n_r} + \epsilon_r \quad (10)$$

where r denotes the network and n_r is the number of observations in each network, Y_r is a $n_r \times 1$ vector of observations of the outcome variable disruptive behavior, X is the $n_r \times k$ matrix of exogenous variables such as age, gender and family characteristics, G^* is the $n_r \times n_r$ row-normalized adjacency matrix that gives the (undirected) connections g_{ij}, $G_r Y_r$ is the $n_r \times 1$ vector of peers' disruptive behavior, μ_{n_r} is the network fixed effect, and ϵ_r is the error term. ϕ, β and γ are the estimated parameters. The adjacency matrix is row-normalized in the average effects model.

5.2 Endogenous network formation and correlated effects

Networks are formed endogenously: who our friends are is not at all random but contingent on both our own characteristics and those of our friends. The famous proverbial expression friends of a feather flock together describes the tendency of individuals with similar backgrounds and preferences to associate with one another. Moreover, contextual effects, i.e. the mean characteristics of friends (or any reference group), could be correlated with school effects. Thus, in order to identify a credible peer effect one must first correct for the endogenous sorting of individuals into schools, classrooms and friendship networks. The challenge is to disentangle the effect of the behavior among friends (endogenous effect) from the effect of friends' characteristics (contextual effect) and the influence of the shared environment (correlated effect).

The identification of peer effects rests on the assumption that the socio-matrix **G** is exogenous (or conditionally exogenous). Peer effect models suffer from a Heckman-style selection bias and the adjacency matrix **G** has built-in endogeneity. For one, there is simultaneity in the outcome variable since individuals choose their disruptiveness level simultaneously. Second, friendship networks are formed endogenously, i.e. there is an omitted variable bias.[24] The main threat to the identification strategy employed in this paper is potential unobservable heterogeneity at the individual,

[24] The friendship networks could be formed based on for example individual disruptiveness. Ideally, one would like to use lagged individual characteristics in peer effect estimations, however, the questions that the disruptiveness measure is based on as well as the cognitive and language ability test scores are only available in the first wave of CILS4EU.

school or network level. For example there may exist network-specific factors that are correlated with individual disruptiveness.

I address the issue of simultaneity (Manski 1993) by using instrumental variables (2SLS/GS2SLS) and Maximum Likelihood (ML) estimation. Different instruments are used in the 2SLS approach in order to take care of potential correlated effects. First, I use characteristics of the friends of friends, under the assumption that own friends, but not friends friends, are actively chosen (Bramoullé et al. 2009). Peers' characteristics are used as an instrument for average peer outcomes, i.e. the matrix $\mathbf{G^2X}$ is used as an instrument for $\mathbf{Gy}$. The structural parameters in model (12) can be identified if $\mathbf{I}$, $\mathbf{G}$ and $\mathbf{G^2}$ are linearly independent and if the friendship network between individuals is intransitive (everyone is not friends with everyone).

The second instrument (Lee et al. 2010, Liu & Lee 2010) is defined as the number of friendship ties. Individuals have different number of friends and the idea here is that the more friends one has, the higher is the aggregate disruptiveness, $\mathbf{JGY}$, in one's friendship network. This instrument is only valid in the case of the aggregate model. Following Tatsi (2015) I also use the Best IV as proposed by Lee (2003) and the results are reported in Appendix B. The Best IV performs only marginally better than the "standard" 2SLS.

The ML strategy tackles the problem of simultaneity by modifying the form of the likelihood function in order to control for the autocorrelation between the observations.[25] Moreover, the ML approach requires that the errors are distributed normally.

I overcome the issue of endogenous group formation by using the control function approach of Heckman et al. (2013). I estimate a spatial Durbin model (Elhorst 2010) and a dyadic network formation process (Graham 2015, Arduini et al. 2015).[26] The root of the omitted variable bias problem is potential correlation between the errors in the model explaining individual disruptiveness and individual behavior in friendship link formation.[27] The control function approach is described in further detail in section 5.4 below.

I avoid the reflection problem Manski (1993) since the analysis in this study is based on network data, meaning that the characteristics of direct friends are not the same for all individuals given the incomplete structure of the network. As a result of this, the contextual effects can be isolated from the peer effect. Furthermore, everything that is common at the classroom and the network level, such as the

[25]The introduction of an log Jacobian term in the likelihood function accounts for simultaneity and disallows the regressors to be correlated with the error terms, thus removing the possible bias in the estimates generated by the simultaneity term (Drukker, Prucha & Raciborski 2013).

[26]See Elhorst (2010) for an overview of different spatial dependence models.

[27]This is discussed in detail in Goldsmith-Pinkham & Imbens (2013).

quality of the teacher, is captured by the network fixed effects (see section 5.3 below). Another potential source of bias is measurement error or incomplete information of friendship links. In line with Lindquist & Zenou (2014) I perform a number of robustness checks in order to asses the validity of the results (see Appendix B).

5.3 Network fixed effects

In the analyses I use fixed effects at the network level where the network is defined as subcomponents (igraph terminology) of the socio-matrix **G**. A subcomponent of **G** consists of all individuals that are weakly connected to each other in a classroom. Thus, the reported direct friends of individual i is a subset of i's network. The number of friendship nominations is restricted to 5 classmates. The restriction is not binding. Links are not necessarily reciprocal hence the degree distribution rages from 1 to 13. Moreover, a network in the analysis sample can consist of up to 28 students.

Common shocks like for example environmental shocks may bias the estimates of peer effects. The fixed effects imply that I only explore variation within networks. By doing so I assume that the relevant interactions take place at the network level. I apply a so-called network-mean transformation by multiplying equation (9) by the matrix $J_r = I_{m_r} - \frac{1}{m_r} l_r l_r'$, where I_{m_r} is the identity matrix and l_r is a vector of ones. I do the same for equation (10). This transformation implies that I subtract the network average from each individual-level variable. Hence, the corresponding aggregate model of peer effects with network fixed effects becomes:

$$J_r Y_r = \phi J_r G_r Y_r + J_r X_r \beta + J_r G_r^* X_r \gamma + J_r \epsilon_r \tag{11}$$

Analogously, I arrive at the following network-mean transformed average model of peer effects:

$$J_r Y_r = \phi J_r G_r^* Y_r + J_r X_r \beta + J_r G_r^* X_r \gamma + J_r \epsilon_r \tag{12}$$

5.4 The control function approach

The control function approach consists of two stages: a selection equation and an outcome equation (Heckman et al. 2013, Wooldridge 2015). Individuals tend to exhibit homophily in covariates such as gender, ethnicity and socio-economic background.[28] The link (or "dyad") formation equation comprises of these variables as predictors of friendship ties. In the first step the binary dependent variable "link"

[28]See the seminal work of McPherson et al. (2001) on homophily in social networks.

(1=reported friendship link) is regressed on individual-specific observable characteristics and dyad attributes. In order to qualify as a valid instrument for link formation the exclusion restriction variable(s) should affect the probability of two individuals forming a friendship tie but not the individual decision to disrupt.

I claim that **G** is exogenous once I correct for possible sorting which is done by including the residuals from the link formation estimation in the outcome equation. Moreover, the links are undirected and the selection correction term is at the individual level as in Graham (2015).[29] The link formation process is modeled as follows:

$$g_{ij} = \alpha_0 + \alpha_d|X_i - X_j| + \alpha_c|X_i - X_j| + \alpha_C C_{ij} + \alpha_f|\varphi_i - \varphi_j| \tag{13}$$

where C_{ij} represents the link characteristics, $|X_i - X_j|$ the distance between the observed characteristics (either dichotomous or continuous indicated by d or c), u_r and $|\varphi_i - \varphi_j|$ the distance between the unobserved characteristics between individuals.

The outcome model in the second stage is either the aggregate or the average model as described above including the estimated residuals, ν_n, in the first stage:

$$J_rY_r = \phi J_rG_rY_r + J_rX_r\beta + J_rG_r^*X_r\gamma + J_r\epsilon_r + \nu_n \tag{14}$$

$$J_rY_r = \phi J_rG_r^*Y_r + J_rX_r\beta + J_rG_r^*X_r\gamma + J_r\epsilon_r + \nu_n \tag{15}$$

Since the second stage model includes the residuals from the fist stage the estimated coefficients are plagued with noise from the first stage (Hardin et al. 2002). One way to examine the bias is to use bootstrapping methods. As this procedure is computationally intensive, at this stage I only present the results with robust standard errors and explore how the estimates and standard errors change when the residuals are included in the outcome model. [30]

6 Empirical results

In this section, I compare different models of peer effects and arrive at an estimate of peer effects in disruptive behavior. The estimate is then used in the key player simulation.

[29] See Arduini et al. (2015) for a model with directed links.

[30] An alternative solution is to follow Murphy & Topel (1985) by adjusting the covariance matrix.Murphy & Topel (1985) provide a consistent estimator of the covariance matrix. See also Del Bello et al. (2015).

6.1 Estimated peer effects

Table 3 shows the results from the regressions for the average, the aggregate and the hybrid model of peer effects estimated by standard OLS. Columns (1) and (2) report the results from the average and the aggregate models of peer effects in disruptiveness. The baseline model, the raw hybrid model of peer effects which incorporates both effects of peer spillovers, is shown in column (3). If both effects are not included there is a potential upward or downward bias (Liu et al. 2014). The average peer effect estimate in column (1) is positive and significant ($p<0.01$). Unconditional on individual and friends' characteristics, a one point increase in the average disruptiveness of friends is, on average, associated with a 0.31 point increase in individual disruptiveness (the mean of the dependent variable is 6.36). The corresponding estimate in the aggregate model is shown in column (2). A one point increase in the aggregate disruptiveness of friends is associated with a 0.02 point increase in individual disruptiveness, on average ($p<0.01$). In sum, both effects are positive and significant in the separate models but when they are both included in the hybrid model the estimate for the aggregate peer effect vanishes and loses significance (see column (3)) while the average peer effect estimate remains unchanged (0.31, $p<0.01$)

[TABLE 3 HERE.]

The results from the different specifications in table 3, columns (1)-(3), suggest that the average model explains the data best.[31] Hence, the preferred specification is the average model of peer effects. Column (4) shows the average peer effect conditional on covariates. Interestingly, none of the friendship characteristics are significant.

In column (4) I also include controls of individual and friends average characteristics. The individual characteristics consist of language and cognitive test scores, gender, socioeconomic background, immigrant background and age. In line with expectations, language and cognitive ability test scores are negatively related to individual disruptiveness. Next, I add fixed effects at the network level (table 3, column (5)). Once I control for possible sorting and common environmental factors, the average peer effect estimate changes sign and loses significance (-0.10, $p<0.10$). A plausible explanation for this result is that too much variation has been lost by introducing the network fixed effect.

The purpose of this exercise is essentially a matter of identifying the transmission mechanism. The question is whether it is operating among direct friends, the

[31] However, the results should be interpreted with caution. A potential issue here is collinearity between the two adjacency matrices.

friendship network (friends of friends) or at the classroom level. The network fixed effect should take care of any extreme cases at the network level. However, if the causal peer effect operates through a channel other than the friendship level, for instance a factor at the classroom level, not controlling for sorting into networks is going to result in a biased peer effect estimate. On the other hand, by introducing a classroom fixed effect the network may capture part of this variation rather than the peer effect estimate at the friendship level. In that case the estimated coefficient could switch sign and still be biased because the effect is carried over to the network level.

Due to simultaneity and omitted variable bias (as discussed in section 5.2) the average peer effect estimate from OLS reported in table 3 is likely biased. In order to address these identification issues I consider two alternative estimation methods: Generalized spatial two-stage least squares (GS2SLS) and Maximum likelihood (ML).[32] A drawback of the ML approach is the restrictive assumptions about the distribution of the error terms. The 2SLS approach will render consistent estimates but in small samples it may be inferior to the biased OLS estimates in terms of goodness of fit (see Anselin (1988)).

Before I estimate the GS2SLS model I need to find a valid instrumental variable. Initially I only consider predetermined characteristics, such as gender, age, ethnicity and parents' socio-economic status. Next, I also include the characteristics of the parents even if the extent to which these covariates are exogenous is debatable.[33] The GS2SLS model is estimated using the spreg command in Stata (sppack).[34] The preferred instrument is then used in the estimations using GS2SLS below. A combination of predetermined individual characteristics and parental attributes results in the highest first stage F-stat, although it is still weak (around 6). A weak instrument could potentially do more harm than good by generating inconsistent estimates and incorrect confidence intervals which is why I extend the analyses with ML estimations.[35]

In the next step, I use the ML and GS2SLS estimators for the parameters of a linear cross-sectional spatial-autoregressive model as suggested by Drukker, Prucha

[32] See Kelejian & Prucha (1998) and Lee (2003). See also Drukker, Peng, Prucha, Raciborski et al. (2013), Drukker, Prucha & Raciborski (2013).

[33] The standard practice is to instrument $\mathbf{G}$ with $\mathbf{G}^2$. However, other instruments are also theoretically motivated (for example $\mathbf{G}^3$ and/or $\mathbf{G}^4$ and/or $\mathbf{G}^3$). Also one could consider parents' characteristics such as marital status, paid job, religion, age, nationality and ISCO 2008.

[34] The output from spreg does not include first stage F-stats hence I try out alternative instruments using two-stage least squares using the Stata package ivreg2. The first stage F-statistic of these estimations ranges between 1 and 6 which is much less than the convention or rule of thumb of at least 10. The results of estimations with the "Best IV" are presented in table B2 in Appendix B.

[35] See Anselin (1988) for a discussion on the finite sample properties of the IV estimator.

& Raciborski (2013). Regression results for each of these estimators for the average model are reported in table 4. The results indicate that GS2SLS is less efficient than the ML.

[TABLE 4 HERE.]

As the spatial-weighing matrix is row-normalized the parameter space of λ is (-1,1). The average peer effect estimate, indicated by λ, is 0.169 and insignificant in the GS2SLS case with network fixed effects (table 4, column (2)), whereas strongly significant when using the ML estimator. Thus, in both cases the peer effect estimate is positive and of moderate size. Column (3) reports the ML results for the aggregate model and the peer effect is highly significant and positive, 0.054 ($p<0.01$). The aggregate model GS2SLS results are found in column (4). The estimate 0.125 is significant ($p<0.01$) but the instrument is invalid (as discussed above). All in all, the results suggest that individual disruptiveness is affected by friends' disruptiveness and that the average peer effect model explains the data best.

Table 4 indicate that in all specifications (columns (1)-(4)) the sign and significance of the individual covariates are consistent. In contrast to the OLS results in table 3, some of the average characteristics of friends are significant. Friends' average language and cognitive test scores are negatively related to individual disruptiveness. In line with expectations, friends' average age is positively related to the outcome variable. Since the model includes spatial lags of the dependent variable the interpretation is less straightforward than in the linear model case. The interpretation of the coefficients for the independent variables are discussed further below.

The two final candidates are the average model and the aggregate model estimated using ML. As a robustness test I compare the Log Likelihood of the average and the aggregate model and they turn out to be almost equal.[36] I continue with the average model since it produces a significant and non-negligible peer effect estimate and is the model that explains the data best as suggested by the results in tables 3 and 4.

Next, I turn to the link formation process reported in table 5. The number of possible links is nearly 18 million. The predictors include the absolute difference in scores on the language test, the difference in scores on the cognitive ability test, male dummy (1=both individuals are male), native dummy (1=both individuals are native-born) and the absolute difference in age.[37] The exclusion restriction

[36] The Log Likelihood for the average model and aggregate model is -9156 and -9159 respectively.

[37] Due to the high non-response rate of both students and parents regarding the parents' occupation I leave out the absolute difference in the highest occupational status of the parents as an explanatory variable in the link formation process.

in the model is an indicator for living within a 5 minute walking distance from a classmate. The geographical proximity variable affects the probability of two individuals forming a friendship tie but not the individual decision to disrupt and should therefore be a valid instrument for link formation. The indicator variable is excluded from the second stage, i.e. the outcome equation.

For the time being, I assume that the errors are following a normal distribution and that they are independent (although there is room to reconsider this). The selection equation is estimated by OLS and the residuals are added up with respect to each individual. Recall that the control function is estimated at the dyad level while the outcome model is at the individual level.

Evidence of the non-randomness in link formation is displayed in table 5. Unsurprisingly, geographical proximity seems to be an important predictor of friendship ties. The estimates reflect probabilities and the coefficient for "5 min distance" is significant and non-negligible. Language and cognitive ability test scores and nativity seem to also be driving friendship formation. The larger the absolute difference in test scores of two individuals the less likely they are to be friends. Homogeneity in terms of region of origin also makes two individuals more likely to form a friendship link.

[TABLE 5 HERE.]

Table 6 reports the outcome equation, namely equation (12) including the selection correction term. Neither the magnitude nor the significance of the peer effect changes by including the selection bias. Furthermore, the size of the standard errors remains unchanged. A plausible explanation for this result is that link formation is as good as random after controlling for sorting using individual and friendship characteristics.[38] Overall the other estimates (and their standard errors) remain fairly unaffected by including the selectivity bias term which suggests that conditional exogeneity holds and that the peer effect can be interpreted in causal terms. Hence, the peer effect estimate that I will use in the key player analysis is 0.167. The preferred model, the average peer effect model, indicates that individual disruptiveness is positively related to the average disruptiveness of best friends.

[TABLE 6 HERE.]

6.2 Interpretation of estimates

The interpretation of the estimates in table 6 is less straightforward than in the OLS case presented in table 3. One way to interpret the coefficients for the independent

[38] See discussion in Del Bello et al. (2015).

variables is to calculate the predicted values at different levels of the dependent variables, as suggested by for example Drukker, Prucha & Raciborski (2013). Due to the built-in simultaneity of the model (SARAR), a change in the dependent variable of one individual can alter the predicted values of all other individuals in the sample. Either the units of the exogenous variable are changed sequentially (average total direct impact, ATDI) or simultaneously (average total impact, ATI). I calculate the predictions using the simultaneous approach. The mean change in the predictions from increasing the individual cognitive ability score by one point is -.0444. The ATI corresponds to about 2.0 percent of a standard deviation in individual disruptiveness (demeaned).[39] The estimated ATI from a one unit change in the individual language test score is -.0224 which corresponds to circa 1.0 percent of a standard deviation in individual disruptiveness.

6.3 Key player simulation

In this section, I proceed by identifying the key player using the concepts presented in section 3. The analysis that follows is based on the average model of peer effects. The estimated peer effect of the average model reported in column (4) in table 6 is positive and statistically significant (0.167, p<0.01).

First, I derive the Katz-Bonacich measure along with the intercentrality (as defined in equation (8)) of each individual using the estimated peer effect of 0.167. I next use all the estimated coefficients in the average model reported in table 6, to derive the disruptive ability a_i of each individual in the network. As defined in equation (2), a_i depends on individual observable attributes, the average observable characteristics of one's friends and the total number of friends. I plug each a_i into the expression (6) and derive the vector of Nash equilibrium disruptiveness levels which corresponds to the Katz-Bonacich of each individual (see Definition 1).

The final part of the exercise involves identifying the key player, i.e. the optimal target. This is done by calculating the intercentrality of all individuals in each network. The key player is the individual with the highest intercentrality. Clearly, the number of key players is the same as the number of networks which is 374. Networks with less than three members have been removed from the key player analysis which leaves us with a total of 329 networks in the analysis sample. Also, the number of most active players is larger than the number of networks since more than one player could have the same level of disruptiveness.

[39] When presented in percentage terms and the denominator is the sample average of individual disruptiveness the absolute ATI from increasing individual cognitive ability by one point corresponds to a very large number. This is because all variables in the preferred specification have been demeaned at the network level and therefore consist of both positive and negative values (including individual disruptiveness).

By definition, key players hold important positions in their network and may act as bridges of both desirable and undesirable behavior.[40] The key player is not necessarily the most active individual in the network. In fact, the key player and the most active individual is the same person in only 28 out of 329 networks (about 8.5 percent). Table 7 shows the observable characteristics of the key player and the most active player. Column (1) reports the results from a logistic regression of a dummy variable, indicating whether an individual is the key player or not, on a selected set of observable characteristics such as gender and parents' immigration background. Column (2) displays the corresponding regression results for the most active player.

According to the results in table 7, the log of odds of a being the key player is positively related language test scores ($p<0.01$) and cognitive ability test scores ($p<0.01$). In other words, the higher the test scores, the more likely it is that an individual is the key player.

The odds ratio of 1.141 indicates that boys are 1.141 times more likely to be the key player but the estimate is insignificant. Thus, I do not find evidence in support of the notion that the key player is more likely to be a boy than a girl; given the same language and cognitive ability test scores, HISEI, age and parents' immigration status, boys are not more likely to be the key player than girls. This seems also be the case for the most active player. Moreover, having two native-born parents is negatively related to one being the key player and positively related to being the most active player (both however insignificant). The log of odds of a being the most active player is negatively related to cognitive ability test scores ($p<0.10$).

[TABLE 7 HERE.]

Next, following the analysis employed in Lindquist & Zenou (2014), I investigate the percentage reduction in disruptiveness from removing the key player, calculated as the intercentrality of the key player times 100 divided by the total Bonacich of that network. I run an OLS regression of this value on a constant and the independent variable network size. The results of these regressions are shown in table 8. I do the same for the most active player and a random player.

[TABLE 8 HERE.]

Table 8 reports the predicted reductions without any baseline. The average reduction in disruptiveness for the average network (size=16) from removing the

[40] Bridges have high betweenness centrality.

key player is roughly 13.2 percent compared to removing the most active player which is about 11.9 percent.[41]

In table 9 the baseline is the most active player or a random player. This approach produces estimates of the performance of the key player strategy relative to other policies such as targeting the most disruptive individual. In the first column of table 9 the dependent variable is the difference in the percentage reduction in disruptiveness from removing the key player compared to removing a random player. In column (2) the dependent variable is the reduction relative the most active player. Networks where the key player and the most active individual or a random player is the same person have been removed from the analysis in table 9 which is why the sample sizes are different in columns (1) and (2).

[TABLE 9 HERE.]

The intercept gives an indication of how much the key player strategy outperforms the other two policies. The key player strategy outperforms the other strategies to a significant extent, although the difference is small: the average reduction in disruptiveness for the average network (size=13.1) from removing the key player is 1.41 percent higher than removing a random player and 1.44 percent higher (size=12.9) than removing the most active player. The estimate of network size is as one would expect negative in both cases.[42] Table 9 shows the relationship between the average predicted reduction and network size. A one point increase in the number of network members is, on average, associated with a 9.7 percentage point decrease in the difference in the average reduction in aggregate disruptiveness.

In summary, the effect of removing the key player is significantly larger than removing the most active player thus removing the most active player is not necessarily the most effective way of lowering aggregate disruptiveness in the network. The difference in the predicted percentage reduction in disruptiveness is however relatively small. Furthermore, the predicted reduction is negatively related to network size which is a mechanical property: removing the key player (or actually any player) in a smaller network will have a larger effect than in a bigger one.

[41] The sample size is the same in all three models since the key player and the most active or random player is allowed to be the same individual.

[42] The number of networks in both column (1) and (2) is less than 374 since the networks in which the key player and a randomly chosen player coincide are removed from the analysis. The same applies to the case when the key player is also the most active player in the network. Also, as previously mentioned, networks with less than three members are excluded from the key player analysis.

7 Discussion

A deeper understanding of how and when peer effects influence adolescent behavior could help both researchers and policy makers create effective policy interventions in education (e.g. how to organize teaching and classrooms optimally) and adolescent risk behavior (e.g. how to reduce delinquent behavior). Should policy be aimed at changing the context (teachers, resources etc.) or the composition of students? Should teachers target the most active individual, i.e. the one making the most noise, or perhaps the most popular individual such as the key player?

Different classroom situations can bring about different behaviors, as noted by McFarland (2001): "changing either the student or the classroom would change the decision to rebel" (p. 617). Disruption could be rectified through organizational changes of the classroom, for example by altering the formats of instruction or the grouping of students.[43] That said, changing classroom size (teacher/student ratio) or introducing remedial classes could be costly compared to altering the groupings of students. The implementation of a policy that changes the configuration of classroom networks of students resistant to learning can prove to be less expensive than other policies and the potential gains could be substantial.

The optimal target for treatment hinges on the underlying behavioral mechanism of disruptive conduct. I find that the average model fits the data best suggesting group-based policies should be more effective than policies aimed at specific individuals. Thus, in order to reduce aggregate disruptiveness the social norm – the behavior of the majority in each network– needs to be changed.

I also find that the key player and the most active individual is the same person in 28 out of 329 networks (approximately 8.5 percent). I find evidence that removing the key player has a significantly larger effect on aggregate disruptiveness in a network than removing the most disruptive individual implying policy aimed at the most noisy individual could be inadequate. However, the difference in the predicted percentage reduction in disruptiveness is small.

Improving the behavior of the worst-behaved (most active) students clearly has a positive effect on other students in the classroom because of the social multiplier. Targeting the most active individuals is likely less demanding than aiming policy at key players. In practice it could be difficult to target key players since they are not as easily identified (compared to the most noisy individuals). An alternative strategy is to "reshuffle" classrooms every semester or school year, thereby potentially changing the classroom norm. A drawback of this approach is that positive spillovers from advantaged to disadvantaged peers could be lost by reorganizing classrooms randomly.

[43] Educators can alter the grouping of student either by mixing, matching or random assignment.

A related question is whether to mix or match students according to specific observable characteristics such as grades. The seminal paper of Lazear (1999) derives optimal class size from a model of educational production that incorporates the disruptive behavior of students in the classroom. Lazear (1999) finds that the effect of classroom size is larger for disruptive than obedient children. From a cost-benefit point of view, reducing the class size by a small number of students may not matter much for individual behavior when the class sizes are relatively large.[44] In Sweden, students often have the same classmates all through the last years of compulsory school hence classroom networks are fairly stable which leaves room for policy on classroom composition.

Are some classroom environments more likely to facilitate or inhibit aggregate disruptiveness? The question opens up new avenues of research on classroom composition and learning environment. Rules on classroom interaction vary across schools and classrooms.[45] Future research could investigate the relationship between the structure of classrooms and specific adolescent undesirable (or desirable?) behaviors. Do classrooms where individuals sort around the most disruptive student stand out in some observable way, for example with respect to density? If so, what makes students in these types of classrooms more susceptible to disruptive conduct? Are the externalities from bad apples larger in dense classrooms? One possibility is to use popularity ranking in the classroom or negative nominations and examine teacher characteristics closer (available in CILS4EU). The next step is to also examine the effect of disruptiveness on individual achievement such as school grades and later educational outcomes.

Finally, this study has a number of limitations that should be mentioned. First, since students who were absent on the day of the network questionnaire or who refused to participate were excluded from the class roaster and the set of potential friend nominees, there is a risk that I underestimate the effect of friends' disruptiveness and the effect of removing the key player (unless these individuals are isolated). As shown in table B1 in Appendix B.2, the cases dropped from the analysis sample due to non-response are more likely to have higher scores on the disruptive measure while lower on the language and cognitive ability tests implying that the analysis sample is positively selected on these characteristics.

Second, a disadvantage of the CILS4EU data is that it is based on individuals' self-reports of problem behavior. Ideally one would like to have data on disruptive behavior collected through classroom observations over time. Furthermore, an

[44] In fact, evidence is inconclusive about the effect of class size on student performance. See discussion in e.g. Hanushek (2002).

[45] Group sizes are important. See for example McFarland et al. (2014), Roman (2016) and Frank et al. (2013).

important question concerns the nature and level of measurement error in the self-reported variables. Is it systematic or random, i.e. do disruptive students tend to misreport their behavior to a larger extent than others? This and related issues could be investigated further using the teacher questionnaire in CILS4EU.[46]

8 Concluding remarks

This paper set out to investigate the peer effect in disruptive behavior using the architecture of the networks in the classroom and to move towards a policy-relevant application of the key player strategy. I find that being the individual that exerts the greatest negative influence on the classroom learning environment is positively related to test scores in cognitive ability and language proficiency. Moreover, the key player is not more likely to be a boy than a girl. I also find evidence that removing the key player has a significantly larger effect on aggregate disruptiveness in a network than removing the most disruptive individual implying policy aimed at the most active and potentially socially isolated individual could be inadequate.

The findings of this study have implications for educational policy on optimal classroom composition. The impact of a policy aimed at key players may prove to be more effective in reducing aggregate disruptiveness and improving the learning environment for all students in a classroom. I suggest a reshuffling policy where students are reassigned to classrooms regularly during the school year along with remedial classes for the most disruptive students.

[46] The sociological study of McFarland (2001) is based on classroom observations of two schools and 36 classrooms followed during two school semesters.

References

Angrist, J. D. (2014), 'The perils of peer effects', *Labour Economics* **30**, 98–108.

Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Vol. 4, Springer Science & Business Media.

Arai, M., Schröder, L. & Vilhelmsson, R. (2000), 'En svartvit arbetsmarknad: en ESO-rapport om vägen från skola till arbete: Rapport till expertgruppen för studier i offentlig ekonomi'.

Arduini, T., Patacchini, E., Rainone, E. et al. (2015), Parametric and semiparametric IV estimation of network models with selectivity, Working paper, Einaudi Institute for Economics and Finance (EIEF).

Ballester, C., Calvó-Armengol, A. & Zenou, Y. (2006), 'Who's who in networks. Wanted: The key player', *Econometrica* **74**(5), 1403–1417.

Ballester, C. & Zenou, Y. (2014), 'Key player policies when contextual effects matter', *The Journal of Mathematical Sociology* **38**(4), 233–248.

Ballester, C., Zenou, Y. & Calvó-Armengol, A. (2010), 'Delinquent networks', *Journal of the European Economic Association* **8**(1), 34–61.

Bifulco, R., Fletcher, J. M. & Ross, S. L. (2011), 'The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health', *American Economic Journal: Economic Policy* **3**(1), 25–53.

Black, S. E., Devereux, P. J. & Salvanes, K. G. (2013), 'Under pressure? The effect of peers on outcomes of young adults', *Journal of Labor Economics* **31**(1), 119–153.

Bonacich, P. (1987), 'Power and centrality: A family of measures', *American journal of sociology* **92**(5), 1170–1182.

Borgatti, S. P. (2003), The key player problem, *in* K. M. C. Ronald Breiger & P. Pattison, eds, 'Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers', National Academy of Sciences Press, pp. 241–252.

Borgatti, S. P. (2006), 'Identifying sets of key players in a network', *Computational, Mathematical and Organizational Theory* **12**(1), 21–34.

Boucher, V. & Fortin, B. (forthcoing), 'Some challenges in the empirics of the effects of networks', *Oxford Handbook on the Economics of Networks, Oxford: Oxford University Press* .

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

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

Calvó-Armengol, A. & Zenou, Y. (2004), 'Social networks and crime decisions: The role of social structure in facilitating delinquent behavior', *International Economic Review* **45**(3), 939–958.

Carrell, S. E., Hoekstra, M. & Kuka, E. (2016), The long-run effects of disruptive peers, Technical report, National Bureau of Economic Research.

Carrell, S. E. & Hoekstra, M. L. (2010), 'Externalities in the classroom: How children exposed to domestic violence affect everyone's kids', *American Economic Journal: Applied Economics* **2**(1), 211–228.

Del Bello, C. L., Patacchini, E. & Zenou, Y. (2015), Neighborhood effects in education, Technical report.

Drukker, D. M., Peng, H., Prucha, I. R., Raciborski, R. et al. (2013), 'Creating and managing spatial-weighting matrices with the spmat command', *Stata Journal* **13**(2), 242–286.

Drukker, D. M., Prucha, I. & Raciborski, R. (2013), 'Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances', *Stata Journal* **13**(2), 221–241.

Elhorst, J. P. (2010), 'Applied spatial econometrics: raising the bar', *Spatial Economic Analysis* **5**(1), 9–28.

Frank, K. A., Muller, C. & Mueller, A. S. (2013), 'The embeddedness of adolescent friendship nominations: The formation of social capital in emergent network structures 1', *American Journal of Sociology* **119**(1), 216–253.

Fruehwirth, J. C. (2013), 'Identifying peer achievement spillovers: Implications for desegregation and the achievement gap', *Quantitative Economics* **4**(1), 85–124.

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

Gould, E. D., Lavy, V. & Daniele Paserman, M. (2009), 'Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence', *The Economic Journal* **119**(540), 1243–1269.

Graham, B. S. (2015), 'Methods of identification in social networks', *Annu. Rev. Econ.* **7**(1), 465–485.

Hahn, Y., Islam, A., Patacchini, E. & Zenou, Y. (2015), Teams, organization and education outcomes: Evidence from a field experiment in Bangladesh, Technical report.

Hanushek, E. A. (2002), 'Evidence, politics, and the class size debate', *The class size debate* pp. 37–65.

Hardin, J. W. et al. (2002), 'The robust variance estimator for two-stage models', *Stata Journal* **2**(3), 253–266.

Heath, A. & Brinbaum, Y. (2014), *Unequal Attainments: Ethnic educational inequalities in ten Western countries*, Oxford University Press.

Heckman, J. et al. (2013), 'Sample selection bias as a specification error', *Applied Econometrics* **31**(3), 129–137.

Hjalmarsson, S. & Mood, C. (2015), 'Do poorer youth have fewer friends? The role of household and child economic resources in adolescent school-class friendships', *Children and Youth Services Review* **57**, 201–211.

Hoxby, C. (2000), Peer effects in the classroom: Learning from gender and race variation, Technical report, National Bureau of Economic Research.

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

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

Jonsson, J. O. & Rudolphi, F. (2011), 'Weak performancestrong determination: school achievement and educational choice among children of immigrants in sweden', *European Sociological Review* **27**(4), 487–508.

Kalter, F., Heath, A. F., Hewstone, M., Jonsson, J., Kalmijn, M., Kogan, I., Van Tubergen, F., Kroneberg, C., Andersson Rydell, L., Brolin Låftman, S. et al. (2013), 'Children of immigrants longitudinal survey in four European countries, GESIS data archive, Cologne, za5656 data file version 1.2.0'.

Katz, L. (1953), 'A new status index derived from sociometric analysis', *Psychometrika* **18**(1), 39–43.

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.

Kristoffersen, J. H. G., Krægpøth, M. V., Nielsen, H. S. & Simonsen, M. (2015), 'Disruptive school peers and student outcomes', *Economics of Education Review* **45**, 1–13.

Kruse, H. & Konstanze, J. (2016), Children of immigrants longitudinal survey in four European countries. Sociometric fieldwork report. Wave 1 2010/2011, v1.2.0. Mannheim, Technical report, Mannheim University.

Lavy, V. & Schlosser, A. (2007), Mechanisms and impacts of gender peer effects at school, Technical report, National Bureau of Economic Research.

Lazear, E. P. (1999), Educational production, Technical report, National bureau of economic research.

Lee, L.-f. (2003), 'Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances', *Econometric Reviews* **22**(4), 307–335.

Lee, L.-f., Liu, X. & Lin, X. (2010), 'Specification and estimation of social interaction models with network structures', *The Econometrics Journal* **13**(2), 145–176.

Lindquist, M. J., Sauermann, J. & Zenou, Y. (2015), Network effects on worker productivity, Working paper.

Lindquist, M. J. & Zenou, Y. (2014), Key players in co-offending networks, Working paper.

Liu, X. & Lee, L.-f. (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 & Organization* **103**, 39–59.

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

McFarland, D. A. (2001), 'Student resistance: How the formal and informal organization of classrooms facilitate everyday forms of student defiance', *American Journal of Sociology* **107**(3), 612–678.

McFarland, D. A., Moody, J., Diehl, D., Smith, J. A. & Thomas, R. J. (2014), 'Network ecology and adolescent social structure', *American sociological review* **79**(6), 1088–1121.

McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001), 'Birds of a feather: Homophily in social networks', *Annual review of sociology* **27**(1), 415–444.

Murphy, K. M. & Topel, R. H. (1985), 'Estimation and inference in two-step econometric models', *Journal of Business & Economic Statistics* **3**(4), 370–379.

Patacchini, E., Rainone, E. & Zenou, Y. (2016), 'Heterogeneous peer effects in education', *Journal of Economic Behavior & Organization* .

Raven, J. J. (2003), Raven progressive matrices, *in* 'Handbook of nonverbal assessment', Springer, pp. 223–237.

Roman, S. (2016), Friendship Dynamics Among Adolescents, PhD thesis, Department of Sociology, Stockholm University.

Sacerdote, B. et al. (2011), 'Peer effects in education: How might they work, how big are they and how much do we know thus far?', *Handbook of the Economics of Education* **3**(3), 249–277.

Tatsi, E. (2015), Endogenous social interactions: Which peers matter?, Working paper.

Wasserman, S. & Faust, K. (1994), *Social network analysis: Methods and applications*, Vol. 8, Cambridge university press.

Wooldridge, J. M. (2015), 'Control function methods in applied econometrics', *Journal of Human Resources* **50**(2), 420–445.

Zenou, Y. (2016), Key players, *in* A. G. Yann Bramoullé & B. Rogers, eds, 'Oxford Handbook on the Economics of Networks', Oxford University Press.

Figure 1: Distribution of degree centrality in the Swedish classroom data. N=4219.

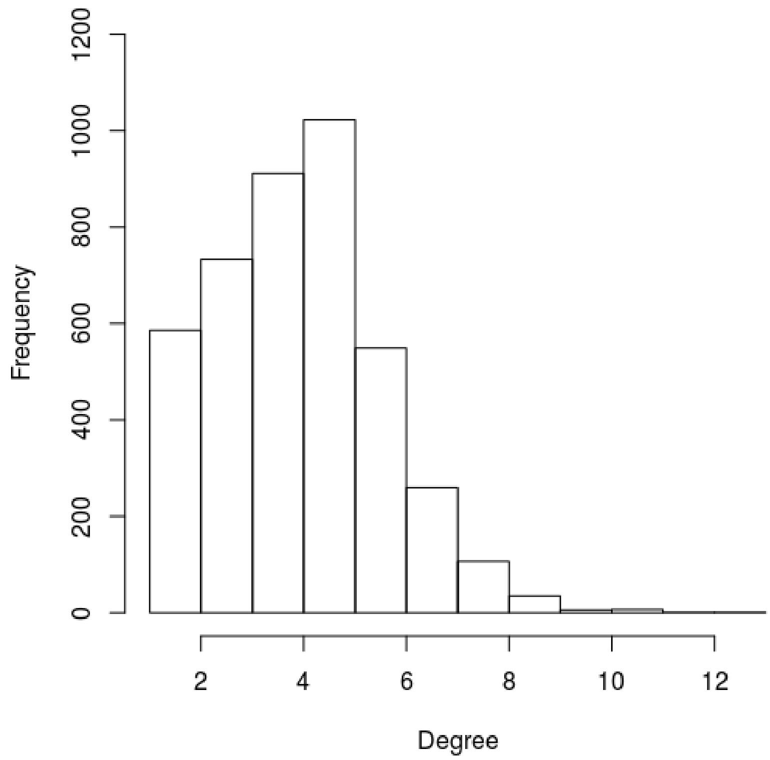

Table 1: Individual level summary statistics.

Variable	Mean	Std. Dev.	Min.	Max.	N
Demographics					
Male	0.486	0.5	0	1	4219
Highest index of occupational status	52.982	20.35	11.74	88.960	4219
Native background	0.677	0.468	0	1	4219
Age	15.029	0.264	13	17	4219
Language test scores	18.654	4.949	0	29	4219
Cognitive ability test scores	17.812	4.751	0	27	4219
Delinquent behavior (1=Never, 5=Every day)					
Arguing with teacher	4.435	0.837	1	5	4209
Getting punished	4.666	0.635	1	5	4204
Skipping school	4.637	0.719	1	5	4196
Late to school	3.9	1.037	1	5	4199
Disruptiveness measure	6.362	2.433	4	20	4219

Notes: Summary statistics on demographics and academic outcomes for the analysis sample and delinquent behavior variables for the full sample.

Figure 2: Distribution of disruptiveness. N=4219.

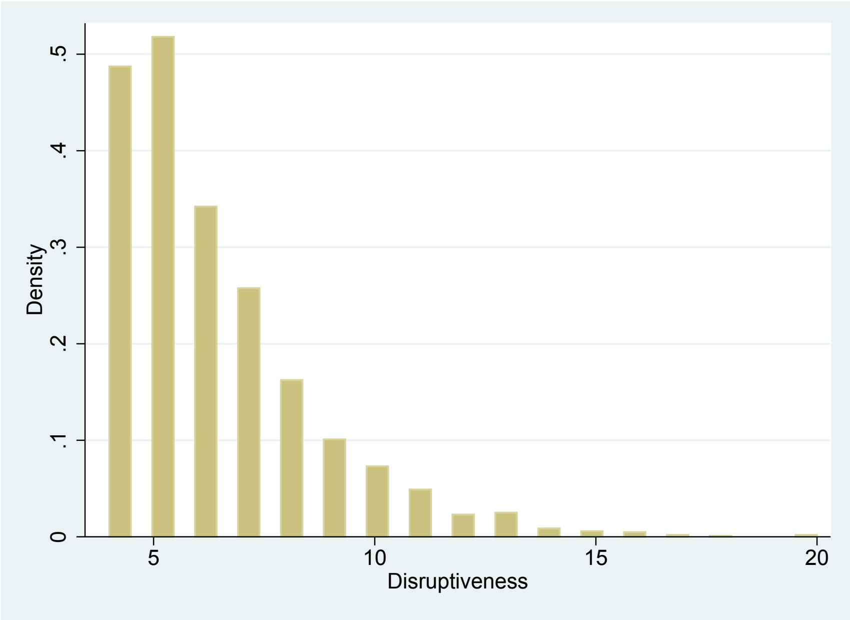

Table 2: Network characteristics.

Variable	Mean	Std. Dev.	Min.	Max.
Network size	15.953	5.92	3	28
Degree	4.433	1.652	1	13
Radius	2.631	0.852	1	4
Eigenvalue	0.571	0.331	0	1
Bonacich	10.3	1.068	7.538	14.977
Betweenness	0.086	0.13	0	1

Notes: Networks with less than three members have been removed from the key player simulation.

Figure 3: Distribution of network size. N=374.

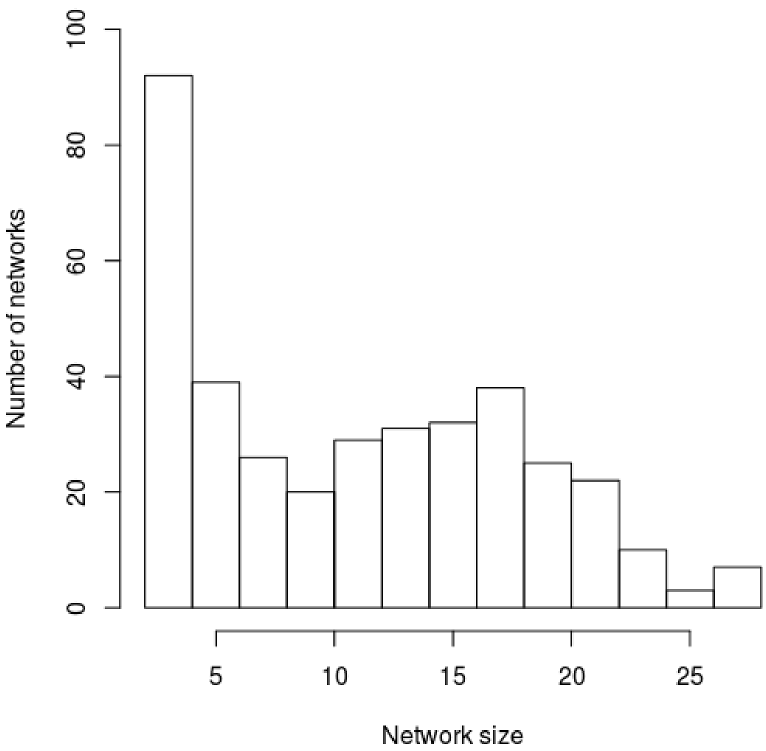

Figure 4: A classroom network of 27 students (undirected links).

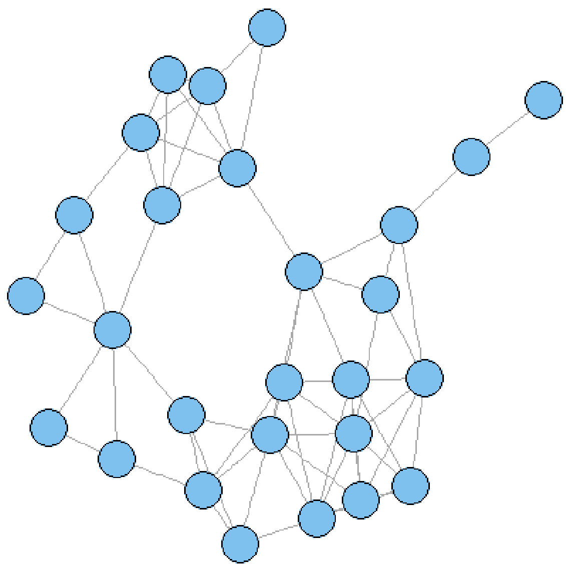

Table 3: Alternative models of peer effects estimated by OLS.

	Average model	Aggregate model	Hybrid model	Average model	
	(1)	(2)	(3)	(4)	(5)
Dependent variable: Disruptiveness					
Intercept	4.39***	5.92***	4.38***	3.54	
	(0.23)	(0.12)	(0.23)	(4.39)	
Average peer effect	0.31***		0.31***	0.30***	−0.10*
	(0.04)		(0.04)	(0.04)	(0.06)
Aggregate peer effect		0.02***	0.00		
		(0.00)	(0.00)		
Language test scores				−0.03***	−0.02**
				(0.01)	(0.01)
Cognitive ability test scores				−0.04***	−0.04***
				(0.01)	(0.01)
Male				0.24**	0.17
				(0.10)	(0.11)
Highest index of occupational status				−0.00	−0.00
				(0.00)	(0.00)
Native				0.18*	0.18
				(0.10)	(0.12)
Age				0.15	0.17
				(0.16)	(0.15)
Missing values: HISEI				0.41**	0.49**
				(0.18)	(0.20)
Friends' average language test scores				0.01	0.02
				(0.02)	(0.02)
Friends average cognitive test scores				−0.00	−0.04
				(0.02)	(0.02)
Proportion male friends				−0.21	−0.04
				(0.13)	(0.15)
Friends' average HISEI				0.00	−0.01*
				(0.00)	(0.00)
Proportion native friends				−0.07	0.09
				(0.15)	(0.23)
Friends' average age				−0.02	−0.12
				(0.27)	(0.35)
Network fixed effects	*NO*	*NO*	*NO*	*NO*	*YES*
Observations	4219	4219	4219	4219	4219
Adj. R^2	0.04	0.01	0.04	0.05	0.14

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: Columns (1)-(3) report the baseline estimates for the average, the aggregate model and the hybrid model of peer effects. Columns (4) and (5) present the results from OLS estimations of the average model including covariates. In column (5) network fixed effects are included. Standard errors are clustered at the network level in all models.

Table 4: The average and the aggregate model of peer effects in disruptive behavior estimated by ML and GS2SLS.

	Average model		**Aggregate model**	
	ML (1)	G2SLS (2)	ML (3)	G2SLS (4)
Dependent variable: Disruptiveness				
Language test scores	-0.0186**	-0.0185*	-0.0190**	-0.0142
	(0.00890)	(0.0102)	(0.00891)	(0.00942)
Cognitive ability test scores	-0.0370***	-0.0370***	-0.0368***	-0.0348***
	(0.00865)	(0.00878)	(0.00866)	(0.00874)
Age	0.208	0.208	0.206	0.193
	(0.133)	(0.134)	(0.133)	(0.134)
Male	0.0474	0.0471	0.0486	0.0306
	(0.150)	(0.151)	(0.150)	(0.150)
Native background	0.220**	0.219**	0.234**	0.224**
	(0.0986)	(0.102)	(0.0986)	(0.0987)
Highest index of occupational status	-0.00164	-0.00164	-0.00161	-0.00173
	(0.00186)	(0.00186)	(0.00186)	(0.00186)
Missing HISEI	0.417***	0.417***	0.423***	0.407***
	(0.147)	(0.148)	(0.147)	(0.147)
Friends' average language test scores	-0.0595***	-0.0593***	-0.0603***	-0.0496***
	(0.0156)	(0.0190)	(0.0156)	(0.0170)
Friends' average cognitive test scores	-0.0384**	-0.0382*	-0.0400***	-0.0283*
	(0.0151)	(0.0199)	(0.0151)	(0.0169)
Friends' average age	0.0746**	0.0743*	0.0762**	0.0503
	(0.0357)	(0.0443)	(0.0357)	(0.0394)
Proportion male friends	0.0696	0.0693	0.0728	0.0581
	(0.172)	(0.172)	(0.172)	(0.172)
Proportion native friends	0.119	0.118	0.130	0.0774
	(0.158)	(0.170)	(0.159)	(0.162)
Friends' average HISEI	1.46e-05	1.56e-05	0.000216	0.000563
	(0.00327)	(0.00327)	(0.00327)	(0.00328)
λ	0.167***	0.169	0.0540***	0.125***
	(0.0182)	(0.206)	(0.00606)	(0.0465)
σ^2	4.460***		4.468***	
	(0.0974)		(0.0976)	
Log-likelihood	-9156.496		-9159.382	
Observations	4,219	4,219	4,219	4,219
Network fixed effects	YES	YES	YES	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Columns (1) and (2) report the average model of peer effects estimated by ML and GS2SLS while columns (3) and (4) report the aggregate model estimated by ML and G2SLS aggregate model. All models include network fixed effects. Standard errors are clustered at the network level.

Table 5: Control function approach: Link formation model.

	Link
Constant	0.00048***
	(0.00002)
Language test scores	−0.00002***
	(0.00000)
Cognitive ability test scores	−0.00001***
	(0.00000)
Male	0.00041***
	(0.00001)
Native	0.00056***
	(0.00001)
Age	0.00002
	(0.00002)
5 min distance	0.52904***
	(0.00036)
R^2	0.10928
Adj. R^2	0.10928
Num. obs.	17799961

Standard errors in parentheses

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: Results from OLS regression. The dependent variable is a dummy indicating whether there is a friendship link between two individuals. The explanatory variables include the absolute difference in scores on the language test, the difference in scores on the cognitive ability test, male dummy (1=both individuals are male), native dummy (1=both individuals are native) and the absolute difference in age and finally whether or not individuals live within a five min distance from each other.

Table 6: Outcome equation (OE) and link formation (LF).

	OE (1)	OE and LF (2)
Dependent variable: Disruptiveness		
Language test scores	-0.0186**	-0.0187**
	(0.00890)	(0.00890)
Cognitive ability test scores	-0.0370***	-0.0370***
	(0.00865)	(0.00865)
Age	0.208	0.208
	(0.133)	(0.133)
Male	0.0474	0.0463
	(0.150)	(0.150)
Native background	0.220**	0.220**
	(0.0986)	(0.0986)
Highest index of occupational status	-0.00164	-0.00165
	(0.00186)	(0.00186)
Missing HISEI	0.417***	0.417***
	(0.147)	(0.147)
Friends' average language test scores	-0.0595***	-0.0594***
	(0.0156)	(0.0156)
Friends' average cognitive test scores	-0.0384**	-0.0385**
	(0.0151)	(0.0151)
Friends' average age	0.0746**	0.0751**
	(0.0357)	(0.0357)
Proportion male friends	0.0696	0.0701
	(0.172)	(0.172)
Proportion native friends	0.119	0.120
	(0.158)	(0.158)
Friends' average HISEI	1.46e-05	1.78e-05
	(0.00327)	(0.00327)
Selectivity bias		8.05e-06
		(1.46e-05)
λ	0.167***	0.167***
	(0.0182)	(0.0182)
σ^2	4.460***	4.460***
	(0.0974)	(0.0974)
Observations	4,219	4,219
Network fixed effects	YES	YES

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Column (1) reports the results from ML estimations of the average peer effect model with network fixed effects, the so-called the outcome equation (OE). In column (2) the model also includes the estimated errors from the link formation model. The *selectivity bias* is reported in column (2). Standard errors are clustered at the network level.

Table 7: Observable characteristics of the key player vs. the most active or a random player.

	(1) Key player	(2) Most active player	(3) Random player
Language test scores	1.302***	0.988	1.020
	(0.027)	(0.014)	(0.015)
Cognitive ability test scores	1.135***	0.977*	0.980
	(0.022)	(0.013)	(0.013)
Male	1.144	1.271**	0.875
	(0.141)	(0.148)	(0.102)
Highest index of occupational status	0.998	0.997	1.000
	(0.003)	(0.003)	(0.003)
Native background	0.879	1.161	0.969
	(0.151)	(0.158)	(0.132)
Age	1.099	1.156	1.099
	(0.343)	(0.244)	(0.243)
HISEI missing	0.700	1.019	1.251
	(0.190)	(0.240)	(0.274)
Observations	4129	4129	4129
Pseudo R^2	0.184	0.006	0.002

Exponentiated coefficients; Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Results from logistic regressions.

Table 8: Predicted reductions from removing the key player, the most active player or a random player without any baseline.

	(1) Key player	(2) Most active player	(3) Random player
Network size (demeaned)	-1.218***	-1.142***	-1.148***
	(0.0416)	(0.0404)	(0.0418)
Constant	13.17***	11.88***	11.95***
	(0.272)	(0.264)	(0.273)
Observations	329	329	329
R-squared	0.724	0.710	0.698

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: Results from OLS regressions of the percentage reduction in disruptiveness from removing either the key player, most active or random player, calculated as the intercentrality of that player times 100 divided by the total Bonacich of that network, regressed on a constant and the independent variable network size.

Table 9: Predicted reductions from removing the key player (KP) when he or she is not the most active player (MA) or a random player (RP) in the network.

	(1) Difference KP and RP	(2) Difference KP and MA
Network size (demeaned)	-0.0973***	-0.0973***
	(0.0120)	(0.0138)
Constant	1.412***	1.440***
	(0.0780)	(0.0881)
Observations	295	301
R-squared	0.183	0.144

Notes: Results from OLS regressions. The dependent variable in column (1) is the difference in the average reduction in aggregate disruptiveness from removing the key player compared to removing a random player while in column (2) it is the difference in the average reduction in aggregate disruptiveness from removing the key player compared to removing the most active player.

Appendices

A Data creation notes

CILS4EU is a multileveled survey containing rich information on the family, teacher, school and classroom. It includes five sub-questionnaires directed at students, parents and teachers, entitled "Parents", "Teachers", "Youth classmates", "Youth friends" and "Youth main". The last three are directed towards students. The network data in this paper is created using "Youth main" the "Youth classmates" questionnaires. The number of respondents in the main questionnaire in the school year 2010-2011 was 5,025.

The analysis is based on the full data set including 249 classrooms, although sample restrictions could be considered in order to increase the proportion of participants per classroom (see important discussion in Hjalmarsson & Mood (2015) on CILS4EU classroom data). Figure A.1 shows the number of classroom if the sample is conditioned with respect to the degree of participation.

Figure A.1: Share of participants and sample restrictions. *Source:* Kruse & Konstanze (2016), Children of Immigrants Longitudinal Survey in Four European Countries. Sociometric Fieldwork Report. Wave 1 2010/2011, v1.2.0.

	ENG	GER	NET	SWE	TOTAL
	N(classes)	N(classes)	N(classes)	N(classes)	N(classes)
>60% partic.	202	243	220	250	915
>75% partic.	191	201	211	235	838
>90% partic.	153	97	158	172	580

The full sample in the "Youth classmates" file consists of 4,794 individuals (249 classrooms and 129 schools). As a first step, I drop all individuals who have not nominated anyone in the "Youth classmates" questionnaire (311 individuals). Based on the reduced sample I then create an edgelist file including all pairs of friendships.

Table A1: Classroom characteristics, full sample

Variable	Mean	Std. Dev.	Min.	Max.	N
Classroom size	20.353	4.287	6	31	4794

Next, I prepare the vertex file with all individual background variables including classid, schoolid, male age, disruptiveness, native, and HISEI. In the following step, I match the vertex file with a datafile with records of the students' language and cognitive ability test scores (4,804 observations). Individuals that performed the language and cognitive ability tests but did not take part in the main questionnaire were excluded (221 individuals in total). Individuals with missing values on HISEI (272 cases) have been given the sample average. In all regressions that include the HISEI variable I add a dummy for missing values on HISEI. I match the vertex file with the achievement file which leaves me with a total of 4,792 distinct cases. Next, I merge the vertex file with the edgelist. Since there are more distinct observations of "friends" (5,149 cases) than of "egos" (4,468 cases) I need to remove cases where egos are not found among the friends. Thus, I remove the observations from the edgelist file that contain an island among all the edges. In this step, 806 individuals are excluded due to matching issues. The analysis sample consists of about 72 percent of the total number of sampled students by CILS4EU.

Table A2: Classroom characteristics, analysis sample

Variable	Mean	Std. Dev.	Min.	Max.	N
Classroom size	18.298	4.445	3	28	4219

The matrix analyses are done in Stata, Mata (sppack) and R. I use Stata to construct the vertex file and the edgelist file which are then exported to R (gplot). In R, I create the network data for the key player simulation. Due to implementation and data memory issues the second stage estimations in the control function approach are done in R. Robustness checks are performed in Stata and Mata (sppack).

B Robustness checks

B.1 Instruments and exclusion restriction

I perform a number of robustness checks in order to asses the validity of the instruments and the exclusion restriction in the control function approach. Table B3 and B4 reports the correlation between individuals' characteristics and the average characteristics of their friends in the classroom conditional and unconditional on their 5 minute distance neighborhood cluster. Several estimates are noticeably reduced once I condition on the 5 minute distance variable.

B.2 Individual non-response

In order to estimate the network model, all isolated individuals (students with no friendship nominations) must be dropped as by construction the adjacency matrix cannot include missing values. As described in the data creation section A above, I drop all individuals who have not nominated anyone in the "Youth classmates" questionnaire (311 individuals). None of these "isolated" individuals filled in the Main questionnaire hence I am unable to explore their observable characteristics.

To get an indication of the degree of non-random selection due to individual non-response I investigate the characteristics of those excluded from the network analysis, in total 806 individuals. I perform this test on the individuals that are not matched with the edgelist file (those who did not take the language and cognitive ability tests are not included since they have already been dropped). It is not unlikely that these 573 individuals stand out in some way (non-random selection). Being absent at the time of the survey could be an indication of school shirking which is likely correlated with individual disruptiveness. I explore their observable characteristics in the descriptives table B1 below.

The cases dropped from the analysis sample are more likely male and have a foreign-born background. They also have higher scores on the disruptive measure while lower ones on the language and cognitive ability tests implying that the analysis sample is positively selected on these characteristics. With regard to the test scores, the means are significantly different from each other. Since the dropped individuals also have, on average, statistically higher self-reported disruptiveness levels, the estimated effect could be downward biased. The mean number of observations per classroom in the analysis sample is roughly 18 (see table A2).

Table B1: Observable characteristics, dropped individuals and the analysis sample

Variable	Mean	Std. Dev.	Min.	Max.	N
PANEL A: Dropped individuals					
Language test scores	15.515	5.829	0	30	573
Cognitive ability test scores	15.439	5.726	0	27	573
Age	15.122	0.387	14	17	573
Male	0.571	0.495	0	1	573
Disruptiveness	7.201	3.104	4	20	573
Native background	0.546	0.498	0	1	573
PANEL B: Analysis sample					
Language test scores	18.654	4.949	0	29	4219
Cognitive ability test scores	17.812	4.751	0	27	4219
Age	15.029	0.264	13	17	4219
Male	0.486	0.5	0	1	4219
Disruptiveness	6.364	2.433	4	20	4219
Native background	0.677	0.468	0	1	4219

Table B2: GJGX versus JGGX estimated using the Best IV approach.

	GJGX (1)	**JGGX** (2)
Intercept	0.00	0.00
	(0.04)	(0.03)
Language test scores	-0.03^{***}	-0.03^{***}
	(0.01)	(0.01)
Cognitive ability test scores	-0.04^{***}	-0.04^{***}
	(0.01)	(0.01)
Male	0.24^{***}	0.24^{***}
	(0.08)	(0.08)
Highest index of occupational status	-0.00	-0.00
	(0.00)	(0.00)
Native	0.20^{*}	0.20^{*}
	(0.10)	(0.10)
Age	0.11	0.10
	(0.16)	(0.17)
HISEI missing	0.45^{***}	0.45^{***}
	(0.16)	(0.16)
Average language friends	0.02	0.02
	(0.03)	(0.03)
Average cognitive friends	-0.00	-0.01
	(0.04)	(0.04)
Proportion male friends	-0.19	-0.18
	(0.17)	(0.18)
Average HISEI friends	-0.01	-0.01
	(0.00)	(0.00)
Proportion native friends	0.00	0.02
	(0.24)	(0.25)
Average age friends	-0.19	-0.20
	(0.33)	(0.33)
Average HISEI missing	-0.10	-0.09
	(0.37)	(0.37)
Local average peer effect	0.34	0.28
	(0.55)	(0.61)
R^2	-0.05	-0.03
Adj. R^2	-0.06	-0.04
Num. obs.	4219	4219
Wald test	5.962	6.057

Standard errors in parenthesis.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: Both column (1) and (2) include network fixed effects.

Table B3: Correlation between an individual's characteristics and the average characteristics of one's self-reported friends in the classroom *unconditional* on their their 5 min distance neighborhood cluster.

	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	9.11***	10.59***	0.25***	0.19***	13.73***
	(0.49)	(0.50)	(0.02)	(0.01)	(0.50)
Language test scores	0.51***				
	(0.03)				
Cognitive ability test scores		0.40***			
		(0.03)			
Male			0.49***		
			(0.03)		
Native				0.72***	
				(0.02)	
Age					0.09**
					(0.03)
R^2	0.11	0.06	0.10	0.00	0.29
Adj. R^2	0.11	0.06	0.10	0.00	0.29
Num. obs.	3253	3253	3253	3253	3253

Standard errors in parentheses
$^{***}p < 0.001$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: Results from OLS regressions of an individual's observable characteristics and the average characteristic of his or her friends unconditional on the distance variable. For example Model 3 shows the correlation between an individual's gender and the average gender of one's self-reported friends in the classroom.

Table B4: Correlation between individuals' characteristics and average characteristics of their friends in the classroom *conditional* on their their 5 min distance neighborhood cluster.

	Model 1	Model 2	Model 3	Model 4	Model 5
Language test scores	0.47^{***}				
	(0.03)				
Cognitive ability test scores		0.39^{***}			
		(0.03)			
Male			0.49^{***}		
			(0.03)		
Native				0.70^{***}	
				(0.02)	
Age					0.07^{**}
					(0.03)
Num. obs.	3253	3253	3253	3253	3253
Adj. R^2	0.14	0.08	0.10	0.30	0.02

Standard errors in parentheses
$^{***}p < 0.001$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: Results from OLS regressions of an individual's observable characteristics and the average characteristic of his or her friends conditional on the distance variable. For example Model 3 shows the correlation between an individual's gender and the average gender of one's self-reported friends in the classroom.

C Questionnaire items, CILS4EU

The Classmates questionnaire wave 1:

- (Q1) Who are your best friends in this class? (Here you may write down no more than five numbers.)
- (Q9) Which of your classmates live within a 5 min walk from your home?
- (Q10) Who do your parents know?

Selected questions from the Main questionnaire wave 1:

- (Q20) How often do you... (Every day, Once or several times a week, Once or several times a month, Less often, Never)
  - ... argue with a teacher?
  - ... get a punishment in school (for example being kept in detention, being sent out of class, writing lines)?
  - ... skip a lesson?
  - ... come late to school?
- (Q81) Have you done the following things in past 3 months? Your answers will be kept secret. (Yes, No)
  - Deliberately damaged things that were not yours?
  - Stolen something from a shop/from someone else?
  - Carried a knife or weapon?
  - Been very drunk?
- (Q93) How often do you... (Every day, Once or several times a week, Once or several times a month, Less often, Never)
  - ... drink alcohol?
  - ... smoke cigarettes?
  - ... use drugs (for example, hash, paddos, ecstasy pills)?

Source: Kalter et al. (2013).