

Neighborhood Effects in Education

Carlo L. Del Bello* Eleonora Patacchini† Yves Zenou‡

March 30, 2015

Abstract

Using unique geo-coded information on the residential address of a representative sample of American adolescents and their friends, we revisit the importance of geographical proximity in shaping education outcomes. Our findings reveal no evidence of residential neighborhood effects. Social proximity, as measured by similarity in religion, race and family income as well as in unobserved characteristics, appears to play a major role in facilitating peer influence. Our empirical strategy is able to control for the endogeneity of both social network and location choices.

Key words: Neighborhood effects, social networks, link formation, education.

JEL Classification: C21, Z13.

*Paris School of Economics, France. Email: cl.delbello@psemail.eu.

†Cornell University, USA and CEPR. Email: ep454@cornell.edu.

‡Stockholm University and IFN, Sweden and CEPR. E-mail: yves.zenou@ne.su.se.

1 Introduction

The impact of neighborhood on education outcomes has been the focus of researchers for a long time. The general consensus is that the neighborhood where individuals grow up matters, although the effects are small after controlling for individual and family characteristics and parental selection of residential neighborhood (see e.g., Durlauf, 2004; Dujardin et al., 2009; Ioannides 2012). For example, Solon et al. (2000), Oreopolous (2003), and the papers using the Moving to Opportunity (MTO) programs (Katz et al., 2001; Rosenbaum and Harris, 2001; Kling et al., 2005, 2007; Sanbonmatsu et al., 2006) find little evidence of neighborhood effects on education. On the contrary, there are strong evidence suggesting that peer effects matter in education outcomes and they mainly operate at schools (Sacerdote, 2011, 2014). These two results suggest that school-based policies could be more effective in enhancing education than place-based policies (Fryer and Katz, 2013; Katz, 2015).

This paper is among the first to examine the relative importance of neighborhood versus peer effects in education outcomes by using the actual social interaction patters of the individuals who live in the same census block group. Our main results show that, when using this more precise definition of neighbors and when accounting for neighbor choices, there is no evidence of neighborhood effects on education. On the contrary, we show that peer effects are important in education and mainly operate at schools.

Our analysis is made possible by the unique information contained in the National Longitudinal Survey of Adolescent Health (AddHealth). While the AddHealth data has been extensively used for its social network information based on friends' nomination, this dataset also contains another unique information that has not been much exploited. Indeed, the AddHealth also provides the geographical coordinates of the residential location of each respondent. This dataset allows us to obtain information not only on the residential Census block, but also on the precise location of individuals within (and across) Census blocks.

Using the Census-block as a measure of residential neighborhood, we begin our analysis by differentiating between *peers at school*, who are the friends nominated by a student but who do not live in the same neighborhood (or Census block) and *peers in the neighborhood*, who are the friends nominated by a student who also reside in the same neighborhood. We estimate the effects of these different peers on own education using a Bayesian methodology that takes into account both the observable and unobservable characteristic differences between the students by explicitly modelling the friendship formation process of these students. When we implement such a strategy, we find that the effect of peers-in-the-neighborhood on own GPA becomes insignificant. This indicates that the correlation in behavior between peers who

reside in the same neighborhood is mainly due to the unobservable common characteristics that drive the link formation process and thus there is no “real” impact of peers-in-the-neighborhood on own GPA. This implies that the standard estimates of peer effects in education can be misleading given that peers are not differentiated by their place of residence and link formation is not included in the model.

We then investigate more closely the relationship between physical distance and peer effects by exploiting the more detailed information on each student’s residential location. For each individual, we construct concentric rings with an increasing radius and study the spatial decay of the peer effects. When all the individuals within each ring are considered as peers (i.e. *geographical peers*), we find the traditional results known in the literature on neighborhood effects: peer effects are generally decreasing with physical distance. Furthermore, when neighborhood fixed effects are included in the model, only highly localized social interactions seem to be conducive of peer effects. When we then consider as peers the *nominated friends*, the decreasing relationship disappears: distant friends seem to be as important as friends who reside close by.

Such evidence is consistent with a social network formation model where there is a trade off between the costs and benefits of friendship. We propose a simple theoretical framework where each agent chooses both education effort and friendship links. Our framework helps understand the mechanism behind our results, i.e. why neighborhood effects or more precisely the importance of geographical distance in shaping peer effects disappears once neighbors are defined as the actual (chosen) friends. Indeed, in our model, an individual would choose a more distant friend only if the expected value of this friendship is higher than the costs of interacting with distant friends. As a result, distant friends may be as important as close friends.

The rest of the paper unfolds as follows. In the next section, we discuss the related literature, while highlighting our contribution. In Section 3, we describe our data while, in Section 4, we discuss the different identification issues and explain how we address them. In Section 5, we present our empirical results. Section 6 proposed a theoretical framework which is consistent with our evidence. In Section 7, we investigate the relative importance of the geographical and social distance between students in facilitating peer effects. Finally, Section 8 concludes and discusses some policy implications of our results.

2 Related Literature

Our paper contributes to different literatures. They are briefly reviewed below, while highlighting our contribution.

2.1 Neighborhood versus network effects

As mentioned in the Introduction, there is a large empirical literature of neighborhood effects on education, with mixed results.¹ A more recent literature has looked at the effect of social networks and peers on education. The empirical evidence here seems to point towards an important role of social contacts.²

There is, however, little research that tries to understand the role of both the neighborhood (geographical space) *and* the social network (social space) on education outcomes.³ Recent empirical evidence has shown that the link between these two spaces is quite strong, especially within community groups (see e.g. Bayer et al., 2008; Hellerstein et al., 2011, 2014; Ioannides, 2012; Ioannides and Topa, 2010; Patacchini and Zenou, 2012; Topa, 2001; Topa and Zenou, 2015). For example, using Census data on residential and employment locations, Bayer et al. (2008) show that individuals residing in the same city block are more likely to work together than those in nearby blocks. Similarly, using matched employer-employee data at the establishment level, Hellerstein et al. (2011) show the presence and importance of labor market networks based on residential proximity, in determining the establishments at which people work.⁴

Our paper contribute to this literature by explicitly examining the role of network and

¹See Dujardin et al. (2009), Ioannides (2011), Ioannides and Topa (2010), Overman et al. (2015) and Topa and Zenou (2015) for recent surveys on neighborhood effects.

²Apart from the classical peer-effect papers of Sacerdote (2001) and Zimmerman (2003), there are recent papers that also found peer effects in education. For example, Lavy et al. (2012) find that high-ability high school students in Israel benefit from the presence of other high ability students. Imberman et al. (2012) use the unexpected arrival of Hurricane Katrina evacuees as a shock to peer groups and find that high achieving students benefit the most from the arrival of high achieving peers and are hurt the most by the arrival of low achieving peers. Surveys on the importance of social networks in education can be found in Ioannides and Loury (2004) and Sacerdote (2011, 2014).

³There are some recent theoretical papers that look at the effect of these two spaces on labor-market outcomes. See Zenou, (2013), Picard and Zenou (2014), Helsley and Zenou (2014), Sato and Zenou (2015).

⁴Beaman (2012) and Edin et al. (2003) exploit natural experiments, consisting of refugee resettlement programs in the U.S. and Sweden, respectively, to try to disentangle social network referral effects from sorting or correlation in unobservable attributes. Beaman (2011) finds that a one standard deviation increase in the number of network members in a given year lowers the employment probability of someone arriving one year later by 4.9 percentage points. She also finds that a greater number of tenured network members improves the probability of employment and raises the hourly wage. Edin et al. (2003) find similar positive results. See also the recent papers by Damm (2014) and Damm and Dustmann (2014).

neighborhood effects for education outcomes. More specifically, this is the first paper that investigates the importance of social and geographical proximity in facilitating peer influence in education.

2.2 The identification of neighborhood effects

The identification of neighborhood effects is difficult because of nonrandom sorting of families into neighborhoods and the likely presence of unobserved individual and neighborhood characteristics. The existing studies have followed a variety of approaches. Some studies base their estimates of neighborhood effects on the random component of neighborhood choice induced by special social experiments. Most notably, Katz et. al. (2001), Kling et al. (2001, 2007), Ludwig et al. (2001, 2013) and Sanbonmatsu et al. (2006) have used the randomized housing voucher allocation associated with the Moving to Opportunity (MTO) programs to examine neighborhood effects on a very wide set of individual outcomes, including health, labor market activity, crime, education, and more.⁵ As stated in the Introduction, these studies find very little effect of neighborhood on education outcomes. For example, Sanbonmatsu et al. (2006), who use data on more than 5,000 children aged six to 20 at all five MTO sites (Boston, Baltimore, Chicago, Los Angeles, and New York), looking at medium-term outcomes four to seven years after random assignment, show that there are no significant effects on test scores for any age group of these children. Another set of papers deal with neighborhood-level sorting by aggregating to a higher level of geography - for example by using metropolitan area level data rather than neighborhood level data.⁶ A different approach is instead followed by studies that adopt a structural approach. Those papers use models generating a rich stochastic structure that can be taken to data for estimation. The parameters of these models can then be estimated by matching moments from the simulated spatial distribution generated by the model with their empirical counterparts from spatial data on neighborhoods or cities. For example, Glaeser et al. (1996) explain the very high variance of crime rates across U.S. cities through a model in which agents' propensity to engage in crime is influenced by neighbors' choices.⁷ Finally, some more recent papers have

⁵Studies that examine neighborhood effects using other social experiments include Jacob (2004) and Oreopoulos (2003).

⁶See for example Cutler and Glaeser (1997), Evans et al. (1992), Ross (1998), Weinberg (2000), and Ross and Zenou (2004)

⁷Their model is a version of the voter model, in which agents' choices regarding criminal activity are positively affected by their social contacts' choices. One important innovation in this paper is to allow for "fixed agents", who are not affected by their neighbors' actions. The variance of crime outcomes across replications of the economy (i.e., cities) is inversely proportional to the fraction of fixed agents in an economy. The distance between pairs of fixed agents in the model yields a measure of the degree of interactions. By

proposed innovative identification strategies to identify neighborhood effects by exploiting the extremely detailed information on geography. The above mentioned papers by Bayer et al. (2008) and Hellerstein et al. (2011) belong to this literature.⁸

Our paper contributes to this latter strand of the literature and makes progress in the following directions. Our data provide information on the precise location of individuals within (and across) Census blocks. For each individual, we can thus define as neighbors all the individuals who live within a given radius and consider the influence of individuals at different distances. Sorting effects (or the presence of unobserved individual and local area correlated characteristics) are controlled for by including *Census Block fixed effects*. Our main assumption (as well as in Bayer et al., 2008) is that, while individuals are able to choose their residential neighborhood (block group), the exact home that they choose within this block group is as good as randomly assigned with respect to the characteristics of their neighbors. This assumption is motivated by two considerations. First, the thinness of the housing market at such small geographic scales restricts an individual's ability to choose a specific block. Secondly, it may be difficult for individuals to identify block-by-block variation in neighbor characteristics at the time of purchase or lease. That is, while an individual may have a reasonable sense of the socio-demographic structure of the neighborhood, the variation across blocks within a neighborhood is less easily observed a priori. In addition, because we have network data within neighborhoods, we can separately identify endogenous from exogenous effects, i.e. the effects of peers that stem from their behavior from the effects that stem from peer characteristics. In so doing, we would overcome the so-called reflection problem (Manski, 1993). As a result, our analysis is able to identify the effects of neighbors' school performance as distinct from the effects of neighbors' characteristics. Furthermore, the geographical connotation of the data allows us also to estimate the spatial reach of the effects. Observe that, while Bayer et al. (2008) document convincing evidence of the presence of peer effects (local referral effect in their context), their data structure does not allow to separately identify endogenous from exogenous effects, nor estimate the spatial reach of the effects.

matching the empirical cross-city variance of various types of crime with that implied by the model, the authors estimate the extent of neighborhood effects for these different types of crime. Other studies following a structural approach are Topa (2001) and Conley and Topa (2002).

⁸Other papers exploiting detailed information on geography to identify neighborhood effects include Hellerstein et al. (2014), Patacchini and Zenou (2012) and Schmutte (2015).

2.3 The identification of social network effects

The most challenging issue faced by all studies using social network data to identify peer effects is that individuals sort into groups in a non-random way.⁹ If the variables that drive this process of selection are not fully observable by the researcher, then potential correlations between (unobserved) group-specific factors and the target regressors are major sources of bias. To address this issue, most of the existing papers (see, in particular, Bramoullé et al., 2009; Calvó-Amengol et al., 2009; Lin, 2010; Lee et al., 2010) use the architecture of the networks by introducing *network fixed effects* in the econometric equation.¹⁰ The underlying assumption is that the unobservable factors that drive friendship formation are common to all individuals belonging to the same network. This means that it is assumed that the structure of interactions is (conditionally) exogenous. However, if there are individual-level unobservables that drive both network formation and outcome choices, this strategy will not work.¹¹ By failing to account for similarities in unobserved characteristics, similar behaviors might mistakenly be attributed to peer influence when they are simply due to similar unobserved characteristics.

In this paper, we model explicitly network formation and estimate a model of link formation and outcomes using a Bayesian approach.¹² By doing so, a possible presence of unobservable individual characteristics affecting both network formation and outcome decisions is taken into account. The importance of this methodological innovation is confirmed by the fact that the results are dramatically different when including or not the network formation.

3 Data description

We use a unique database on friendship networks from the National Longitudinal Survey of Adolescent Health (AddHealth).¹³ The AddHealth survey has been designed to study

⁹This is similar to the nonrandom sorting into geographical locations addressed above. However, the sorting is now into social networks and thus deal with friendship and network formation.

¹⁰This is a similar approach to the neighborhood fixed-effect approach mentioned above.

¹¹For a general discussion and overview on these issues, see Blume et al. (2011), Goldsmith-Pinkham and Imbens (2013), Graham (2015), and Jackson et al. (2015).

¹²A similar modeling approach is used by Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2013)

¹³This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in

the impact of the social environment (i.e. friends, family, neighborhood and school) on adolescents' behavior in the United States by collecting data on students in grades 7-12 from a nationally representative sample of roughly 130 private and public schools in years 1994-95. A random subset of those adolescents, about 12,000 individuals, is also asked to compile a longer questionnaire containing sensitive individual and household information (*core sample survey data*), and includes the geographical coordinates of the residential address.

From a network perspective, the most interesting aspect of the AddHealth data is the friendship information, which is based upon actual friends nominations. Indeed, pupils were asked to identify their best friends from a school roster (up to five males and five females).¹⁴ As a result, one can reconstruct the whole geometric structure of the friendship networks, summarized in the adjacency matrix $\mathbf{G} = \{g_{ij}\}$, where $g_{ij} = 1$ if i and j are friends, and $g_{ij} = 0$, otherwise. Such a detailed information on social interaction patterns allows us to measure the peer group precisely by knowing exactly who nominates whom in a network (i.e. who interacts with whom in a social group).

For each individual, we calculate an index of school performance using the respondent's scores received in the more recent grading period in several subjects, namely mathematics, English, history and science. The scores are coded as 1=D or lower, 2=C, 3=B, 4=A. The final average score (*GPA index* or grade point average index) has mean equal to 2.93 and standard deviation equal to 0.75.

To analyze the relationship between *neighborhood* and *social interactions*, we use the clustering of respondents into statistical spatial units made possible by the AddHealth data. The spatial file of the AddHealth dataset provides information on the residential area where each student of the core sample resides at an *extremely high level of spatial disaggregation*. For each student, we use the residential Census block-group identifier and merge this information with the nomination data.¹⁵ This information allows us to decompose the peers (friends) of student i between his/her *peers at school*, who are the friends nominated by i at school but who do *not* live in the same neighborhood as i , and the *peers in the neighborhood*, who are the peers nominated by i at school who also live in the same neighborhood as i .

To illustrate our decomposition, consider the following simple example (Figure 1) of an economy composed of 3 individuals and 2 neighborhoods, \mathcal{N}_1 and \mathcal{N}_2 . In the *social space*,

the original design. Information on how to obtain the Add Health data files is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>). No direct support was received from grant P01-HD31921 for this analysis.

¹⁴The limit in the number of nominations is not binding (even by gender). Less than 1% of the students in our sample show a list of ten best friends.

¹⁵For confidentiality concerns, respondents can only be grouped into neighborhoods by using pseudo-IDs, which correspond to Census block groups, but they cannot be directly matched to Census data.

individuals 1 and 2 as well as individuals 1 and 3 are friends (i.e. peers). In the *geographical space*, individuals 1 and 2 live in the same neighborhood while individual 3 resides in another neighborhood. As a result, individuals 1 and 2 are considered as *peers in the neighborhood* while individuals 1 and 3 are considered as *peers at school*.

[Insert Figure 1 here]

Let us go back to the description of our dataset. Our definition of *neighborhood* is the Census block-group. A Census block-group is a collection of approximately ten contiguous city blocks, with population ranging from 600 to 3,000 inhabitants. Figure 2 shows the distributions of census blocks by their size in our sample. It appears that almost 40% of the neighborhoods have less than 20 (sampled) respondents. Networks of contacts typically include individuals belonging to different neighborhoods.¹⁶

[Insert Figure 2 here]

Figure 3 presents an heuristic description of our data structure. Green boxes represents networks, while blue boxes neighborhoods. The larger is the box, the wider is the size of the network. Red dots represent friendships between student i (y -axis) and student j (x -axis). For example, the first network consists of 45 individuals who nominate on average 3 friends. In total, we have 67 friendship ties. Among these ties, 30% are between students residing in the same neighborhood (red dots in blue boxes) and 60% are between students who reside in different neighborhoods (red dots outside blue boxes). In this network, friendship links in the same neighborhoods are distributed over 7 different neighborhoods. Network 2 is denser and consists of 36 individuals who have nominated on average 3.5 friends; a total of 125 friendship links. The ties between students residing in the same neighborhood (56% of the total) are between students residing in only two neighborhoods, a small and a large one. The last network, which has approximately the same size (38 individuals), is more sparse and the ties within the neighborhoods are distributed over two neighborhoods of approximately the same size.

[Insert Figure 3 here]

By excluding the students with missing or inadequate information, we obtain a final sample of 3,126 students distributed over 568 networks. This large reduction in sample size

¹⁶Less than 5% of our networks are within a single neighborhood.

with respect to the original core sample is mainly due to the network construction procedure - roughly 20 percent of the students do not nominate any friend and another 20 percent cannot be correctly linked.¹⁷ In addition, we need to minimize the missing data in the average level of activity of peers in each neighborhood by ensuring that, for the majority of our students, there is at least one nominated friend residing in her/his residential neighborhood. This requirement led to the exclusion of neighborhoods with less than 4 people and of networks with less than 10 members.¹⁸ Our final sample consists of 838 individuals distributed over 29 networks. Table 1 provides information on the 29 networks of our study and gives precise definitions and summary statistics of the variables used in our analysis. The mean and the standard deviation of network size are roughly 29 and 26 students, respectively. Students, on average, nominate slightly more than two friends, one of whom is also residing in the same neighborhood. Among the individuals selected in our sample, 54% are female and 19% are blacks. The average parental education is more than high-school graduate. Roughly 25% of our students come from a household with only one parent. On average, our adolescents live in a household of about 3.3 people. Less than 45% of the students in our sample attend religious services more than once a month. Finally, the average family income is approximately 40,000 in 1995 dollars.

[Insert Table 1 here]

Figure 4 displays the distribution of friends by geographical distance. The distance ranges between 0 and 50 kilometers. The distribution is skewed to the left, having 50% of the students living within about 8 kilometers from their friends (median value equal to 8.26).

Figure 5 shows the distribution of neighborhoods by average distance between residents. The average distance is about 10.5 kilometers with a notable dispersion around this mean value.

[Insert Figures 4 and 5 here]

Table 2, panels a and b, compare the characteristics of friends along the geographical dimension. We use the median value as a threshold to discriminate between close and distant friends. Specifically, we define as distant friends those who reside at more than 8.26 km from each other and as close friends those who reside at less than 8.26 km from each other. The table does not reveal pronounced differences between distant and close friends in

¹⁷This is common when working with AddHealth data. The representativeness of the sample is, however, preserved.

¹⁸The maximum network size is 106 individuals.

terms of observable characteristics, apart from GPA and grade. Indeed, we see that *distant friends tend to be older and have lower GPA than close friends*. For example, if we consider panel a, we see that 38% of distant friends have the same GPA (against 34% for close friends) while 70% are in the same grade (against 74% for close friends). If we now consider panel b, then we see that 31% of distant friends have a higher GPA (against 35% for distant friends) and 17% are in a higher grade (against 14% for distant friends).

[Insert Table 2 here]

4 Econometric framework and identification issues

Let \bar{r} be the total number of networks in the sample (i.e. $\bar{r} = 29$), n_r the number of individuals in the r th network, and $n = \sum_{r=1}^{\bar{r}} n_r$ the total number of individuals (i.e. $n = 838$). Let $\mathbf{x}_{il,r} = (x_{il,r}^m, \dots, x_{il,r}^M)'$ and denote the vector of individual observable characteristics of individual i residing in neighborhood l belonging to network r . Assume that each network r contains \bar{l} mutually exclusive neighborhoods. For each network r , we know where each student lives and thus we can group together all students from the same neighborhood l belonging to the same network r . Let us denote the adjacency matrix of the peers at school by $\mathbf{G}^{PS} = \{g_{ij}^{PS}\}$, where $g_{ij}^{PS} = 1$ if i and j are friends but do not live in the same neighborhood and 0 otherwise. Similarly, let the adjacency matrix of the peers in the neighborhood be $\mathbf{G}^{PN} = \{g_{ij}^{PN}\}$, where $g_{ij}^{PN} = 1$ if i and j are friends and reside in the same neighborhood. Our empirical model of agent i belonging to network r and residing in neighborhood l can then be written as:

$$\begin{aligned}
 y_{il,r} = & \phi^{PS} \sum_{j=1}^{n_r} g_{ij,r}^{PS} y_{jl,r} + \phi^{PN} \sum_{j=1}^{n_r} g_{ij,r}^{PN} y_{jl,r} + \mathbf{x}'_{il,r} \beta \\
 & + \frac{1}{g_{i,r}^{PS}} \sum_{j=1}^{n_r} g_{ij,r}^{PS} x'_{jl,r} \gamma^{PS} + \frac{1}{g_{i,r}^{PN}} \sum_{j=1}^{n_r} g_{ij,r}^{PN} x'_{jl,r} \gamma^{PN} + \eta_{l,r} + \epsilon_{il,r}
 \end{aligned} \tag{1}$$

where $y_{il,r}$ is the GPA of student i residing in neighborhood l belonging to network r , $\eta_{l,r}$ represents the neighborhood fixed effect and $\epsilon_{il,r}$ is an error term. In this model, ϕ^{PS} and ϕ^{PN} represent *the endogenous effects*, which capture the impact of peers' outcomes on own outcome, while γ^L and γ^S represent *the contextual effect*, which represents the impact of the exogenous characteristics of peers on own outcome. The behavioral foundation of this model

is an extension of the network model developed in Calvó-Armengol et al. (2009) to the case of heterogenous peer effects. It is presented in Appendix A. It shows that both the position in the network and the interactions in the neighborhood play an important role in shaping educational outcomes.

Model (1) is a spatial autoregressive model with two adjacency matrices and heterogeneous autoregressive parameters. It can be estimated using a Bayesian approach (see LeSage and Pace, 2009, for a review on Bayesian SAR models).

Let us now describe the traditional challenges in the estimation of peer effects and explain how we tackle them.

Reflection problem It is well-known that when estimating neighborhood/peer effects using a linear-in-means model (i.e. when regressing individual activity level on the average activity level in the neighborhood/among the peers), the endogenous and contextual effects cannot be separately identified due to *the reflection problem*, first formulated by Manski (1993). With explicit social network data, as we have here, this problem is eluded (see, e.g. Bramoullé et al. 2009; Liu and Lee, 2010; Calvó-Armengol et al., 2009; Lin, 2010; Lee et al., 2010). Indeed, the reflection problem arises in linear-in-means models because individuals interact in groups - individuals are affected by all individuals belonging to their group and by nobody outside the group. In the case of social networks, instead, since the reference group is individual specific, this is not true because peer groups are overlapping. Formally, social effects are identified (i.e. no reflection problem) if at least two individuals in the same network have different links (Bramoullé et al. 2009). This condition is generally satisfied in any real-world network.

While a network approach allows us to separate endogenous effects from contextual effects, it does not necessarily produce the causal effect of peers' influence on individual behavior. In our context, we could in fact potentially face two sources of bias: correlated neighborhood effects and endogenous network formation.

Correlated neighborhood effects

The proper identification of neighborhood effects could be complicated by the nonrandom sorting of households into neighborhoods and the likely presence of unobserved individual and neighborhood attributes. Indeed, there may exist (i) some unobservable household or individual characteristics that influence both the propensity of having good grades and neighborhood choices and/or (ii) neighborhood specific *unobservable* factors affecting both individual and peer behavior. For example, two peers (friends) who reside in the same neighborhood (peers in the neighborhood) could be facing the same school quality, which

could influence their grades and this has nothing to do with peer effects.

Regarding (i), it should be noted that our dataset is about adolescents who have not chosen themselves the location where they live; their parents did. This is, however, not perfectly satisfying because parents' and children's unobservables may be correlated. The extremely high level of geographical detail of our data allow us to identify peer effects while including *neighborhood fixed effects* in our model. The strategy is similar to Bayer et al. (2008). The idea is that although households are able to select a given area in which they want to live, they are, however, unable to select a precise neighborhood within this given area. As a consequence, once we include neighborhood fixed effects, the remaining spatial variance of education activities within the neighborhood is supposed to be exogenous. In addition, this fixed-effect strategy washes out everything that acts at the neighborhood level and thus common unobservable factors that may affect both individual and peer behavior. As a result, this strategy addresses both (i) and (ii).

Endogenous network formation As mentioned in the Introduction, the most challenging issue faced by all studies attempting to identify peer effects is a possible endogeneity of the network. While it might be reasonable to assume that teenagers are not choosing where to live (their parents do), they certainly choose their friends, that is the nominations in the AddHealth data. If there are individual-level unobservables that drive both network formation and outcome choices, the estimates of peer effects are biased.

Let $\xi_{il,r}$ denote such an *unobserved characteristic* of individual i residing in neighborhood l belonging to network r , which influences the link formation process. Let us also assume that $\xi_{il,r}$ is correlated with $\epsilon_{il,r}$ in Model (1) according to a bivariate normal distribution

$$(\xi_{il,r}, \epsilon_{il,r}) \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_{\epsilon_{il,r}}^2 & \sigma_{\epsilon_{il,r}\xi_{il,r}} \\ \sigma_{\epsilon_{il,r}\xi_{il,r}} & \sigma_{\xi_{il,r}}^2 \end{bmatrix} \right). \quad (2)$$

Joint normality implies that the error term $\epsilon_{il,r}$ in equation (1) can be replaced with its expected value $\sigma_{\epsilon_{il,r}\xi_{il,r}}\xi_{il,r}$, yielding:

$$\begin{aligned} y_{il,r} &= \phi^{PS} \sum_{j=1}^{n_r} g_{ij,r}^{PS} y_{jl,r} + \phi^{PN} \sum_{j=1}^{n_r} g_{ij,r}^{PN} y_{jl,r} + \mathbf{x}'_{il,r} \beta \\ &+ \frac{1}{g_{i,r}^{PS}} \sum_{j=1}^{n_r} g_{ij,r}^{PS} x'_{jl,r} \gamma^{PS} + \frac{1}{g_{i,r}^{PN}} \sum_{j=1}^{n_r} g_{ij,r}^{PN} x'_{jl,r} \gamma^{PN} + \eta_{l,r} + \sigma_{\epsilon_{il,r}\xi_{il,r}} \xi_{il,r} + u_{il,r} \end{aligned} \quad (3)$$

where $u_{il,r}$ is now an i.i.d. error term uncorrelated with the $x_{il,r}$ s and the unobservable $\xi_{il,r}$. Observe that $\sigma_{\epsilon_{il,r}\xi_{il,r}} \neq 0$ implies that the networks in model (1), \mathbf{G}_r^{PS} and \mathbf{G}_r^{PN}

are endogenous.¹⁹ In this case, individual-level correlated unobservables would motivate the use of parametric assumptions and Bayesian inferential methods to integrate a network formation model with the study of behavior and outcomes over the endogenous networks (see Goldsmith-Pinkham and Imbens, 2013, Hsieh and Lee, 2015).

Let us consider a network formation model based on homophily behaviors, where the variables that explain friendship ties between students i and j belonging to network r , i.e. $g_{ij,r}^p$, with $p = PS, PN$, are the distances between them in terms of observed and unobserved characteristics. Let us assume that the probability of two individuals being friends $g_{ij,r}^p$ follows a logit specification of the form

$$P(g_{ij,r}^p = 1 | x_{il,r}, x_{jl,r}, \xi_{il,r}, \xi_{jl,r}) = \Lambda(\delta_0 + \sum_m |x_{il,r}^m - x_{jl,r}^m| \delta^m + \delta_2 |\xi_{il,r} - \xi_{jl,r}|) \quad (4)$$

where $\Lambda(\cdot)$ is the logistic distribution and $\delta_0, \delta_1, \delta_2$ are parameters governing friendship formation. Among the observable individual characteristics (x variables), we also include a dummy variable taking a value of 1 if students i and j reside in the same neighborhood and zero otherwise. Equation (4) explains the link formation process between individuals i and j in network r by their difference in observable characteristics (i.e. $|x_{il,r}^m - x_{jl,r}^m|$) and unobservable characteristics (i.e. $|\xi_{il,r} - \xi_{jl,r}|$). This is a standard model of *homophily* (see e.g. Currarini et al., 2009, 2010).

Equations (4) and (1) form a structural model of link formation and outcomes, which can be estimated using a Bayesian approach. A Bayesian approach will produce marginal posterior distributions of the parameters, conditioned on the data and the set of individual-level nuisance parameters $\xi_{il,r}$. We can interpret $\xi_{il,r}$ as an individual fixed-effect that affects the outcome and also explains the probability of two individuals i and j being friends. Details on the Bayesian estimation procedure can be found in Appendix B.

5 Empirical results

Table 3 reports the results of the estimation of equation (1) only, i.e. when assuming an exogenous choice of friends, with an increasing set of controls. The last row of this table shows the percentage of the variance that is explained by peer effects. Column (1) shows that when conditioning on individual characteristics, peer effects are important and the

¹⁹For simplicity, we consider only one unobserved characteristics governing the link formation process. The introduction of different unobservables affecting differently different types of ties does not change anything, it just adds more notations.

percentage of the variance explained by peer behavior is 22.7 % and 33.8% for peers at school and in the neighborhood, respectively. However, a correlation between individual and peers' behavior may be due to similar individual and peer characteristics, rather than to peer effects (i.e. endogenous effects). The uniqueness of our data where both respondents and friends are interviewed allows us to control for peers' characteristics, thus disentangling the endogenous effects from the exogenous ones. Observe that peers' characteristics also include the characteristics of the parents of the friends (in particular peers' family income, parental education, race). Compared to column (1), column (2) shows that, indeed, a portion of the effects attributed to peers' behavior is in fact due to peers' characteristics. The remaining concerns relate to the presence of unobserved factors since the observed characteristics of peers may not capture all the aspects of the social environment. The fact that, in our data, we observe individuals over neighborhoods (in addition to networks) allows us to include neighborhood fixed effects.²⁰ By doing so, we purge our estimates from the effects of unobserved factors that are common to all individuals living in the same neighborhood. Column (3) shows that the percentage of explained variance falls by approximately two third for both types of peers, thus revealing the presence of important unobserved factors in each individual's social circle. The presence of unobserved individual-level unobserved factors driving both social contacts formation and behavior is instead addressed in Table 4 (simultaneous network and outcome equations model).

Importantly, Table 3 shows that, in all specifications, both peers in the neighborhood and peers at school have a positive and significant impact on own GPA but the magnitude of the effect of the latter is always higher than that of the former.²¹ This is an interesting result, which shows that friends who are nominated at school but who do not live in the same neighborhood have a stronger impact on own education activities than those who reside in the same neighborhood. A possible explanation could be the fact that social norms about education are often established at school and kids tend to study with other kids at school rather than in the neighborhood.

Interestingly, when neighborhood fixed effects are included, then the coefficient of peers-in-the neighborhood becomes much smaller (the coefficient is divided by two from 0.08 to 0.0486) while the coefficient of peers-at-school is nearly not affected. We report in Appendix C, Table C.1, the results that are obtained when we separate the two peer-related variables in two different models. It appears that the results are qualitatively unchanged, even though

²⁰This is a pseudo panel data within-group strategy, where the group mean (here network mean) is removed from each individual observation.

²¹Observe that, since all the peer-related variables are standardized to have mean zero and variance 1, the coefficients are comparable in magnitude.

the effects are higher in magnitude. For completeness, Table C.2 reports the estimation results of a traditional peer effects model in education, i.e. when we do not differentiate between peers at school and peers in the neighborhood. If we compare Table 3 and Table C.2, we see that the magnitude of the coefficient of the endogenous peer effect in the latter is in between the corresponding coefficients of peers in the neighborhood and peers at school in the former.

[Insert Table 3 here]

Let us now turn to the main novelty of this paper. In Table 4, we report in column (1) the estimates of the outcome equation when the network is assumed exogenous, which simply corresponds to column (3) in Table 3, whereas, in columns (2) and (3), we report the estimates of the full model where the outcome equation (3) and the link formation equation (4) are estimated simultaneously. We see that, compared to Table 3, the results are substantially different. The coefficients are now smaller but, more importantly, the effect of peers-in-the-neighborhood on own GPA becomes insignificant. This suggests that the effects of peers who reside in the same neighborhood are mainly due to the *unobservable individual characteristics that drive the friendship formation of neighbors*. Therefore, there is no “real” impact of peers-in-the-neighborhood on own GPA. In other words, it is not because my friends who reside in my neighborhood are having good grades that I have myself good a grade but it is because we share some common unobservable characteristics that make us friends (for example, parents living in specific neighborhood have education and living standards that affect both the social life and the school performance of their children).

[Insert Table 4 here]

For completeness, Figure 6 shows the simulated posterior distributions for the peer effect parameters ϕ^{PS} (green curve) and ϕ^{PN} (blue curve) in grades (GPA). In Tables 3 and 4 the posterior distributions are summarized using means and standard deviations. Figure 6 provides confirmation that the estimated school peer effects parameter (ϕ^{PS}) is located to the right compared to the neighborhood effects parameter (ϕ^{PN}), i.e. the distribution of the estimates of ϕ^{PS} is shifted to the right. This means that the impact of peers at school is overall (i.e. for nearly the whole distribution) higher than that of the peers in the neighborhood.

[Insert Figure 6 here]

In order to further understand our results, we now refine our definition of neighborhood

by using the detailed information on each student’s residential location. To model the effects of distance between two individuals, we stratify our sample into four groups based on the quartiles of the distance distribution. Using a dyadic formulation, our model can be written in the following way:

$$y_{ir} = \sum_{\tau=1}^4 \delta_{\tau} D_{\tau} + \sum_{\tau=1}^4 \delta_{\tau} D_{\tau} \times y_{jr} + \gamma' X_{ir} + \eta_r + \epsilon_{ijr} \quad (5)$$

Let us refer to the student who is nominating as the “ego” and to the student who is receiving the nomination, as the “alter”. In such specification, individual i is the “ego” and j the “alter”. D_{τ} denotes a set of dummies that take value 1 if the alter lives within a distance τ from i and 0 otherwise. The threshold τ takes values equal to the quartiles of the empirical distribution of distance. In our case, the thresholds are 4, 9 and 15 kilometers. So the first dummy activates if the alter lives within 4 km from the ego, the second is equal to 1 if j lives between 4 and 9 km, and so on.

We begin our analysis by estimating this model on our sample while ignoring the friendship information. That is, in this specification, all the individuals who reside within τ kilometers from individual i are considered as i ’s peers. Table 5 contains the estimation results, with an increasing set of controls. Standard errors are clustered at the neighborhood level to remove any possible residual spatial autocorrelation. In columns (1), (2) and (3), we estimate (5) where friends are those who reside at a distance τ from student i , i.e. they are “fake” friends since they are not nominated by a student but just reside closer to him/her. On the contrary, in columns (4), (5) and (6), we estimate a peer effect model where friends are those nominated by student i . Even though we do not claim causality here, the patterns of the correlations are interesting. First, inspection of columns (1), (2) and (3) indicates that the correlation between individual and peers’ school performance weakens as distance increases. If we look at column (3), we see that the students whose gpa has an association with own grades are only the students who reside within 4 kilometers from each other. In other words, when *neighborhood fixed effects* are introduced, only highly localized interactions appear to be relevant.²² Second, if we now look at “real” peers, i.e. students who have been nominated as friends by student i (columns (4), (5) and (6) in Table 5), then we have an interesting result. Conditional on being friends, the correlation between friends’ GPA scores *is no longer decreasing with distance*. The introduction of neighborhood fixed effects (column

²²This is what is usually obtained in the urban/spatial economics literature where information about direct friendship is not available and only information about neighbors is available. See Andersson et al. (2009), Rosenthal and Strange (2008), Patacchini and Zenou (2012) and Wahba and Zenou (2005).

(3)) roughly reduces the coefficients by half but keeps the results qualitatively unchanged. In other words, being a “close” or “distant” friend does not change the peers’ correlation with own GPA.

In the following section, we present a theoretical mechanism that explains why this may be the case.

[Insert Table 5 here]

6 Theoretical framework: Choosing education effort and link formation

Our empirical investigation reveals that geographical proximity is not important in shaping peer effects. In other words, when considering as peers the chosen neighbors rather than the ones simply imputed on the basis of geographical boundaries, closer or distance peers have the same importance in shaping individual decisions.

The question that we would now like to investigate is what are the possible behavioral mechanisms behind our results.

Let us develop a simple model where individuals (students) reside in two distinct neighborhood l and k . As above, each individual i residing in neighborhood l is referred to as individual il . Assume that the *cost of interacting* with other individuals mainly depends on the physical distance between students.²³ We say that two students are *close friends* if both live in the same neighborhood, and *distant friends* if they reside in different neighborhood. We assume $C(il, jl) = 0$ and $C(il, jk) = c > 0$, i.e. the cost of interacting with close friends is normalized to zero. The *benefit of interacting* between i and j is denoted by $\Lambda(x_{il,jl}, \xi_{il,jl})$, where $x_{il,jl} \equiv |x_{il} - x_{jl}|$ is the difference between individuals il and jl in terms of an *observable* characteristic²⁴ and $\xi_{il,jl} \equiv |\xi_{il} - \xi_{jl}|$ is the difference between individuals il and jl in terms of an *unobservable* characteristic. We assume *homophily* in both observable and unobservable characteristics so that $\Lambda_1(\cdot) < 0$ and $\Lambda_2(\cdot) < 0$, which means that the higher is the distance in terms of observables and unobservables, the lower is the benefit of

²³See, for example, Patacchini et al. (2015) who found empirically that the intensity of social interactions between students negatively depends on the geographical distance between them.

²⁴For simplicity and without loss of generality, we only consider one characteristic $m = 1$ and one network $r = 1$.

interacting. As a result, the utility of individual il is given by:

$$\begin{aligned}
u_{il}(\mathbf{y}, g) &= \sum_{j=1}^n g_{ij}^{PS} [\Lambda(x_{il,jk}, \xi_{il,jk}) - c + \phi^{PS} y_{jk} y_{il}] \\
&\quad + \sum_{j=1}^n g_{ij}^{PN} [\Lambda(x_{il,jl}, \xi_{il,jl}) + \phi^{PN} y_{jl} y_{il}] \\
&\quad + (x_{il} + \eta_l + \xi_{il}) y_{il} - \frac{1}{2} y_{il}^2
\end{aligned}$$

where $g_{ij}^{PS} \equiv g_{il,jk}$ denotes a friendship link between individuals il and jk who live in different neighborhoods (neighborhood l for individual i and neighborhood k for individual j) and $g_{ij}^{PN} \equiv g_{il,jl}$ denotes a friendship link between students il and jl who reside in the same neighborhood. All the other variables have the same meaning as above. An explicit form of this utility function is given by:

$$\begin{aligned}
u_{il}(\mathbf{y}, g) &= \sum_{j=1}^n g_{ij}^{PS} \left[\frac{\beta^{PS}}{|x_{il} - x_{jk}|} + \frac{\delta^{PS}}{|\xi_{il} - \xi_{jk}|} - c + \phi^{PS} y_{il} y_{jk} \right] \\
&\quad + \sum_{j=1}^n g_{ij}^{PN} \left[\frac{\beta^{PS}}{|x_{il} - x_{jl}|} + \frac{\delta^{PN}}{|\xi_{il} - \xi_{jl}|} + \phi^{PN} y_{il} y_{jl} \right] \\
&\quad + (x_{il} + \eta_l + \xi_{il}) y_{il} - \frac{1}{2} y_{il}^2
\end{aligned} \tag{6}$$

If there were *no social interactions*, i.e. individual il has no link with anybody else ($g_{ij} = 0, \forall j$), then the optimal education effort would be given by:

$$y_{il} = x_{il} + \eta_l + \xi_{il}$$

This choice only depends on own observable and unobservable characteristics and on the characteristics of the neighborhood l .

If there are social interactions and some links are formed, then each individual il will optimally choose both education effort and friendship links.

Effort choice Assume that individual il chooses effort y_{il} . By maximizing (6), we obtain:

$$y_{il} = \phi^{PS} \sum_{j=1}^n g_{ij}^{PS} y_{j,k} + \phi^{PN} \sum_{j=1}^n g_{ij}^{PN} y_{j,l} + x_{il} + \eta_l + \xi_{il} \tag{7}$$

This is the outcome equation (1) that we empirically estimated in Section 4. In Appendix A, we develop the model when only this effort choice is made.

Link formation choice In order to determine the link-formation choice, we need to calculate the incremental payoff for individual il from changing her friendship status with individual jk who lives in *another neighborhood* from $g_{ij}^{PS} = 0$ to $g_{ij}^{PS} = 1$. By using (6), we obtain:

$$\Delta u_{il}^{PS} \equiv u_{il}(\mathbf{y}, g_{ij}^{PS} = 1) - u_{il}(\mathbf{y}, g_{ij}^{PS} = 0) = \frac{\beta^{PS}}{|x_{il} - x_{jk}|} + \frac{\delta^{PS}}{|\xi_{il} - \xi_{jk}|} - c + \phi^{PS} y_{jk} y_{il} \quad (8)$$

where y_{il} and y_{jk} are given by (6).

Similarly, the incremental payoff for individual il from changing her friendship status with individual jl who lives in the *same neighborhood* from $g_{ij}^{PN} = 0$ to $g_{ij}^{PN} = 1$ is given by:

$$\Delta u_{il}^{PN} \equiv u_{il}(\mathbf{y}, g_{ij}^{PN} = 1) - u_{il}(\mathbf{y}, g_{ij}^{PN} = 0) = \frac{\beta^{PN}}{|x_{il} - x_{jl}|} + \frac{\delta^{PN}}{|\xi_{il} - \xi_{jl}|} + \phi^{PN} y_{jl} y_{il} \quad (9)$$

This model is helpful for understanding our results. We have seen that if we estimate the outcome equation (7) taking as given the g_{ijs} , then both ϕ^{PS} and ϕ^{PN} are significant and positive. When instead we estimate together the outcome and the link formation equations, ϕ^{PN} became insignificant. Given our model of link formation, this means that there must be some advantage of being friend with students who are not residing in the same neighborhood (peers at school) compared to those who do (peers in the neighborhood). If we compare (9) and (8), we see that there is a disadvantage in terms of costs of interacting since $\Delta C_{il} \equiv C(il, jk) - C(il, jl) = c > 0$. It must then be that, compared to the peers in the neighborhood, the chosen friends in different neighborhoods (peers at school) must have an advantage in terms of observable characteristics, i.e. $|x_{il} - x_{jk}| < |x_{il} - x_{jl}|$, and/or in terms of unobservable characteristics, i.e. $|\xi_{il} - \xi_{jk}| < |\xi_{il} - \xi_{jl}|$. Given that we control for peers' characteristics in Table 4, and that Table 2 reveals that there are no substantial differences in terms of observable characteristics between close and distant friends, it has to be that the differences in unobservables, i.e. $|\xi_{il} - \xi_{jk}| < |\xi_{il} - \xi_{jl}|$, are driving the results. The more marked similarity with distant friends in terms of unobservable characteristics may result in similarity in behavior (effort) between distant friends. This mechanism may result in a failure to observe an attenuation of peer effects with geographical distance.

7 Social versus geographical distance

Let us now look more closely at the relative importance of the *geographical and social distance* between students in facilitating peer effects. In doing so, we also examine asymmetries in peer influences stemming from hierarchical characteristics in peer dyads, that is from the fact that an individual has an higher value of a characteristic (e.g. age, family income) than his/her peers.²⁵ Our baseline dyadic specification is:

$$y_{ir} = \alpha + \beta y_{jr} + \gamma' X_{ir} + \eta_r + \epsilon_{ijr} \quad (10)$$

where the coefficient β represents the effect of the alter's GPA on the ego's GPA.

We interact the alter's baseline GPA with two index variables, Δ^H and Δ^L , where Δ^H equals one in cases when the ego's *characteristic* has a *higher* status relative to her alter, and Δ^L equals one in cases when the ego's baseline characteristics has a *lower* status relative to her alter. The omitted category is therefore defined as “the similarity of the friend (i.e. the alter) at baseline”. Thus, our econometric specification, derived from (10) can be written as:

$$y_{ir} = \alpha + \beta y_{jr} + \delta \Delta^L \times y_{jr} + \lambda \Delta^H \times y_{jr} + \gamma' X_{ir} + \eta_r + \epsilon_{ijr} \quad (11)$$

We estimate model (11) separately for all characteristics. We say that a dyad of peers is *similar* if the ego and the alter are similar in terms of baseline characteristics and are *dissimilar* if they are not. We also say that there is an *ego-dominated dyad* when $\Delta^H = 1$ and an *alter-dominated dyad* when $\Delta^L = 1$. As a result, the coefficient β estimates the size of the peer effect in peer dyads where the ego and the alter are *similar* (i.e. $\partial y_{ir} / \partial y_{jr} = \beta$ when $\Delta^H = \Delta^L = 0$) while, in the two dissimilar cases, the peer effect coefficients is given by $\beta + \delta$ for alter-dominated dyads (i.e. $\partial y_{ir} / \partial y_{jr} = \beta + \delta$ when $\Delta^L = 1$ and $\Delta^H = 0$) and by $\beta + \lambda$ for ego-dominated dyads (i.e. $\partial y_{ir} / \partial y_{jr} = \beta + \lambda$ when $\Delta^L = 0$ and $\Delta^H = 1$).

If the peer influence is hierarchical, we expect that the size of the effect will be greater in dyads that are alter-dominated in related characteristics (such as income or parental education). Also, if there is similarity in characteristics such as gender or race in non-hierarchical dyads, then we would expect that the size of the peer effect in similar dyads will be greater than in dissimilar dyads. For the non-hierarchical characteristics, i.e. dichotomous variables that takes a value of 1 or 0 such as gender, race and distance (close or distant peers),

²⁵This exercise is similar in spirit to the paper by Yakusheva et al. (2014).

model (11) can be written as:

$$y_{ir} = \alpha + \beta y_{jr} + \xi \Delta^D \times y_{jr} + \gamma' X_{ir} + \eta_r + \epsilon_{ijr}, \quad (12)$$

where Δ^D is a dummy variable that takes value 1 if i and j are different in term of the binary characteristic (for example if i is a boy and j is a girl or if i is black and j is white, etc.). As a result, the coefficient $\beta + \xi$ estimates the size of the peer effect in peer dyads with dichotomous variables where the ego and the alter are *different*, i.e. $\partial y_{ir} / \partial y_{jr} = \beta + \xi$ when $\Delta^D = 1$.

Table 6 collects the results. One can see interesting asymmetries in the size of peer effects according to the values of the different characteristics. Baseline regression results are shown in Table C.3 in Appendix C.

Column (1) shows that adolescents who are younger (i.e. in lower grades) than their peers (*alter-dominated dyads* with $\Delta^L = 1$ and $\Delta^H = 0$) are more likely to be subject to peer influences than the opposite (i.e. *ego-dominated dyads*). The difference in magnitude, however, is not large (0.1577 for friends of different grades as compared to 0.128 for friends in the same grades). Peer influence seems to be hierarchical also with respect to religious practice (column (2)). For family income (column (3)), it appears that peer influence only exists for students with family income that lies within one standard deviation from one another. Interestingly, for gender (column (4)), we see that alter-ego in dissimilar dyads show higher peer effects than alter-ego in similar dyads (i.e. same gender), i.e. 0.1303 and 0.0949, respectively. For race (column (5)), we find no significant peer influences in interracial friendships. The last column of Table 6 reports the importance of geographical proximity for peer effects. Peer effects in *distant alter-ego dyads* are slightly higher than in closer ones. This evidence is consistent with our theoretical mechanism where friendships with distant friends are formed only if the expected value of the subsequent social interaction is sufficiently high. However, the role of geographic distance as opposed to the one of other socio-economic characteristics in shaping peer influence is limited. Indeed, the last row of Table 6 (% of total variance explained) reports how much similarity on a given characteristics contributes in explaining similarities in outcome (GPA). One can see that the geographical distance only explains 2.18% of the correlation in GPAs between students, which is lower than any other observable characteristics reported in Table 6. Race, family income or religion explains much more of the correlation in GPAs between friends.

[Insert Table 6 here]

8 Concluding remarks and policy implications

It is common wisdom that the Information Technology (IT) revolution has reduced the importance of geographical proximity. Many studies have argued that the emergence of the new technologies would make it possible to think of the world as a single global village. Recent research on social networks have shown evidence of contagious behavior over the chains of social contacts, even if the contacts are online acquaintances (Bond et al., 2012). However, as it is extremely difficult to obtain detailed data on social contacts as a function of geographical distance between them together with information on relevant socio-economic outcomes, there is no research about the relative importance of social versus geographical distance in shaping peer effects.²⁶

In this paper, we investigate the importance of social and geographical proximity in facilitating peer influence in education. For that, we propose an econometric methodology that addresses both neighborhood and network formation endogeneity. In particular, we address the non-random formation of students' friendship links by modelling the role of observable as well as unobservable characteristics in the link formation process and propose a model where education outcomes and friendship formation are jointly estimated. We find that the effect of peers-in-the-neighborhood on own grades (GPA) becomes insignificant once we control for individual-level unobserved characteristics. This suggests that the effect of peers who reside in the same neighborhood is mainly due to the unobservable common characteristics that drive the link formation process so that there is no "real" impact of peers-in-the-neighborhood on own GPA. On the contrary, peers at school have a strong and significant impact on own grades. This suggests that, in terms of education, social interactions between friends at school are more important than between friends who also reside close to each other.²⁷

In terms of policy implications, our analysis suggests that school-based policies could be more effective than place-based policies. We have seen in the Introduction that neighborhood policies such as the MTO programs produced no sustained improvements in academic achievement, educational attainment, risky behaviors, or labor market outcomes for either

²⁶Using an extensive dataset of 100,000 Facebook network users, Goldberg and Levy (2009) find that both the volume of email traffic and the density of social network contacts decrease with the distance. They do not investigate, however, whether short-range links and long-range links have a differential influence on individual decision making. Indeed, face-to-to-face meetings and social activities are important to develop friendship relationships and they are activities which are more costly for long-distance friends. As a result, even if the volume of contacts is decreasing with physical distance, the few long-distance friends that each individual keeps may be very important.

²⁷This is in accordance with Liu et al. (2014) who show the importance of social norms at school for adolescents in the United States.

female or male children, including those that were below school age at the time of random assignment (Fryer and Katz, 2013). Peer effects do matter in education outcomes and they mainly operate at schools. As a result, a school-based policy should improve the positive impact of peers. An example of a policy that has tried to change the social norm of students in terms of education is the *charter-school* policy. The charter schools are very good in screening teachers and at selecting the best ones. In particular, the “No Excuses policy” (Angrist et al., 2010, 2012) is a highly standardized and widely replicated charter model that features a long school day, an extended school year, selective teacher hiring, strict behavior norms, and emphasizes traditional reading and math skills. The main objective is to change the social norms of disadvantage kids by being very strict on discipline. Angrist et al. (2012) focus on special needs students that may be underserved. Their results show average achievement gains of 0.36 standard deviations in math and 0.12 standard deviations in reading for each year spent at a charter school.

Other school-based policies that aim at changing the social norms at school are the SEED schools, which combine a “No Excuses” charter model with a 5-day-a-week *boarding* program. These are America’s only urban public boarding schools for the poor. Using admission lotteries, Curto and Fryer (2014) show that attending a SEED school increases achievement by 0.211 standard deviation in reading and 0.229 standard deviation in math per year.

The Harlem Children’s Zone (HCZ) is a 97-block area in Harlem, New York, that combines “No Excuses” charter schools with neighborhood services designed to ensure the social environment outside of school is positive and supportive for children from birth to college graduation. The approach of HCZ is based on the assumption that one must improve both neighborhoods and schools to affect student achievement (Dobbie and Fryer 2011). Dobbie and Fryer (2012) use the random assignment nature of lottery admissions to determine the causal effect of being offered admission to the HCZ Promise Academy charter school on academic achievement and medium-term life outcomes. Dobbie and Fryer (2012) find that lottery winners have large and significant increases in math performance and marginal improvements in reading, and are 14.1 percentage points more likely to enroll in college.

We believe that our paper is one of the first that provides an explicit empirical analysis on the impact of social interactions at school and in the neighborhood on education outcomes. It provides a first stab at a very important question in both social networks, urban economics and education. More research should be performed to better understand these issues and to tease out precise policy recommendations.

References

- [1] Andersson, F., Burgess, S. and J. Lane (2009), “Do as the neighbors do: The impact of social networks on immigrant employment,” IZA Discussion Paper No. 4423.
- [2] Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A. and C.R. Walters (2010), “Inputs and impacts in charter schools: KIPP Lynn,” *American Economic Review Papers and Proceedings* 100, 239-243.
- [3] Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A. and C.R. Walters (2012), “Who benefits from KIPP?” *Journal of Policy Analysis and Management* 31, 837-860.
- [4] Ballester, C., Calvó-Armengol, A. and Y. Zenou (2006), “Who’s who in networks. Wanted: the key player,” *Econometrica* 74, 1403-1417.
- [5] Bayer, P., S.L. Ross and G. Topa (2008), “Place of work and place of residence: Informal hiring networks and labor market outcomes,” *Journal of Political Economy* 116, 1150-1196.
- [6] Beaman, L.A. (2012), “Social networks and the dynamics of labor market outcomes: Evidence from refugees resettled in the U.S.,” *Review of Economic Studies* 79, 128-161.
- [7] Blume, L.E., Brock, W.A., Durlauf, S.N. and Y.M. Ioannides (2011), “Identification of social interactions,” In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics*, Amsterdam: Elsevier Science, pp. 853-964.
- [8] Bond, R.M., Fariss, C.J., Jones, J. J., Kramer, A.D.I., Marlow, C., Settle, J. E. and J.H. Fowler (2012), “A 61-million-person experiment in social influence and political mobilization,” *Nature* 489, 295-298.
- [9] Bramoullé, Y., Djebbari, H. and B. Fortin (2009), “Identification of peer effects through social networks,” *Journal of Econometrics* 150, 41-55.
- [10] Bramoullé, Y., Kranton, R. and M. d’Amours (2014), “Strategic interaction and networks,” *American Economic Review* 104, 898-930.
- [11] Calvó-Armengol, A., Patacchini, E. and Y. Zenou (2009), “Peer effects and social networks in education,” *Review of Economic Studies* 76, 1239-1267.

- [12] Chib, S. (2001), “Markov chain Monte Carlo methods: Computation and inference,” In: J.J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Vol. 5*, Amsterdam: Elsevier Science, pp. 3569-3649.
- [13] Conley, T.G. and G. Topa (2002), “Socio-economic distance and spatial patterns in unemployment,” *Journal of Applied Econometrics* 17, 303-327.
- [14] Currarini, S., Jackson, M.O., and P. Pin (2009), “An economic model of friendship: Homophily, minorities, and segregation,” *Econometrica* 77, 1003-1045.
- [15] Currarini, S., Jackson, M.O., and P. Pin (2010), “Identifying the roles of race-based choice and chance in high school friendship network formation,” *Proceedings of the National Academy of Sciences of the USA* 107, 4857-4861.
- [16] Curto, V.E. and R.G. Fryer Jr. (2014), “The potential of urban boarding schools for the poor: Evidence from SEED,” *Journal of Labor Economics* 32, 65-93.
- [17] Cutler, D. and E.L. Glaeser (1997), “Are ghettos good or bad?” *Quarterly Journal of Economics* 112, 826-872.
- [18] Damm, A.P. (2014), “Neighborhood quality and labor market outcomes : Evidence from quasi-random neighborhood assignment of immigrants,” *Journal of Urban Economics* 79, 139-166.
- [19] Damm, A.P. and C. Dustmann (2014), “Does growing up in a high crime neighborhood affect youth criminal behavior?” *American Economic Review* 104, 1806-1832.
- [20] Dobbie, W. and R.G. Fryer (2011), “Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children’s Zone.” *American Economic Journal: Applied Economics* 3, 158-187.
- [21] Dobbie, W. and R.G. Fryer (2012), “Are high-quality schools enough to reduce social disparities? Evidence from the Harlem Children’s Zone?” Unpublished manuscript, Harvard University.
- [22] Dujardin, C., Peeters, D. and I. Thomas (2009), “Neighbourhood effects and endogeneity issues,” In: F. Bavaud and C. Mager (Eds.), *Handbook of Theoretical and Quantitative Geography*, Lausanne: Université

- [23] Durlauf, S.E. (2004), “Neighborhood effects,” In: J.V. Henderson and J-F. Thisse (Eds.), *Handbook of Regional and Urban Economics Vol. 4*, Amsterdam: Elsevier Science, pp. 2173-2242.
- [24] Edin, P.-A., P. Fredriksson and O. Aslund (2003), “Ethnic enclaves and the economic success of immigrants. Evidence from a natural experiment,” *Quarterly Journal of Economics* 118, 329-357.
- [25] Evans, W., Oates, W. and R. Schwab (1992), “Measuring peer group effects: A study of teenage behavior,” *Journal of Political Economy* 10, 966-991.
- [26] Fryer, R.G. and L.F. Katz (2013), “Achieving escape velocity: Neighborhood and school interventions to reduce persistent inequality,” *American Economic Review: Papers & Proceedings* 103, 232-237.
- [27] Glaeser, E.L., Sacerdote, B. and J. Scheinkman (1996), “Crime and social interactions,” *Quarterly Journal of Economics* 111, 508-548.
- [28] Goldenberg, J. and M. Levy (2009), “Distance Is not dead: Social interaction and geographical distance in the Internet era,” *Computers and Society* 2, 1-22.
- [29] Goldsmith-Pinkham, P. and G.W. Imbens (2013), “Social networks and the identification of peer effects,” *Journal of Business and Economic Statistics* 31, 253-264.
- [30] Graham, B.S. (2015), “Methods of identification in social networks,” *Annual Review of Economics*, forthcoming.
- [31] Hellerstein, J.K., McInerney, M. and D. Neumark (2011), “Neighbors and coworkers: The importance of residential labor market networks,” *Journal of Labor Economics* 29, 659-695.
- [32] Hellerstein, J.K., Kutzbach, M.J. and D. Neumark (2014), “Do labor market networks have an important spatial dimension?” *Journal of Urban Economics* 79, 39-58.
- [33] Helsley, R.W. and Y. Zenou (2014), “Social networks and interactions in cities,” *Journal of Economic Theory* 150, 426-466.
- [34] Hsieh, C.-S. and L.-F. Lee (2015), “A social interaction model with endogenous friendship formation and selectivity,” *Journal of Applied Econometrics*, forthcoming.

- [35] Imberman, S., Kugler, A. and B. Sacerdote (2012), “Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees,” *American Economic Review* 102, 2048-2082.
- [36] Ioannides, Y.M. (2011), “Neighborhood effects and housing,” In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics, Vol. 1B*, Amsterdam: Elsevier Science, pp. 1281-1340.
- [37] Ioannides, Y.M. (2012), *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton: Princeton University Press.
- [38] Ioannides, Y.M. and D.L. Loury (2004), “Job information networks, neighborhood effects, and inequality,” *Journal of Economic Literature* 42, 1056-1093.
- [39] Ioannides, Y.M. and G. Topa (2010), “Neighborhood effects: Accomplishments and looking beyond them,” *Journal of Regional Science* 50, 343-362.
- [40] Jackson, M.O. (2008), *Social and Economic Networks*, Princeton: Princeton University Press.
- [41] Jackson, M.O., Rogers, B.W. and Y. Zenou (2015), “The economic consequences of social network structure,” CEPR Discussion Paper 10406.
- [42] Jackson, M.O. and Y. Zenou (2015), “Games on networks,” In: P. Young and S. Zamir (Eds.), *Handbook of Game Theory, Vol. 4*, Amsterdam: Elsevier, pp. 91-157.
- [43] Jacob, B. (2004), “Public housing, housing vouchers and student achievement: Evidence from public housing demolitions in Chicago,” *American Economic Review* 94, 233-258.
- [44] Katz, L.F., Kling, J.R. and J.B. Liebman (2001), “Moving to opportunity in Boston: Early results of a randomized mobility experiment,” *Quarterly Journal of Economics* 116, 607-654.
- [45] Katz, L.F. (2015), “Reducing inequality: Neighborhood and school interventions,” *Focus* 31, 12-17.
- [46] Kling, J.R., Ludwig, J. and L.F. Katz (2005), “Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment,” *Quarterly Journal of Economics* 120, 87-130.

- [47] Kling, J.R., Liebman, J.B. and L.F. Katz (2007), “Experimental analysis of neighborhood effects,” *Econometrica* 75, 83-119.
- [48] Lavy, V.M., Paserman, D. and A. Schlosser (2012), “Inside the black box of ability peer Effects: Evidence from variation in low achievers in the classroom,” *Economic Journal* 122, 208-237.
- [49] Lee, L-F., Liu, X. and X. Lin (2010), “Specification and estimation of social interaction models with network structures,” *Econometrics Journal* 13, 145-176.
- [50] LeSage, J. P. and R.K. Pace (2009), *Introduction to Spatial Econometrics*, Boca Raton, FL: Taylor & Francis Group.
- [51] Lin, X. (2010), “Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables,” *Journal of Labor Economics* 28, 825-860.
- [52] Liu, X., Patacchini, E. and Y. Zenou (2014), “Endogenous peer effects: Local aggregate or local average?” *Journal of Economic Behavior and Organization* 103, 39-59.
- [53] Ludwig, J., Duncan, G.J., Gennetian, L.A., Katz, L.F., Kessler, R.C., Kling, J.R. and L. Sanbonmatsu (2013), “Long-term neighborhood effects on low-income families: Evidence from Moving to Opportunity,” *American Economic Review: Papers & Proceedings* 103, 226-231.
- [54] Manski, C.F. (1993), “Identification of endogenous effects: The reflection problem,” *Review of Economic Studies* 60, 531-542.
- [55] Meyer, C.D. (2000), *Matrix Analysis and Applied Linear Algebra*, Philadelphia: SIAM.
- [56] Oreopoulos, P. (2003), “The long-run consequences of living in a poor neighborhood,” *Quarterly Journal of Economics* 118, 1533-1575.
- [57] Overman, H.G., Gibbons, S. and E. Patacchini (2015), “Spatial methods,” In: G. Duranton, V. Henderson and W. Strange (Eds.), *Handbook of Regional and Urban Economics*, Vol. 5, Amsterdam: Elsevier Publisher, forthcoming.
- [58] Patacchini, E. and Y. Zenou (2012), “Ethnic networks and employment outcomes,” *Regional Science and Urban Economics* 42, 938-949.

- [59] Patacchini, E., Picard, P.M. and Y. Zenou, “Urban social structure, social capital and spatial proximity”, CEPR Discussion Paper No. 10501.
- [60] Picard, P.M. and Y. Zenou (2014), “Urban spatial structure, employment and social ties,” CEPR Discussion Paper No. 10030.
- [61] Robert, C.P. and G. Casella (2004), *Monte Carlo Statistical Methods*, New York: Springer-Verlag.
- [62] Rosenbaum, E. and L.E. Harris (2001), “Residential mobility and opportunities: Early impacts of the Moving to Opportunity demonstration program in Chicago,” *Housing Policy Debate* 12, 321-346.
- [63] Rosenthal, S.S. and W. C. Strange (2008), “The attenuation of human capital externalities,” *Journal of Urban Economics* 64, 373-389.
- [64] Ross, S. (1998), “Racial differences in residential and job mobility: Evidence concerning the spatial mismatch hypothesis,” *Journal of Urban Economics* 43, 112-135.
- [65] Ross, S. and Y. Zenou (2008), “Are shirking and leisure substitutable? An empirical test of efficiency wages based on urban economic theory,” *Regional Science and Urban Economics* 38, 498-517.
- [66] Sacerdote, B. (2001), “Peer effects with random assignment: Results from Dartmouth roommates,” *Quarterly Journal of Economics* 116, 681-704.
- [67] Sacerdote, B. (2011), “Peer effects in education: How might they work, how big are they and how much do we know thus far?”, In: E.A. Hanushek, S. Machin and L. Woessmann (Eds.), *Handbook of Economics of Education, Vol. 3*, Amsterdam: Elsevier Science, pp. 249-277.
- [68] Sacerdote, B. (2014), “Experimental and quasi-experimental analysis of peer effects: Two steps forward,” *Annual Review of Economics* 6, 253-272.
- [69] Sanbonmatsu, L., Kling, J.R., Duncan, G.J. and J. Brooks-Gunn (2006), “Neighborhoods and academic achievement: Results from the Moving to Opportunity experiment,” *Journal of Human Resources* 41, 649-691.
- [70] Sato, Y. and Y. Zenou (2015), “How urbanization affect employment and social interactions”, *European Economic Review* 75, 131-155

- [71] Schmutte, I.M. (2015), “Job referral networks and the determination of earnings in local labor markets,” *Journal of Labor Economics* 33, 1-32.
- [72] Solon, G., Page, M.E. and G.J. Duncan (2000), “Correlations between neighboring children in their subsequent educational attainment,” *Review of Economics and Statistics* 82, 383-392.
- [73] Topa, G. (2001), “Social interactions, local spillovers, and unemployment,” *Review of Economic Studies* 68, 261-295.
- [74] Topa, G. and Y. Zenou (2015), “Neighborhood versus network effects”, In: G. Duranton, V. Henderson and W. Strange (Eds.), *Handbook of Regional and Urban Economics*, Vol. 5, Amsterdam: Elsevier Publisher, forthcoming.
- [75] Wahba, J. and Y. Zenou (2005), “Density, social networks and job search methods: Theory and applications to Egypt,” *Journal of Development Economics* 78, 443-473.
- [76] Weinberg, B.A. (2000), “Black residential centralization and the spatial mismatch hypothesis,” *Journal of Urban Economics* 48, 110-134.
- [77] Yakusheva, O., Kapinos, K. A. and D. Eisenberg (2014), “Estimating heterogeneous and hierarchical peer effects on body weight using roommate assignments as a natural experiment,” *Journal of Human Resources* 49, 234-261.
- [78] Zenou, Y. (2013), “Spatial versus social mismatch,” *Journal of Urban Economics* 74, 113-132.
- [79] Zimmerman, D. (2003), “Peer effects in academic outcomes: Evidence from a natural experiment,” *Review of Economics and Statistics* 9-23.

Appendix

A Theoretical framework: Education effort only

Consider a population of n individuals.

The network $N = \{1, \dots, n\}$ is a finite set of agents. We keep track of social connections by a network g whose adjacency matrix is $\mathbf{G} = \{g_{ij}\}$, where $g_{ij} = 1$ if i and j are direct friends, and $g_{ij} = 0$, otherwise. Friendship is reciprocal so that $g_{ij} = g_{ji}$. We also set $g_{ii} = 0$ so that individuals are not linked to themselves. The adjacency matrix is thus a 0–1 symmetric matrix.

The *direct friends* of an individual i in a network g are denoted by $N_i(g)$. The *degree* of an agent i in a network g is the number of neighbors (here friends) that i has in the network, so that $d_i(g) = |N_i(g)|$.

In our data, for each individual i , his/her peers are the students that i has nominated as friends at school. This is captured by the adjacency matrix \mathbf{G} . We consider here two types of relationships in the network depending whether or not friends live in the same neighborhood. In other words, for each individual i , we decompose his/her peers (i.e. friends) between his/her *peers at school*, i.e. peers nominated by i at school but who do *not* live in the same neighborhood as i , and *peers in the neighborhood*, i.e. peers nominated by i at school who also reside in the same neighborhood.

Denote by \mathbf{G}^{PS} , the adjacency matrix of the peers at school and by \mathbf{G}^{PN} , the adjacency matrix of the peers in the neighborhood, with $\mathbf{G}^{PS} + \mathbf{G}^{PN} = \mathbf{G}$. To construct the peers-at-school adjacency matrix \mathbf{G}^{PS} , we put a 1 in the cell (i, j) of \mathbf{G} only if i and j are friends but do not live in the same neighborhood, i.e. $g_{ij}^{PS} = 1$. Similarly, to construct the peers-in-the-neighborhood adjacency matrix \mathbf{G}^{PN} , we put a 1 in the cell (i, j) of \mathbf{G} only if i and j are friends and reside in the same neighborhood, i.e. $g_{ij}^{PN} = 1$. Clearly, $\mathbf{G} = \mathbf{G}^{PS} + \mathbf{G}^{PN}$.

Preferences Individuals decide how much effort to exert in some activity. In the empirical section below, we will consider educational activities, i.e. grades. We denote the effort level of individual i by y_i and the population effort profile by $\mathbf{y} = (y_1, \dots, y_n)'$. We characterize the *peers-at-school* and *peers-in-the-neighborhood* relationships between two individuals by the *strength* of their relationship, denoted by ϕ so that there are two ϕ s denoted by ϕ^{PS} and ϕ^{PN} . Each agent i selects an effort $y_i \geq 0$, and obtains a payoff $u_i(\mathbf{y}, g)$

that depends on the effort profile \mathbf{y} and the underlying network g , in the following way:

$$u_i(\mathbf{y}, g) = (a_i + \eta + \varepsilon_i) y_i - \frac{1}{2} y_i^2 + \left(\phi^{PS} \sum_{j \in N_i^{PS}(g)} y_j + \phi^{PN} \sum_{j \in N_i^{PN}(g)} y_j \right) y_i \quad (13)$$

where $\phi^{PS}, \phi^{PN} > 0$. The structure of this utility function is now relatively standard in games on networks (Ballester et al., 2006; Calvó-Armengol et al., 2009; Jackson, 2008; Bramoullé et al., 2014; Jackson and Zenou, 2015) where there is an idiosyncratic exogenous part $(a_i + \eta + \varepsilon_i) y_i - \frac{1}{2} y_i^2$ and an endogenous peer effect aspect $\phi^{PS} \sum_{j \in N_i^{PS}(g)} y_i y_j + \phi^{PN} \sum_{j \in N_i^{PN}(g)} y_i y_j$. The main difference to the standard approach is that here, we have *heterogenous* peer effects since peers-at-school friends and peers-in-the-neighborhood friends have different impacts on own utility. Indeed, we have:

$$\frac{\partial u_i(\mathbf{y}, g)}{\partial y_i \partial y_j} = g_{ij}^{PS} \phi^{PS} + g_{ij}^{PN} \phi^{PN} \geq 0$$

Observe that η denotes the unobservable neighborhood characteristics, ε_i is an error term (observable by all individuals but not by the researcher) and there is also an ex ante *idiosyncratic heterogeneity*, a_i , which is assumed to be deterministic, perfectly *observable* by all individuals in the network and corresponds to the observable characteristics of individual i (like e.g. sex, race, parental education, etc.) and to the observable average characteristics of individual i 's best friends, i.e. the average level of parental education of i 's friends, etc. (contextual effects). To be more precise, a_i can be written as:

$$a_i = \sum_{m=1}^M \beta_m x_i^m + \frac{1}{g_i^{PS}} \sum_{m=1}^M \sum_{j=1}^n g_{ij}^{PS} x_j^m \gamma_m^{PS} + \frac{1}{g_i^{PN}} \sum_{m=1}^M \sum_{j=1}^n g_{ij}^{PN} x_j^m \gamma_m^{PN} \quad (14)$$

where x_i^m is a set of M variables accounting for observable differences in the individual characteristics of individual i , $\beta_m, \gamma_m^{PS}, \gamma_m^{PN}$ are parameters and $g_i^{PS} = \sum_{j=1}^n g_{ij}^{PS}$ and $g_i^{PN} = \sum_{j=1}^n g_{ij}^{PN}$ constitute the total number of peers-at-school friends and peers-in-the-neighborhood friends of individual i .

To summarize, when individual i exerts effort in some activity, the benefits of the activity depend on own effects (i.e. on individual characteristics a_i , some network characteristics η and some random element ε_i , which is specific to individual i and non-observable by the researcher) and on peer effects, where the strength of interactions differs between peers depending on whether or not they live in the same neighborhood. Note that the utility (13)

is concave in own decisions, and displays decreasing marginal returns in own effort levels. In sum,

$$u_i(\mathbf{y}, g) = \underbrace{(a_i + \eta + \varepsilon_i) y_i}_{\text{Benefits from own effort}} - \underbrace{\frac{1}{2} y_i^2}_{\text{Costs}} + \underbrace{\phi^{PS} \sum_{j=1}^n g_{ij}^{PS} y_i y_j}_{\text{Benefits from peers-at-school efforts}} + \underbrace{\phi^{PN} \sum_{j=1}^n g_{ij}^{PN} y_i y_j}_{\text{Benefits from peers-in-the-neighborhood efforts}}$$

Nash equilibrium We now characterize the Nash equilibrium of the game where agents choose their effort level $y_i \geq 0$ simultaneously. Let $\mathbf{x}_i = (x_i^1, \dots, x_i^M)'$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_M)'$, $\boldsymbol{\gamma}^{PS} = (\gamma_1^{PS}, \dots, \gamma_M^{PS})'$ and $\boldsymbol{\gamma}^{PN} = (\gamma_1^{PN}, \dots, \gamma_M^{PN})'$. In equilibrium, each agent maximizes her utility (13) and, using (14), we obtain the following best-reply function for each $i = 1, \dots, n$:

$$y_i = \phi^{PS} \sum_{j=1}^n g_{ij}^{PS} y_j + \phi^{PN} \sum_{j=1}^n g_{ij}^{PN} y_j + \mathbf{x}'_i \boldsymbol{\beta} + \frac{1}{g_i^{PS}} \sum_{m=1}^M \sum_{j=1}^{n^L} g_{ij}^{PS} \mathbf{x}'_j \boldsymbol{\gamma}^{PS} + \frac{1}{g_i^{PN}} \sum_{m=1}^M \sum_{j=1}^n g_{ij}^{PN} \mathbf{x}'_j \boldsymbol{\gamma}^{PN} + \eta + \varepsilon_i \quad (15)$$

Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$, $\mathbf{1}_n$ a column vector of 1 and \mathbf{X} an $(M \times M)$ matrix defined as

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}'_1 \\ \dots \\ \dots \\ \mathbf{x}'_M \end{pmatrix}$$

In matrix form, the best-reply function can be written as:

$$\mathbf{y} = \phi^{PS} \mathbf{G}^{PS} \mathbf{y} + \phi^{PN} \mathbf{G}^{PN} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{G}^{PS*} \mathbf{X} \boldsymbol{\gamma}^{PS} + \mathbf{G}^{PN*} \mathbf{X} \boldsymbol{\gamma}^{PN} + \eta \mathbf{1}_n + \boldsymbol{\varepsilon} \quad (16)$$

where \mathbf{G}^{PS*} and \mathbf{G}^{PN*} are the row-normalized adjacency matrices of \mathbf{G}^{PS} and \mathbf{G}^{PN} , respectively.

To determine the existence and uniqueness of the Nash equilibrium of this game, denote $\alpha_i = a_i + \eta + \varepsilon_i$ and the corresponding $(1 \times n)$ vector by $\boldsymbol{\alpha}$. Denote also by \mathbf{I} the $(n \times n)$ identity matrix. Then, the matrix form equivalent of (16) is:

$$\mathbf{y} = (\mathbf{I} - \phi^{PS} \mathbf{G}^{PS} - \phi^{PN} \mathbf{G}^{PN})^{-1} \boldsymbol{\alpha} \quad (17)$$

Denote by $\mu_1(\mathbf{G})$ the spectral radius of a matrix \mathbf{G} . We have the following result:

Proposition 1 *If $\mu_1(\phi^{PS}\mathbf{G}^{PS} + \phi^{PN}\mathbf{G}^{PN}) < 1$, then the peer effect game with payoffs (13) has a unique interior Nash equilibrium in pure strategies given by (16) or by (17).*

Proof of Proposition 1: We need to show that $\mathbf{I} - \mathbf{A}$ is non-singular (i.e. invertible), where $\mathbf{A} \equiv \phi^{PS}\mathbf{G}^{PS} + \phi^{PN}\mathbf{G}^{PN}$. We know that $\mathbf{I} - \mathbf{A}$ is non-singular if $\mu_1(\phi^{PS}\mathbf{G}^{PS} + \phi^{PN}\mathbf{G}^{PN}) < 1$ (see, e.g. Meyer, 2000, page 618). To prove the interiority of the solution, we can use exactly the same arguments as in the proof of Theorem 1 in Ballester et al. (2006). ■

This proposition totally characterizes the Nash equilibrium and gives a condition that guarantees the existence, uniqueness and interiority of this equilibrium.

To illustrate this result, consider the network described in Figure 1. The largest eigenvalue of $\phi^{PS}\mathbf{G}^{PS} + \phi^{PN}\mathbf{G}^{PN}$ is $\sqrt{(\phi^{PS})^2 + (\phi^{PN})^2}$ and thus the condition for existence and uniqueness is: $(\phi^{PS})^2 + (\phi^{PN})^2 < 1$. In that case, there exists a unique Nash equilibrium given by (17). Assume that $\alpha_1 = \alpha_2 = \alpha_3 = 1$ so that all individuals are ex ante identical, the only differences between them being their position in the network and the neighborhood where they live. Assume also that $\phi^{PN} = 0.8$ and $\phi^{PS} = 0.2$, which means that the benefits from social interactions are higher when the individuals reside in the same neighborhood.²⁸ The condition for the Nash equilibrium is satisfied since $(\phi^{PS})^2 + (\phi^{PN})^2 = 0.68 < 1$. It is easily shown that the unique Nash equilibrium is given by:

$$y_1^* = 6.25, y_2^* = 6 \text{ and } y_3^* = 2.25$$

Interestingly, even though individuals 2 and 3 have exactly the same position in the network by having one link each with individual 1, individual 2's effort is much higher than that of individual 3 because he/she lives in the same neighborhood as individual 1 and thus obtains more benefits from interacting with 1 than individual 3. This shows the importance of the neighborhood versus the network effect. If individual 3 had a network advantage because he/she had a more central position in the network (for example, by having more links) but still resided in a neighborhood isolated from his/her friends, then they will be a trade off between the network advantage and the neighborhood disadvantage. This is the main idea that we want to investigate in this paper by testing for ϕ^{PS} and ϕ^{PN} .

²⁸We don't assume it in the model. We will, however, test this hypothesis in the empirical analysis.

B Bayesian estimation

B.1 Setup of the bayesian model

For the Bayesian estimation of the parameters of our model, we follow closely Hsieh and Lee (2015). We extend, however, their model to account for multiple matrices and neighborhood fixed effects.²⁹ The model we are estimating is given by

$$\begin{aligned} y_{i,l,r} &= \phi^{PS} \sum_{j=1}^{n_r} g_{ij,r}^{PS} y_{j,l,r} + \phi^{PN} \sum_{j=1}^{n_r} g_{ij,r}^{PN} y_{j,l,r} + \mathbf{x}'_{i,l,r} \beta \\ &\quad + \frac{1}{g_{i,r}^{PS}} \sum_{j=1}^{n_r} g_{ij,r}^{PS} x'_{j,l,r} \gamma^{PS} + \frac{1}{g_{i,r}^{PN}} \sum_{j=1}^{n_r} g_{ij,r}^{PN} x'_{j,l,r} \gamma^{PN} + \eta_{l,r} + \epsilon_{i,l,r} \end{aligned}$$

which is (1) and (see (4))

$$P(g_{ij,r}^p = 1 | \nu_{ij,r} x_{i,l,r}, x_{j,l,r}, \xi_{i,l,r}, \xi_{j,l,r}) = \Lambda(\delta_0 + \delta_1 \nu_{ij,r} + \sum_{m=2}^{M+1} |x_{i,l,r}^m - x_{j,l,r}^m| \delta^m + \delta_{M+2} |\xi_{i,l,r} - \xi_{j,l,r}|) \quad (18)$$

We showed that joint normality for the unobservable $\xi_{i,l,r}$ and $\epsilon_{i,l,r}$ implies that (1) becomes

$$\begin{aligned} y_{i,l,r} &= \phi^{PS} \sum_{j=1}^{n_r} g_{ij,r}^{PS} y_{j,l,r} + \phi^{PN} \sum_{j=1}^{n_r} g_{ij,r}^{PN} y_{j,l,r} + \mathbf{x}'_{i,l,r} \beta \\ &\quad + \frac{1}{g_{i,r}^{PS}} \sum_{j=1}^{n_r} g_{ij,r}^{PS} x'_{j,l,r} \gamma^{PS} + \frac{1}{g_{i,r}^{PN}} \sum_{j=1}^{n_r} g_{ij,r}^{PN} x'_{j,l,r} \gamma^{PN} + \eta_{l,r} + \sigma_{\epsilon_{i,l,r} \xi_{i,l,r}} \xi_{i,l,r} + u_{i,l,r} \end{aligned}$$

which is (3).

Let $\Theta = (\phi^{PS}, \phi^{PN}, \beta', \gamma', \sigma_\epsilon^2, \sigma_{\epsilon\xi})'$ and $\delta = (\delta_0, \delta_1, \dots, \delta_{Q+2})$ be our parameter vectors of interest and let $\mathcal{D} = (y, X, G, N)$ be our observed data. The likelihood function for the whole model (3) and (18) is $\mathcal{L}(\mathcal{D} | \Theta, \delta) = \mathcal{L}(y, X | \Theta) \times \mathcal{L}(G | \delta, \xi)$. Let also $\pi(\Theta, \delta)$ a prior density for the parameters. The standard priors independence assumption allows us to write $\pi(\Theta, \delta) = \pi(\delta) \times \prod_{t=1}^T \pi_t(\Theta_t)$, where T is the dimension of Θ . By Bayes' theorem the posterior distribution is then

²⁹We thank Chih-Sheng Hsieh for sharing with us his code, which we modified and generalized.

$$p(\Theta, \delta | \mathcal{D}) \propto \mathcal{L}(\mathcal{D} | \Theta, \delta) \times \pi(\Theta, \delta). \quad (19)$$

Bayesian inference requires the computation of marginal distribution for all parameters. However, since this requires the integration of complicated distributions over several variables, a closed form solution is not readily available and Markov Chain Monte Carlo (MCMC) techniques are usually employed to obtain random draws from posterior distributions. The Gibbs sampling algorithm allows us to draw random sample values for each parameter from their marginal posterior distribution, given previous samples of other parameters. Once stationarity of the Markov Chain has been achieved, the random draws can be used to study empirical distributions of the posterior.³⁰

B.2 Prior choices and posterior distributions

For ϕ^{PS} and ϕ^{PN} , we choose an uninformative prior, that is a uniform distribution on the feasible parameter space. Such restriction, required by Gershgorin's theorem for the existence of a reduced form solution of the outcome equation implies that

$$|\phi^{PS}| + |\phi^{PN}| < \frac{1}{\max \left\{ \min \left[\max_i \sum_j g_{ij}^{PS}, \max_j \sum_i g_{ij}^{PS} \right], \min \left[\max_i \sum_j g_{ij}^{PN}, \max_j \sum_i g_{ij}^{PN} \right] \right\}}$$

For $B = (\beta', \gamma^{PN'}, \gamma^{PS'})'$ we choose a multivariate normal with mean zero and variance S_B . For δ , we also choose a multivariate normal, with mean zero and variance S_δ . In order to have a positive definite matrix

$$\Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\xi} \\ \sigma_{\epsilon\xi} & \sigma_\xi^2 \end{pmatrix}$$

we choose the prior for $\sigma_\epsilon^2, \sigma_{\epsilon\xi}$ a truncated bivariate normal over a feasible space with hyperparameters σ_0, Σ_0 , while σ_ξ^2 is restricted to 1 for identification purposes.³¹ For each neighborhood η_l , we choose a normal with mean zero and variance S_η . Finally, for $\xi_{i,r,l}$ we choose a standard normal. What follows are the marginal distributions we use for our gibbs sampling scheme.

By Bayes' theorem the marginal posterior distributions are:

$$P(\xi_{i,l,r} | \mathcal{D}, \Theta, \delta) \propto \prod_{r=1}^{\bar{r}} \prod_i^{n_r} \mathcal{N}(\xi_{i,l,r}) P(\mathcal{D} | \Theta, \delta, \xi) \quad (20)$$

³⁰For a complete reference on Monte Carlo methods in Bayesian econometrics, we refer to Chib (2001).

³¹See Hsieh and Lee (2013) for a discussion on identification of the model with selection on unobservables.

$$P(\delta|\mathcal{D}, \xi) \propto \mathcal{N}_{M+2}(\delta; 0, S_\delta) \prod_{r=1}^{\bar{r}} P(\mathbf{G}_r, \mathbf{X}_r | \xi, \delta) \quad (21)$$

$$P(\phi^{PN}, \phi^{PS} | \mathcal{D}, \xi, \Theta) \propto \prod_{r=1}^{\bar{r}} P(\mathbf{y}_r | \mathbf{G}_r^{PN}, \mathbf{G}_r^{PS}, \eta_{l,r}, \Sigma) \quad (22)$$

$$P(B | \mathcal{D}, \xi, \phi^{PN}, \phi^{PS}, \Sigma, \eta) \propto \mathcal{N}_{3M}(B; \tilde{B}, \tilde{S}_B) \quad (23)$$

$$P(\sigma_\epsilon^2, \sigma_{\xi\epsilon} | \mathcal{D}, \xi, \phi^{PN}, \phi^{PS}, B) \propto \mathcal{N}_2^T(\sigma_\epsilon^2, \sigma_{\xi\epsilon}; \sigma_0, \Sigma_0) \prod_{r=1}^{\bar{r}} P(\mathbf{y}_r | \mathbf{G}_r^{PN}, \mathbf{G}_r^{PS}, \eta_{l,r}, \Sigma) \quad (24)$$

$$P(\eta_{l,r} | \mathcal{D}, \xi, \phi^{PN}, \phi^{PS}, \Sigma, B) \propto \mathcal{N}(\eta_{l,r}, \tilde{\eta}_{l,r}; \tilde{S}_{\eta,l,r}) \quad (25)$$

A normal distribution for a general k -dimensional vector α with parameters α_0 and S_α is denoted by $\mathcal{N}_k(\alpha; \alpha_0, S_\alpha)$. Note that with normal priors and normal likelihood for the errors in the outcome equations, the distributions for B and $\eta_{l,r}$ have a normal distribution, as in the usual Normal-Normal bayesian linear model.

The posterior means and variance for the parameter B will thus be $\tilde{B} = \tilde{S}_B(\sum_{r=1}^{\bar{r}} \mathbf{X}_r' \mathbf{u}_r)$, where

$$\mathbf{u}_r = (I_r - \phi^{PN} \mathbf{G}_r^{PN} - \phi^{PS} \mathbf{G}_r^{PS}) \mathbf{y}_r - \mathbf{X}_r \beta - \mathbf{G}_r^{PN} \mathbf{X}_r \gamma^{PN} - \mathbf{G}_r^{PS} \mathbf{X}_r \gamma^{PS} - \sigma_{\epsilon\xi} \xi_r - \bar{\eta}_r$$

and $\tilde{S}_B = (S_B^{-1} + \sum_{r=1}^{\bar{r}} \mathbf{X}_r' \mathbf{X}_r)^{-1}$. We stacked the individual characteristics $\mathbf{x}_{i,r,l}$, the unobservable $\xi_{i,r,l}$, the outcome variable $y_{i,r,l}$ and the neighborhood fixed effect $\eta_{r,l}$ by neighborhoods to obtain the matrices \mathbf{X}_r and the vectors \mathbf{y}_r , ξ_r , η_r . Similarly, the means and variance for each neighborhood fixed-effect $\eta_{l,r}$ in network r are

$$\tilde{\eta}_{l,r} = \tilde{S}_{\eta,l,r} (\iota_{l,r}' ((I_{r,l} - \phi^{PN} \mathbf{G}_{r,l}^{PN} - \phi^{PS} \mathbf{G}_{r,l}^{PS}) \mathbf{y}_{r,l} - \mathbf{X}_{r,l} \beta - \mathbf{G}_{r,l}^{PN} \mathbf{X}_{r,l} \gamma^{PN} - \mathbf{G}_{r,l}^{PS} \mathbf{X}_{r,l} \gamma^{PS} - \sigma_{\epsilon\xi} \xi_{r,l}))$$

and

$$\tilde{S}_{\eta,l,r} = (S_\eta^{-1} + \iota_{l,r} \iota_{l,r}')^{-1}$$

B.3 Sampling algorithm

B.3.1 MCMC sampling

We can see from the previous subsection, that except the posterior distributions for B and η_r which are normal, every other distribution is of unknown form, so moments or quantiles cannot be computed. To overcome this issue, we use a MCMC algorithm in order to draw a random sample from target marginal distributions: when the Markov Chain converges to a stationary distribution, the chain itself can be used as a sample of draws from the target distribution: sample moments (i.e. mean and standard deviation) and empirical quantiles can be computed. The Gibbs sampling scheme allows us to marginalize a multivariate joint posterior like (19) without calculating multidimensional integrals, a task that is possible only for very specific functional forms. When the posterior distribution of some parameter is not easy to draw samples from, the Metropolis-Hastings (M-H) algorithm is usually employed. It has been shown (see for example Robert and Casella, 2004) that sampling a distribution with M-H within a Gibbs sampling scheme still produces an ergodic Markov Chain. What follows is a brief description of MCMC sampling methods with an example.

Let us suppose we have a distribution $q^*(x)$ of which we don't know its normalizing constant or its moments are very difficult (or even impossible) to compute. As said, the goal of MCMC methods is to draw samples from such target distribution in order to explore its characteristics empirically instead of analytically: sample moments and sample quantiles can be used instead of theoretical ones when the simulated distribution converges to the target one.

Gibbs sampling While Gibbs sampling can be shown to be a special case of the M-H algorithm, it's quicker and simpler although it needs the full conditional distribution to be known. Clearly this makes the Gibbs sampler alone less applicable. Let us suppose we need to draw a sample from $q(x_1, x_2, x_3)$. In our case this could be for example the full conditional distribution (19). The algorithm works as follows.

1. Chose x_1^0, x_2^0, x_3^0 as starting values of the chain and set an index $t = 0$
2. Draw x_1^{t+1} from $q(x_1|x_2^t, x_3^t)$
3. Draw x_2^{t+1} from $q(x_2|x_1^{t+1}, x_3^t)$
4. Draw x_3^{t+1} from $q(x_3|x_1^{t+1}, x_2^{t+1})$
5. Set $t = t + 1$ and get back to step 2.

When the chain $\{x_1\}$ converges to its stationary distribution $q(\cdot)$, its elements can be considered as a sample from the marginal conditional densities $q(x_i|x_{-i})$.

Metropolis-Hastings within Gibbs If any of the conditional distribution needed for the Gibbs sampling are not suitable for direct sampling (in our example it could be $q(x_3|x_1, x_2)$), Metropolis-Hastings (or Metropolis) can be employed in that step.

Metropolis-Hastings Let us suppose we cannot draw a random observation from $q(x_3|x_1, x_2)$ in the example above. Given a starting draw x_3^t , the M-H algorithm uses a *candidate point* y from a suitable *proposal distribution* $f(y|x_3^t)$. The candidate y is accepted as the *new* x_3^{t+1} with probability given by

$$\alpha(y, x_3^t) = \min \left[\frac{f(y|x_3^t) q(x_3^t|x_1^t, x_2^t)}{f(x_3^t|y) q(y|x_1^t, x_2^t)}, 1 \right].$$

Note that since the target distribution is on both the numerator and the denominator, normalizing constant does not matter at all for the sampling.

B.3.2 The algorithm step-by-step

Let $t = 1$ and let Θ^1, δ^1 be the initial state of the chain. We set all values at zero except σ_ϵ^2 set at 1. We denote as $\phi_k(z; \mu, \Sigma)$ the k -variate normal density with parameters μ, Σ evaluated at z , while $\phi(z)$ is the standard normal evaluated at z .

Step 1: Sample ξ^t from $\prod_{r=1}^{\bar{r}} \prod_i^{n_r} \phi(\xi_{i,l,r}) P(\mathcal{D}|\Theta, \delta, \xi)$ by a M-H step. Propose a candidate $\xi_{i,l,r}^c$ by drawing it from a Normal with mean $\xi_{i,r}^{t-1}$ and variance τ_ξ , which is a tuning parameters whose meaning will be explained below. Accept $\xi_{i,l,r}^c$ with probability α_ξ given by

$$\alpha_\xi = \min \left[\frac{\prod_i^{n_r} \phi(\xi_{i,l,r}^c) P(\mathcal{D}_r|\Theta^{t-1}, \delta^{t-1}, \xi^c)}{\prod_i^{n_r} \phi(\xi_{i,l,r}^{t-1}) P(\mathcal{D}_r|\Theta^{t-1}, \delta^{t-1}, \xi^{t-1})}, 1 \right]$$

Step 2: Sample δ^t from $\phi_{M+2}(\delta; \delta_0, S_\delta) \prod_{r=1}^{\bar{r}} P(\mathbf{G}_r, \mathbf{X}_r|\xi, \delta)$ by M-H. Propose δ^c drawing it from $\mathcal{N}_{M+2}(\delta^{t-1}; \tau_\delta)$ and accept it as the next state of the chain with probability α_δ given by

$$\alpha_\delta = \min \left[\frac{\phi_{M+2}(\delta^c; \delta_0, S_\delta) \prod_{r=1}^{\bar{r}} P(\mathbf{G}_r, \mathbf{X}_r|\xi^t, \delta^c)}{\phi_{M+2}(\delta^{t-1}; \delta_0, S_\delta) \prod_{r=1}^{\bar{r}} P(\mathbf{G}_r, \mathbf{X}_r|\xi^t, \delta^{t-1})}, 1 \right]$$

Step 3: Sample ϕ^{PN}, ϕ^{PS} from $\prod_{r=1}^{\bar{r}} P(\mathbf{y}_r | \mathbf{G}_r^{PN}, \mathbf{G}_r^{PS}, \bar{\eta}_r, \Sigma)$ by M-H. Draw ϕ^{PN^c}, ϕ^{PS^c} from $\mathcal{N}(\phi^{PN^{t-1}}, \phi^{PS^{t-1}}, \tau_\phi)$ until it satisfies the restriction discussed before. Accept ϕ^{PN^c}, ϕ^{PS^c} with probability α_ϕ given by

$$\alpha_\phi = \min \left[\frac{\prod_{r=1}^{\bar{r}} P(\mathbf{y}_r | \mathbf{G}_r^{PN}, \mathbf{G}_r^{PS}, \mathbf{X}_r, \phi^{PN^c}, \phi^{PS^c}, B^{t-1}, \bar{\eta}_r^{t-1}, \Sigma^{t-1})}{\prod_{r=1}^{\bar{r}} P(\mathbf{y}_r | \mathbf{G}_r^{PN}, \mathbf{G}_r^{PS}, \mathbf{X}_r, \phi^{PN^{t-1}}, \phi^{PS^{t-1}}, B^{t-1}, \bar{\eta}_r^{t-1}, \Sigma^{t-1})} \right]$$

Step 4: Sample B^t from $\mathcal{N}_{3M}(\tilde{B}, \tilde{S}_B)$, where $\tilde{B} = \tilde{S}_B(\sum_{r=1}^{\bar{r}} \mathbf{X}_r' \mathbf{u}_r^{t-1})$. Here \mathbf{u}_r^{t-1} denotes the residual vector calculated by using $\phi^{PN^t}, \phi^{PS^t}, \eta_r^{t-1}, \sigma_{\epsilon\xi}$

Step 5: Sample $(\sigma_\epsilon^2, \sigma_{\epsilon\xi})$ from $\mathcal{N}_2^T(\sigma_\epsilon^2, \sigma_{\epsilon\xi}; \sigma_0, \Sigma_0) \prod_{r=1}^{\bar{r}} P(\mathbf{y}_r | \mathbf{G}_r^{PN}, \mathbf{G}_r^{PS}, \eta_{l,r}, \Sigma)$. To do this, propose a candidate $(\sigma_\epsilon^{2^c}, \sigma_{\epsilon\xi}^c; \tau_\Sigma)$ from $\mathcal{N}_2(\sigma_\epsilon^{2^{t-1}}, \sigma_{\epsilon\xi}^{t-1})$, verify if it belongs to a set \mathcal{O} where Σ^c is positive definite and accept it with probability

$$\alpha_\Sigma = \min \left[\frac{\phi_2(\sigma_\epsilon^{2^c}, \sigma_{\epsilon\xi}^c; \sigma_0, \Sigma_0) \prod_{r=1}^{\bar{r}} P(\mathbf{y}_r | \mathbf{G}_r^{PN}, \mathbf{G}_r^{PS}, \mathbf{X}_r, \phi^{PN^t}, \phi^{PS^t}, B^t, \bar{\eta}_r^{t-1}, \Sigma^c)}{\phi_2(\sigma_\epsilon^{2^{t-1}}, \sigma_{\epsilon\xi}^{t-1}; \sigma_0, \Sigma_0) \prod_{r=1}^{\bar{r}} P(\mathbf{y}_r | \mathbf{G}_r^{PN}, \mathbf{G}_r^{PS}, \mathbf{X}_r, \phi^{PN^t}, \phi^{PS^t}, B^t, \bar{\eta}_{l,r}^{t-1}, \Sigma^{t-1})}, 1 \right]$$

Step 6: Sample $\eta_{l,r}$ from $\mathcal{N}(\eta_{l,r}, \tilde{\eta}_{l,r}; \tilde{S}_{\eta,l,r})$, where

$$\tilde{\eta}_{l,r} = \tilde{S}_{\eta,l,r} (l'_{l,r} ((I_{r,l} - \phi^{PN,t} \mathbf{G}_{r,l}^{PN} - \phi^{PS,t} \mathbf{G}_{r,l}^{PS}) \mathbf{y}_{r,l} - \mathbf{X}_{r,l} \beta^t - \mathbf{G}_{r,l}^{PN} \mathbf{X}_{r,l} \gamma^{PN,t} - \mathbf{G}_{r,l}^{PS} \mathbf{X}_{r,l} \gamma^{PS,t} - \sigma_{\epsilon\xi}^t \xi_{r,l}^t))$$

The tuning parameters $\tau_\xi, \tau_\phi, \tau_\delta, \tau_\Sigma$ are needed to achieve of roughly 50% of the M-H steps. To do this The Markov chains are 110,000 observations long with 20,000 observations of burn-in period. To test whether convergence has been achieved we performed the standard diagnostics such as autocorrelations, Rafterty-Lewis MCMC diagnostics, Geweke's numerical standard errors and relative numerical efficiency estimates, freely available within the CODA package.

C Extra tables

Table C.1. Estimation results: Outcome equation only (separated models)

	Dep. Var.: School Performance Index					
	(1)	(2)	(3)	(1)	(2)	(3)
Peers in the neighborhood (φ^{PN})				0.1477*** (0.0125)	0.1227*** (0.0132)	0.0671*** (0.0135)
Peers at school (φ^{PS})	0.1522*** (0.0095)	0.1466*** (0.0112)	0.1225*** (0.0117)			
Individual characteristics	yes	yes	Yes	yes	yes	yes
Peers' characteristics	no	yes	Yes	no	yes	yes
Neighborhood fixed effects	no	no	Yes	no	no	yes
N. obs.	838	838	838	838	838	838
N. networks	29	29	29	29	29	29
N. neighborhoods	82	82	82	82	82	82

Notes. We report estimates from a Bayesian spatial autoregressive model. Posterior means and standard deviation (in parentheses) are reported. Control variables are those listed in Table 1. *, **, *** indicate that zero is not contained in a 90, 95 and 99 percent confidence interval.

Table C.2. Estimation results: Outcome equation only (homogeneous model)

Dep. Var.: GPA Index	
	(1)
Endogenous peer effect (ϕ)	0.0937*** (0.0090)
Individual characteristics	yes
Peers' characteristics	yes
Neighborhood fixed effects	yes
N. obs.	838
N. networks	29
N. neighborhoods	82

Notes. We report estimates from a Bayesian spatial autoregressive model. Posterior means and standard deviation (in parentheses) are reported. Control variables are those listed in Table 1. *, **, *** indicate that zero is not contained in a 90, 95 and 99 percent confidence interval.

Table C.3 Heterogeneous peer effects-baseline.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Female	Black	Religion practice	Family income	Resident building quality	Student grade	Distance
β	0.130*** (0.0270)	0.125*** (0.0298)	0.149*** (0.0317)	0.121*** (0.0284)	0.183*** (0.0312)	0.128*** (0.0401)	0.106*** (0.0374)
ξ	-0.0354 (0.0404)	-0.102 (0.0989)	-	-	-	-	0.0178 (0.0424)
λ	-	-	-0.0811 (0.0589)	-0.0318 (0.0634)	-0.0979 (0.0612)	-0.123 (0.0777)	-
δ	-	-	-0.0422 (0.0424)	-0.0335 (0.140)	-0.182** (0.0715)	0.0296 (0.0724)	-
Female	0.185*** (0.0461)	0.188*** (0.0458)	0.183*** (0.0456)	0.187*** (0.0460)	0.180*** (0.0445)	0.185*** (0.0459)	0.189*** (0.0461)
Black	-0.0413 (0.0549)	-0.0437 (0.0540)	-0.0392 (0.0544)	-0.0429 (0.0552)	-0.0455 (0.0545)	-0.0417 (0.0555)	-0.0438 (0.0551)
Parental education	0.118* (0.0616)	0.119* (0.0617)	0.118* (0.0614)	0.118* (0.0619)	0.118* (0.0605)	0.120* (0.0617)	0.119* (0.0619)
Two parents	0.0718 (0.0486)	0.0717 (0.0484)	0.0730 (0.0484)	0.0719 (0.0479)	0.0687 (0.0499)	0.0741 (0.0487)	0.0712 (0.0484)
Family size	0.0194 (0.0456)	0.0199 (0.0452)	0.0210 (0.0460)	0.0188 (0.0450)	0.0195 (0.0441)	0.0179 (0.0450)	0.0191 (0.0454)
Religion practice	-0.0965** (0.0414)	-0.0974** (0.0416)	-0.0965** (0.0414)	-0.0974** (0.0415)	-0.0908** (0.0413)	-0.0955** (0.0417)	-0.0991** (0.0417)
Familyincome	0.209 (0.151)	0.205 (0.150)	0.214 (0.153)	0.208 (0.149)	0.197 (0.149)	0.206 (0.150)	0.205 (0.150)
Familyincomerefused	0.115 (0.131)	0.111 (0.130)	0.121 (0.132)	0.114 (0.130)	0.103 (0.131)	0.113 (0.130)	0.111 (0.131)
Student grade	0.0357 (0.0424)	0.0378 (0.0425)	0.0356 (0.0425)	0.0354 (0.0424)	0.0392 (0.0429)	0.0341 (0.0423)	0.0366 (0.0425)
Constant	0.0437*** (0.00602)	0.0414*** (0.00654)	0.0437*** (0.00580)	0.0423*** (0.00610)	0.0376*** (0.00598)	0.0422*** (0.00608)	0.0411*** (0.00652)
Observations	1,853	1,853	1,853	1,853	1,853	1,853	1,853
R-squared	0.075	0.074	0.075	0.074	0.081	0.076	0.074
Number of neighborhoods	82	82	82	82	82	82	82

Notes: We report OLS estimation results from model in equation (11): $y_i = \beta y_j + \delta \Delta_{ij}^L y_j + \lambda \Delta_{ij}^H y_j + \gamma' X_i + \varepsilon_{ij}$ in columns (3), (4), (5), (6), and model in equation (12): $y_i = \beta y_j + \xi \Delta_{ij}^D y_j + \gamma' X_i + \varepsilon_{ij}$ in columns (1), (2) and (7), respectively. Robust standard errors clustered at the neighborhood level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 1. Star network with two neighborhoods

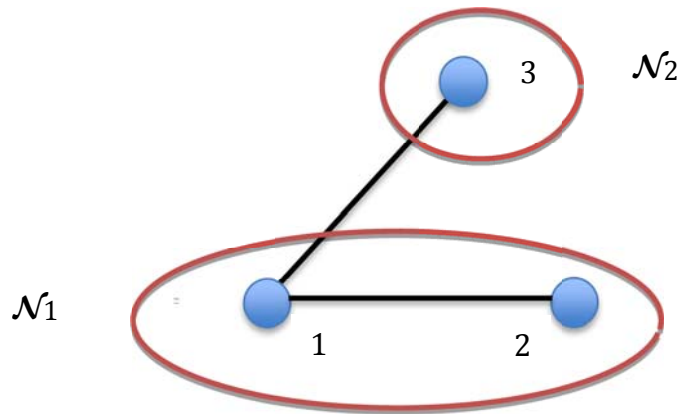


Figure 2. Distribution of neighborhoods by neighborhoods size

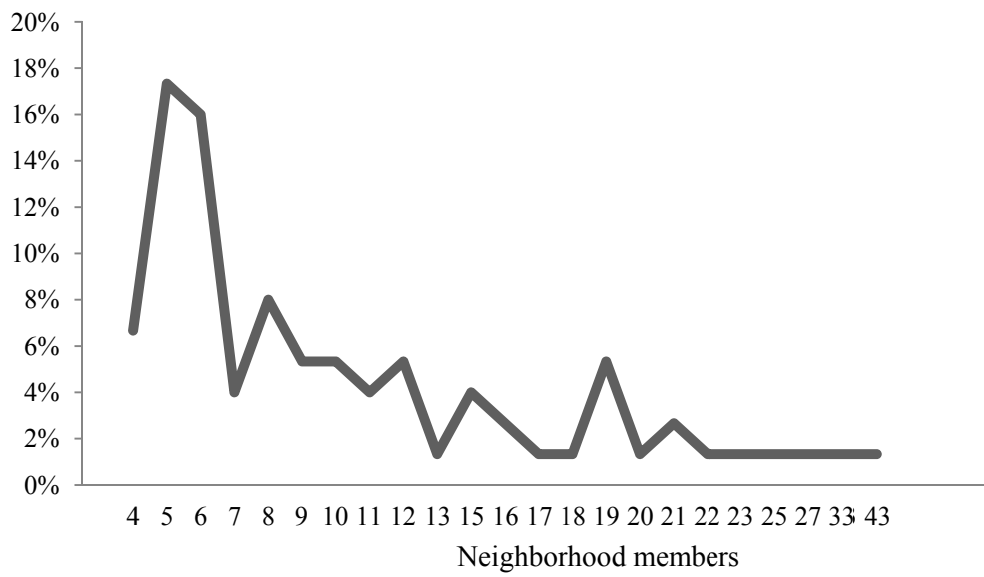


Figure 3. An heuristic description of our data structure

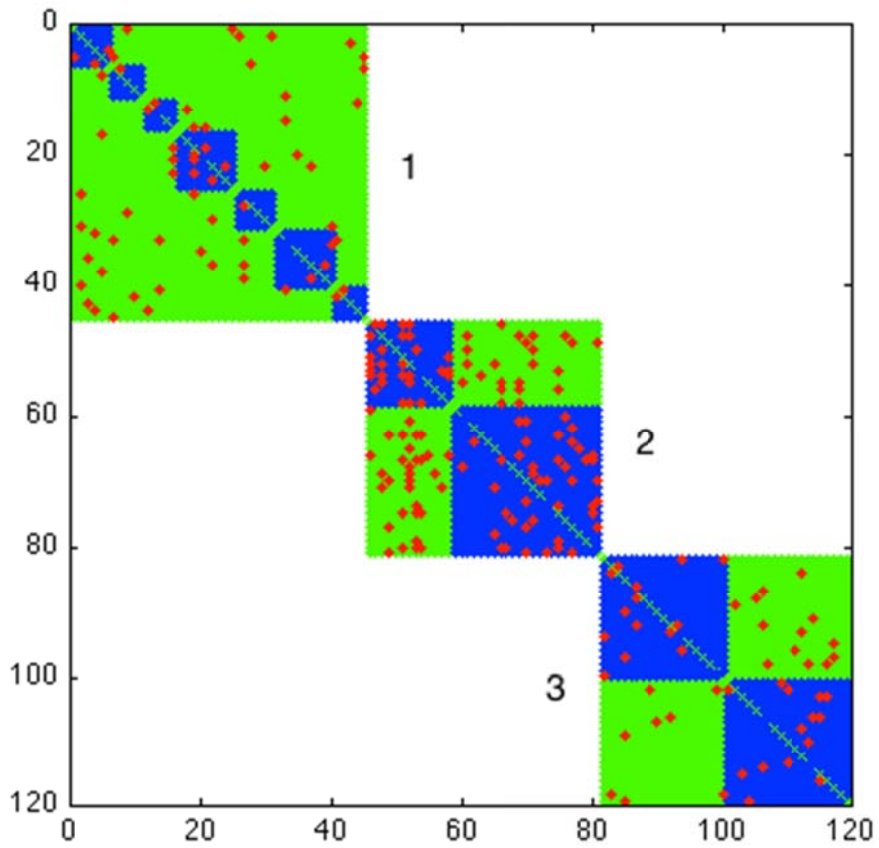


Figure 4. Distribution of friends by geographic distance

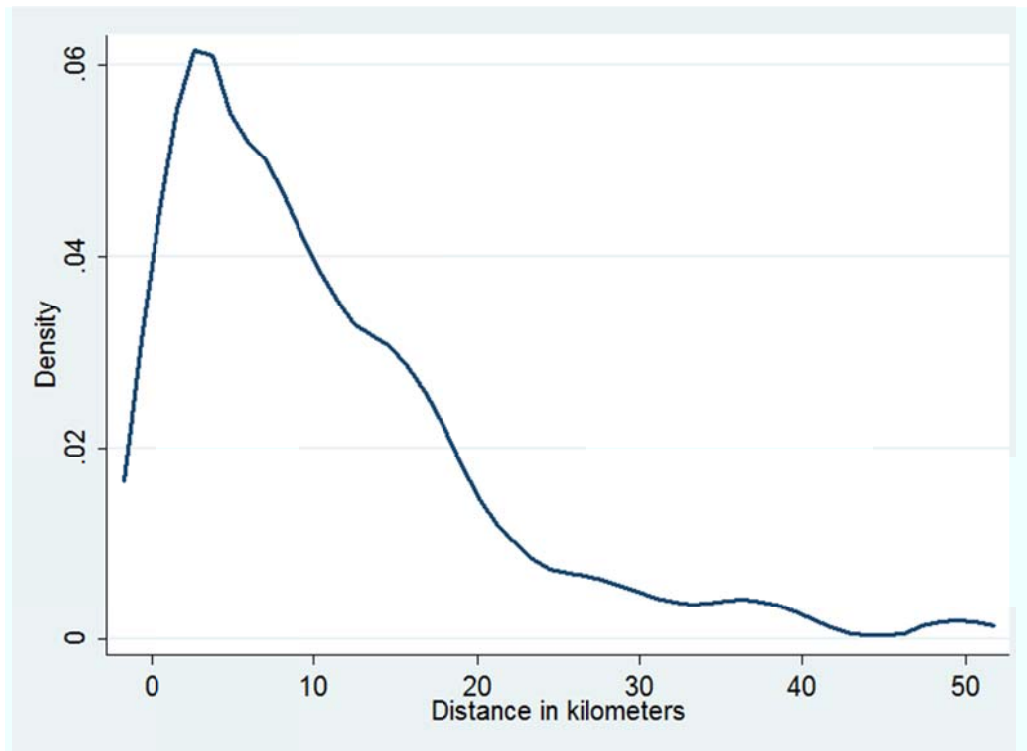


Figure 5. Distribution of neighborhoods by average distance between residents

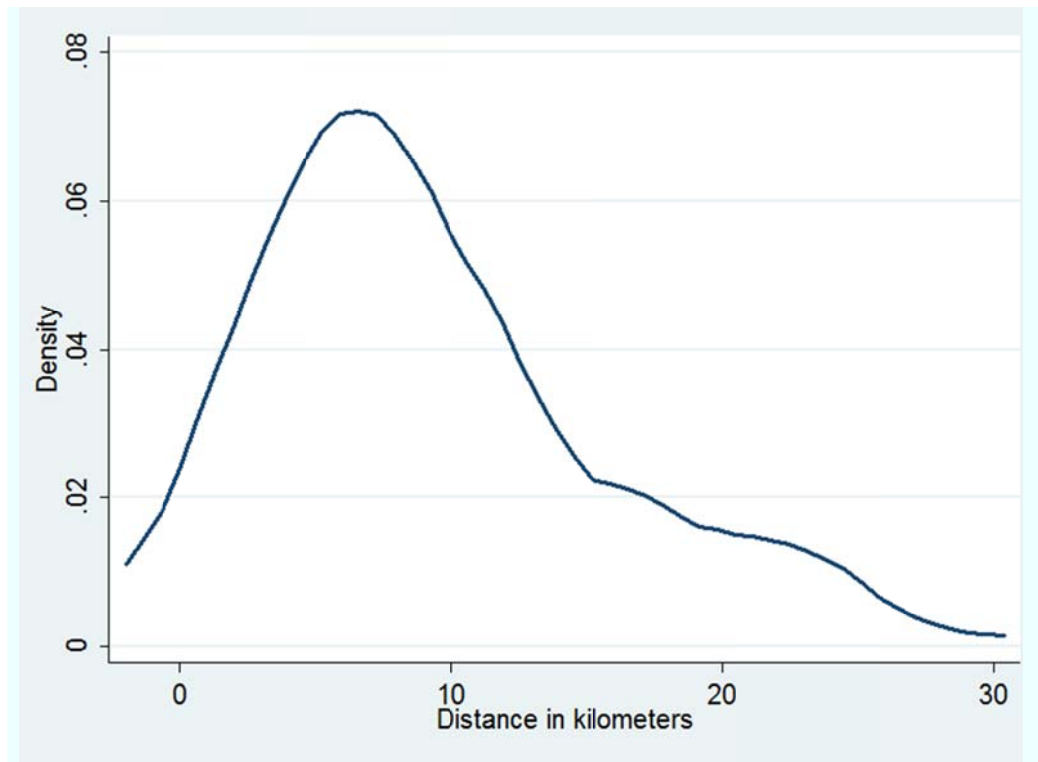


Figure 6. Kernel density estimates of posterior distributions
 φ^{PN} (Blue line) φ^{PS} (Green line)

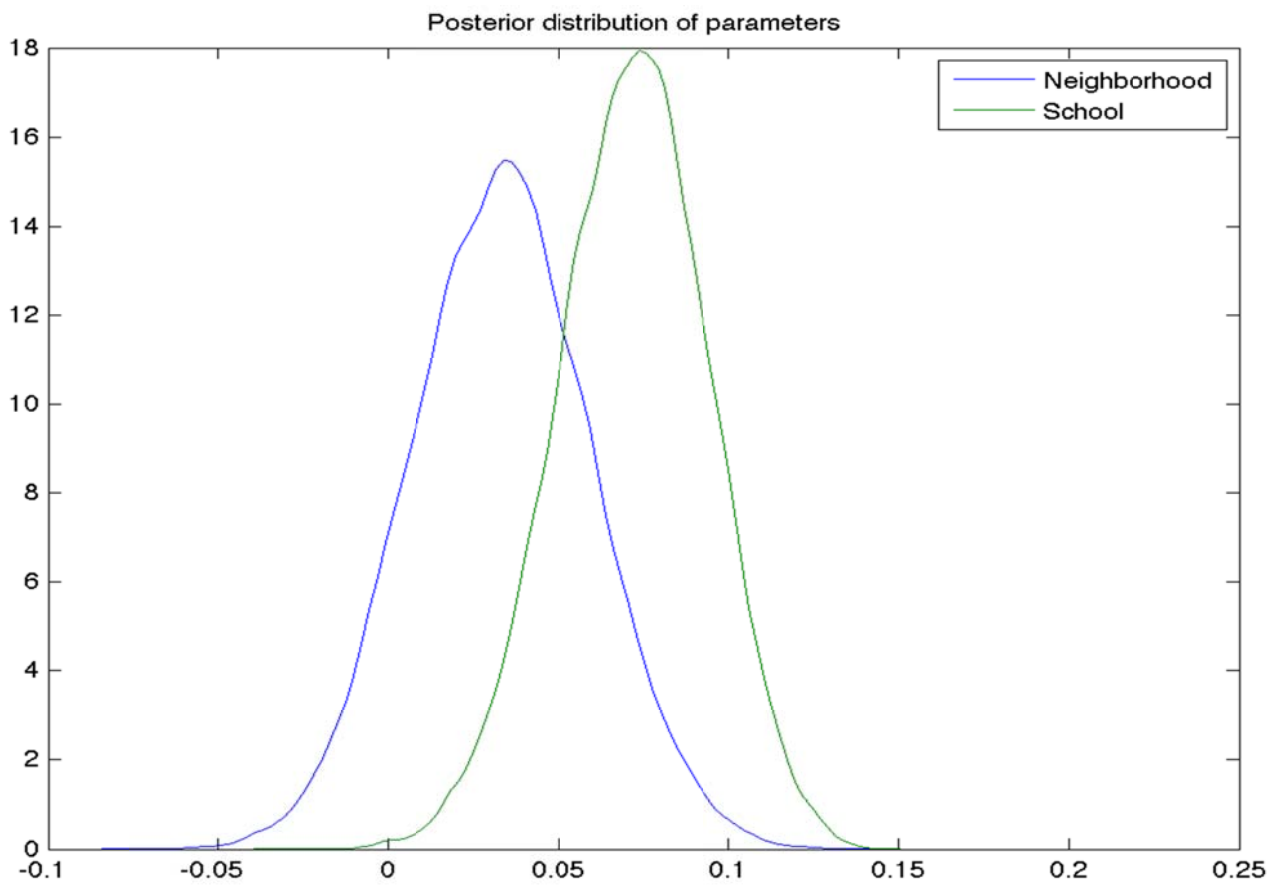
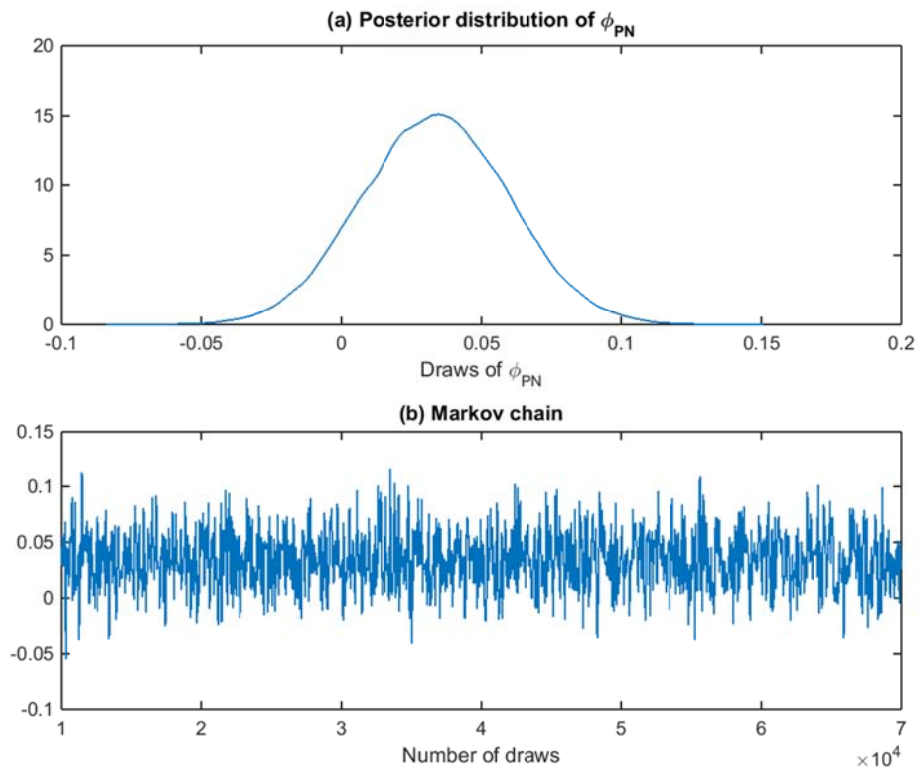
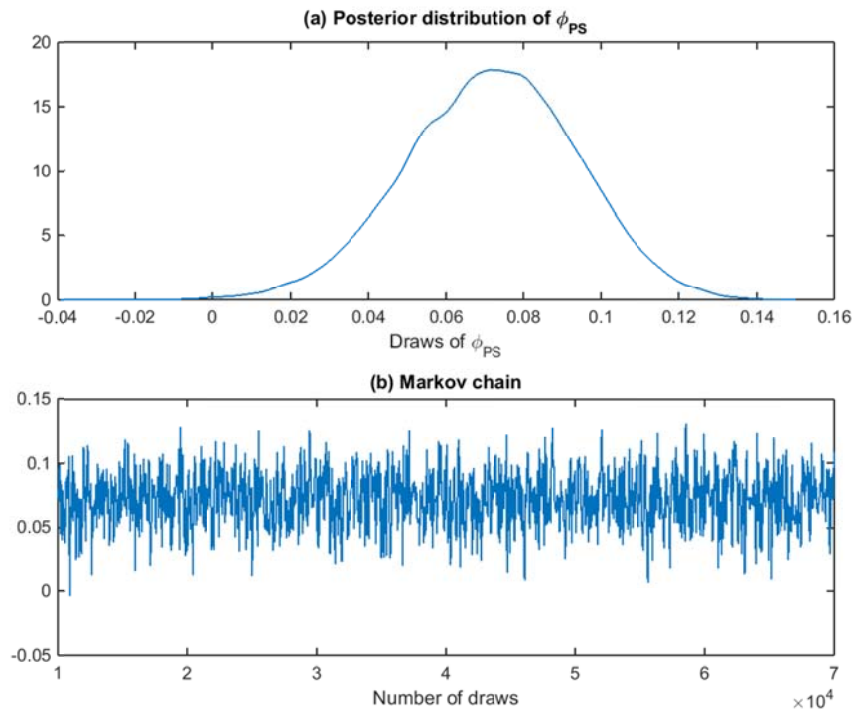


Figure 7. Kernel density estimates of posterior distributions (a) and draws from the ergodic Markov Chain (b) for parameter ϕ^{PN}



Notes: panel (a) shows the kernel density estimate of the posterior distribution of the parameter for peers in the neighborhood. Panel (b) depicts the Markov chain draws of the parameter. We show one value every 50 draws.

Figure 8. Kernel density estimates of posterior distributions (a) and draws from the ergodic Markov Chain (b) for parameter ϕ^{PS}



Notes: Panel (a) shows the kernel density estimate of the posterior distribution of the parameter for peers in the neighborhood. Panel (b) depicts the Markov chain draws of the parameter. We show one value every 50 draws.

Table 1. Description of Data (838 individuals, 29 networks)

	Variable definition	Mean	St.dev
Outcome variable			
School Performance (GPA) Index	Grade Point Average (GPA) for mathematics, English, history and science	2.93	0.75
Individual characteristics			
Female	Dummy variable taking value one if the respondent is female.	0.47	0.50
Black	Race dummy. "White" is the reference group.	0.21	0.41
Student grade	Grade of student in the current year.	9.34	3.02
Religion practice	Response to the question: "In the past 12 months, how often did you attend religious services", coded as 1= never, 2= less than once a month, 3= once a month or more, but less than once a week, 4= once a week or more.	2.89	1.16
Household size	Number of people living in the household.	3.55	1.91
Family income	Total family income in thousands of dollars, before taxes. It includes income of everybody in the household, and income from welfare benefits, dividends, and all other sources	49.40	52.77
Two parent family	Dummy taking value one if the respondent lives in a household with two parents(both biological and non biological).	0.17	0.38
Parent education	Schooling level of the (biological or non-biological) parent who is living with the child, distinguishing between "never went to school", "not graduate from high school", "high school graduate", "graduated from college or a university", "professional training beyond a four-year college", coded as 1 to 5. We considering only the education of the father if both parents are in the household.	3.65	2.23
Physical development	Answer to the question "How advanced is your physical development compared to other boys your age?" Coded from 1 (younger) to 4 (older).	3.31	1.11
Residential building quality	Interviewer response to the question "How well kept is the building in which the respondent lives", coded as 4= very poorly kept (needs major repairs), 3= poorly kept (needs minor repairs), 2= fairly well kept (needs cosmetic work), 1= very well kept.	1.44	0.75
Distance from school	Distance from school in kilometers.	9.18	1.32
Neighbors	Dummy that takes the value 1 if two individuals i and j live in the same Census block group.	0.28	0.45
Distance from friends	Distance between individual i and j if they are friends	10.51	9.27
Network characteristics			
Network size	Number of network members	28.89	25.88
Number of friends (outdegree)	Number of friends	2.26	1.88
Friends in the neighborhood	Number of friends in each neighborhood	0.91	1.11
Friends only at school	Number or friends only at school	1.35	1.35
Neighborhood characteristics			
Neighborhood size	Number of neighborhood (Census block group) members (in 82 neighborhoods)	10.21	7.47

Table 2. Percentage of friends sharing the same characteristics by geographical distance

<i>Panel 2a</i>	<i>Close friends</i>	<i>Distant friends</i>	<i>p-value</i>
% of friends with the same			
Female	0.58	0.56	0.392
Black	0.96	0.95	0.462
Parent education	0.43	0.44	0.503
GPA index	0.34	0.38	0.099*
Two parent family	0.64	0.63	0.591
Household size	0.27	0.25	0.520
Religion practice	0.42	0.44	0.503
Family income	0.78	0.76	0.228
Residential building quality	0.57	0.51	0.016**
Student grade	0.74	0.70	0.055*
<hr/>			
<i>Panel 2b</i>			
<hr/>			
% of friends with higher			
Parent education	0.28	0.28	0.924
GPA index	0.35	0.31	0.043**
Household size	0.37	0.36	0.813
Religion practice	0.29	0.28	0.674
Family income	0.09	0.10	0.624
Residential building quality	0.23	0.25	0.239
Student grade	0.14	0.17	0.070*

Notes. For family income, the category “same” is assigned when the difference between ego’s and alter’s family income is less than one standard deviation. GPA score is rounded to the closest integer. We report the p-values of tests for equality in means between the close and distant friends. *** p < 0.01, ** p < 0.05, * p < 0.10

Table 3. Estimation results: Outcome equation only

	Dep. Var.: GPA Index		
	(1)	(2)	(3)
Peers in the neighborhood (φ^{PN})	0.1289*** (0.0118)	0.0800*** (0.0094)	0.0486*** (0.0164)
Peers at school (φ^{PS})	0.1415*** (0.0095)	0.1195*** (0.0094)	0.1102*** (0.01)
Individual characteristics	yes	yes	yes
Peers' characteristics	no	yes	yes
Neighborhood fixed effects	no	no	yes
% variance explained by peers in the neighborhood	22.69%	16.64%	2.28%
% variance explained by peers at school	33.82%	27.80%	5.15%

Notes. We report estimates from a Bayesian spatial autoregressive model. Posterior means and standard deviation (in parentheses) are reported. Control variables are those listed in Table 1. *, **, *** indicate that zero is not contained in a 90, 95 and 99 percent confidence interval. N. obs. 838, n. networks 29, n. neighborhoods 82.

Table 4. Estimation results: Outcome equation and link formation

	Dep. Var.: GPA Index		
	(1)	(2)	(3)
	Outcome equation without link formation	Link formation	Outcome equation with link formation
Peers in the neighborhood (φ^{PN})	0.0486*** (0.0164)		0.0333 (0.0261)
Peers at school (φ^{PS})	0.1102*** (0.01)		0.0715*** (0.022)
Female	0.1744 *** (0.0202)	-0.153*** (0.0402)	0.1725*** (0.0552)
Black	0.1403*** (0.0263)	-0.5897*** (0.059)	0.1403*** (0.0705)
Physical development	0.0501*** (0.0139)	-0.0689*** (0.0307)	0.0478*** (0.0375)
Parent education	0.0818*** (0.0154)	-0.0153 (0.0328)	0.0834 *** (0.0419)
Two parent family	0.1347*** (0.0155)	0.003 (0.0252)	0.1346 *** (0.042)
Household size	-0.0224* (0.0147)	-0.0749*** (0.0334)	-0.0197 (0.0399)
Residential building quality	-0.0856*** (0.0151)	-0.0574* (0.0291)	-0.0889 ** (0.0405)
Religion practice	-0.025 (0.0211)	-0.0583 (0.0436)	-0.0248 (0.0574)
Family income	0.0885*** (0.0177)	-0.057 (0.0462)	0.0864 * (0.0473)
Distance from school	0.0117 (0.0211)	-0.1631*** (0.0425)	0.0098 (0.0567)
Student grade	0.0235 (0.0231)	-2.3471*** (0.0661)	0.0245 (0.063)
Neighbors		0.3439 *** (0.0571)	
Constant		-2.247*** (0.1061)	
Unobservables ($\sigma_{\varepsilon\xi}$)		0.8181*** (0.0488)	-0.0013 (0.0316)
Peers' characteristics	yes	-	yes
Neighborhood fixed effects	yes	-	yes
Obs	837	42012	837
N of networks	29	29	29
N of neighborhoods	82	82	82

Notes. Columns (2) and (3) report the means and the standard deviations of the posterior distributions of the parameters. We draw random samples from each parameter's marginal conditional distribution using Markov Chain Monte Carlo (MCMC) techniques. We let our chain run for 70,000 iterations, discarding the first 10,000 iterations. Ergodicity of the Markov Chain is achieved quite fast (see Figures 7 and 8). Control variables are described in Table 1. In column (2) the regressors are differences in terms of the listed variables between friends. Column (2) reports results on the dyadic model (18), the covariates are differences in terms of the listed characteristics. Columns (1) and (3) show results on model (20).

*, **, *** indicate that zero is not contained in a 90, 95 and 99 percent confidence interval.

Table 5. Spatial correlation decay - dyadic model

	Dep. Var.: Ego's GPA Index					
	(1)	(2)	(3)	(4)	(5)	(6)
	Whole sample			Only friends		
D1 (d <4km)	-0.0456 (0.0566)	-0.0208 (0.0547)		-0.0576 (0.0657)	-0.0604 (0.0610)	
D1 * alter's gpa	0.134*** (0.0225)	0.119*** (0.0211)	0.0341*** (0.00975)	0.244*** (0.0508)	0.223*** (0.0457)	0.124*** (0.0423)
D2 (4km<d<9km)	-0.0125 (0.0659)	0.0135 (0.0651)	0.0143 (0.0249)	0.0382 (0.0654)	0.0281 (0.0589)	0.00674 (0.0592)
D2 * alter's gpa	0.0875*** (0.0292)	0.0712*** (0.0269)	-0.0103 (0.00741)	0.230*** (0.0721)	0.195*** (0.0722)	0.108* (0.0592)
D3 (9km<d<15km)	0.0384 (0.0611)	0.0756 (0.0610)	0.0381 (0.0315)	0.0700 (0.0741)	0.0894 (0.0791)	0.0619 (0.0718)
D3 * alter's gpa	0.0628*** (0.0152)	0.0568*** (0.0163)	-0.0141 (0.00856)	0.232*** (0.0423)	0.232*** (0.0413)	0.159*** (0.0409)
D4 (d>15km)	0.0388 (0.0807)	0.0633 (0.0760)	0.0418 (0.0349)	0.102 (0.0627)	0.0917 (0.0683)	0.0911 (0.0806)
D4 * alter's gpa	0.0764*** (0.0196)	0.0667*** (0.0169)	-0.0109 (0.00901)	0.193*** (0.0561)	0.196*** (0.0570)	0.0912* (0.0474)
Controls	no	yes	yes	no	yes	yes
Neighborhood fixed effects	no	no	yes	no	no	yes
Obs	42,134	42,134	42,134	1,853	1,853	1,853
N of neighborhoods			82			82

Notes. Robust standard errors clustered at the neighborhood level are in parentheses. *, **, *** indicate statistical significance at the 10, 5 and 1 percent levels. For each network of size n_i , there are $n_i(n_i - 1)$ links, thus for 29 networks the whole sample size is $\sum_i n_i(n_i - 1)$.

Table 6. Heterogeneity in peer influences – dyadic model

	(1)	(2)	(3)	(4)	(5)	(6)
	Student grade	Religion practice	Family income	Female	Black	Distance
Ego = Alter (β)	0.1280*** (0.0401)	0.149*** (0.0316)	0.1208*** (0.0282)	0.0949** (0.040)	0.1247*** (0.0298)	0.1061*** (0.0374)
Ego \neq Alter ($\beta + \xi$)	-	-	-	0.1303*** (0.027)	0.0227 (0.0921)	0.1241*** (0.0353)
Ego > Alter ($\beta + \lambda$)	0.0048 (0.0490)	0.0679 (0.0524)	0.0931 (0.0682)			
Ego < Alter ($\beta + \delta$)	0.1577*** (0.568)	0.1068*** (0.0389)	0.0877 (0.1333)			
% of total variance explained	5.2%	4.22%	4.48%	3.79%	4.77%	2.18%

Notes. We report estimated peer effect sizes and standard errors for each ego / alter category (“ego < alter”, “ego > alter”, and “ego = alter”), based on regression coefficients from the regressions in Table C.3. We test for significance of the sum of parameters. The estimates of the regression coefficients are shown in Table C.3. For family income, category “similar” is assigned when the friends’ family income is within one standard deviation of one another. Student grade and religion practice are categorical variables with greater values corresponding to being in a higher grade and attending religious services more frequently, respectively. For distance, category “similar” is assigned when alter lives within 8.3 kilometers, which is the sample median. All models control for ego’s sex, race, parental education, religion practice, parental income, student grade, family size, two-parents households. The percentage of total variance is the cumulative effect of the variables related to the listed characteristics. Standard errors clustered at the neighborhood level in parentheses. *** p<0.01, ** p<0.05, * p<0.1