Heterogeneous Peer Effects in Education*

Eleonora Patacchini† Edoardo Rainone‡ Yves Zenou§

January 27, 2016

Abstract

We investigate whether, how, and why individual education attainment depends on the educational attainment of schoolmates. Specifically, using longitudinal data on students and their friends during the school years in a nationally representative set of US schools, we consider the role of different types of peers on education outcomes. We find that there are strong and persistent peer effects in education, but peers tend to be influential in the long run only when their friendships last more than a year. This evidence is consistent with a network model where convergence of preferences and the emergence of social norms among peers need long-term interactions.

Key words: Spatial autoregressive model, heterogenous spillovers, 2SLS estimation, Bayesian estimation, education

JEL Classification: C31, D85, Z13.

---

*We thank Francine Blau, Larry Blume, Alfonso Flores-Lagunes, Xiaodong Liu, Lawrence Khan, Guido Kuersteiner, Francesca Molinari, Ingman Prucha, Theodoros Rapanos, John Rust and Frank Vella for valuable comments and discussions. This 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 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 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.

†Cornell University, EIEF, IZA and CEPR. E-mail: ep454@cornell.edu.

‡Bank of Italy. E-mail: edoardo.rainone@bancaditalia.it. The views expressed here do not necessarily reflect those of the Bank of Italy.

§Stockholm University, IFN and GAINS. E-mail: yves.zenou@ne.su.se.
1 Introduction

The influence of peers on education outcomes has been widely studied both in economics and sociology (Sacerdote, 2011). Yet many questions remain unanswered. In particular, very little is known about the effect of school peers on the long-run outcomes of students. This is primarily due to the absence of information on peers during teenage years together with long-run outcomes of individuals in most existing data. Besides, the mechanisms by which the peer effects affect education are unclear.

In this paper, we provide an analysis on the long-run effects of high-school peers on years of schooling and put forward the role of different types of ties for educational outcomes.

For that, we use the unique information on friendship networks among students in the United States provided by the Addhealth data. We exploit three unique features of the AddHealth data: (i) the nomination-based friendship information, which allows us to reconstruct the precise geometry of social contacts during high-school years, (ii) the variation in friendship network topology between Wave I and Wave II, which enables us to distinguish between short-lived ties and long-lived ties and (iii) the longitudinal dimension, which provides information about each individual and her/his friends’ outcomes during the adulthood.

More specifically, we use the different waves of the AddHealth data by looking at the impact of school friends nominated in the first two waves in 1994-1995 and in 1995-1996 on own educational outcome (when adult) reported in the fourth wave in 2007-2008 (measured by the number of completed years of full time education). We define a long-lived tie relationship between two students if they have nominated each other in both waves (i.e. in Wave I in 1994-1995 and in Wave II in 1995-1996) and a short-lived tie relationship if they have nominated each other in one wave only. We also study the robustness of our results in terms of the definition and measurement of long and short-lived ties.

Our results show that there are strong and persistent peer effects in education. When looking at the role of short-lived and long-lived ties in educational decisions, it appears that the education decisions of short-lived ties have no significant effect on individual long-run outcomes, regardless of whether peers are interacting in lower or higher grades. On the

1 The constraints imposed by the available disaggregated data force many studies to analyze peer effects in education at a quite aggregate and arbitrary level, such as at the high school (Evans et al., 1992), the census tract (Brooks-Gunn et al., 1993), and the ZIP code level (Datcher, 1982; Corcoran et al., 1992) where individuals reside. The importance of peer effects as distinct from neighborhood influences is still a matter of debate in many fields (see, e.g., the literature surveys by Durlauf, 2004, Ioannides and Topa, 2010, and Ioannides, 2011, 2012).

2 The economics of networks is a growing field. For overviews, see Jackson (2008), Blume et al. (2011), Ioannides (2012), Boucher and Fortin (2015), Graham (2015), and Jackson and Zenou (2015).
contrary, we find that the educational choices of long-lived ties have a positive and significant effect on own educational outcome.

There is a large literature on the role of different ties in the labor market. In particular, Granovetter (1973, 1974, 1983) initiated a strand of studies looking at the effects of weak versus strong ties. Strong ties are viewed as stable relationships and weak ties as unstable relationships.\(^3\) Interestingly, compared to the literature on the labor market, we find the opposite result for education outcomes.\(^4,5\) Indeed, we show that stable rather than unstable ties matter for education. This is reasonable given that outcomes and mechanisms are different in the two contexts. While random encounters may be helpful in providing information about jobs, they typically do not contribute to shape social norms, values and attitudes (see, e.g. Coleman, 1988, Wellman and Wortley, 1990). The collective value of “social networks”, which is a relevant driver of long-run influences, need time and repeated interactions to be established (Putnam, 2000).

In line with these ideas, we propose a theoretical model that is able to interpret our evidence. We consider a dynamic network model (DeGroot, 1974) where there are two states of the world (or social norms): \{It is worth continuing studying\} and \{It is not worth continuing studying\}, which are unknown to the agents. Agents embedded in a network update their beliefs by repeatedly taking the weighted average of their neighbors’ beliefs. We extend the DeGroot model by differentiating between short-lived friends and long-lived friends. We define short-lived friends as students who interact with each other only once and long-lived friends as students who interact repeatedly. Because short-lived friendships only interact once, they will influence the beliefs of each other only in the initial period. On the contrary, long-lived friends interact repeatedly and thus update their beliefs all the time as in the standard DeGroot model by repeatedly taking the weighted average of their (long-lived) neighbors’ beliefs. We show how all students in the network reach a consensus in the long run and why long-lived friends have more impact over the resulting social norm,

---

\(^3\) In his seminal papers, Granovetter defines weak ties in terms of lack of overlap in personal networks between any two agents, i.e. weak ties refer to a network of acquaintances who are less likely to be socially involved with one another. Formally, two agents A and B have a weak tie if there is little or no overlap between their respective personal networks. Vice versa, the tie is strong if most of A’s contacts also appear in B’s network.

\(^4\) Yakubovich (2005) uses a large scale survey of hires made in 1998 in a major Russian metropolitan area and finds that a worker is more likely to find a job through weak ties than through strong ties. These results come from a within-agent fixed effect analysis, so they are independent of workers’ individual characteristics. Using data from a survey of male workers from the Albany NY area in 1975, Lin et al. (1981) find similar results. Lai et al. (1998) and Marsden and Hurlbert (1988) also find that weak ties facilitate reaching a contact person with higher occupational status who, in turn, leads to better jobs, on average.

\(^5\) See also Patacchini and Zenou (2008) who find evidence of the strength of weak ties in crime.
as shown by our empirical results.

We also collect additional evidence, which remains in line with this mechanism. First, we differentiate between short-lived friends only nominated in Wave I (lower grades) and those only nominated Wave II (later grades). We find that short-lived friends have no impact on future education studies, irrespective of whether short-lived friends have been nominated earlier or later in the school years. This is in accordance with our theoretical model where short-lived friends only affect others’ beliefs for one period, independently if it is the first or second period. Second, we investigate the difference between long-run and short-run effects of peers on education. We show that, while in the long run, only long-lived ties matter, we find that, in the short run, both short-lived and long-lived ties are important in determining a student’s performance at school. Our theoretical model can explain this result since both short-lived and long-lived friends affect the initial beliefs of studying (short run) but only long-lived friends affect the emergence of a long-term social norm favorable to higher-education studies.

There are very few studies looking at the long-run effects of friendship on human capital accumulation. Using the Wisconsin Longitudinal Study of Social and Psychological Factors in Aspiration and Attainment (WLS), Zax and Rees (2002) were the first to analyze the role of friendships in school on future earnings. Using the AddHealth data, Bifulco et al. (2011) study the effect of school composition (percentage of minorities and college educated mothers among the students in one’s school cohort) on high-school graduation and post-secondary outcomes.

In the existing (enormous) literature on peer effects, the wording "heterogenous peer effects" is typically used to indicate that the individual response to peers’ characteristics vary by the level of the characteristic or that peer effects are different for different types of individuals (e.g. males versus females, whites versus blacks). The situation that the individual response to peers’ behavior may vary by peer type is commonly overlooked. A notable exception is Goldsmith-Pinkham and Imbens (2013) who use different network structures as a statistical exercise to investigate measurement errors in peer status. They estimate the model using a bayesian approach.

In this paper, we extend the Liu and Lee (2010) 2SLS approach to a network model with different interaction matrices. The asymptotic consistency and efficiency of the proposed estimators are proved. We also employ a Bayesian inferential method to integrate a network formation model with the study of behavior over the formed networks. Finally, we consider

---

6 See Sacerdote (2014) for a recent review.
7 See e.g. Griffith and Rask (2014), Tincani (2015) and Yakusheva et al. (2014).
possible measurement errors in peer groups using a simulation experiment. Our results are robust to various types of network topology misspecification.

The paper unfolds as follows. Our data are described in Section 2, while the estimation and identification strategy is discussed in Section 3. Section 4 collects the empirical evidence. Section 5 investigates the economic mechanisms behind our peer-effects results by first proposing a theoretical model (Section 5.1) and then by providing more empirical results differentiating long-lived friends between those only nominated in Wave I (lower grades) and those only nominated Wave II (later grades) (Section 5.2). Section 6 shows the robustness of our results with respect to network formation and network topology misspecification, while Section 7 consider short-run and long-run effects of peers on education. Finally, Section 8 concludes the paper.

2 Data description

Our analysis is made possible by the use of a unique database on friendship networks from the National Longitudinal Survey of Adolescent Health (AddHealth). The AddHealth survey has been designed to study 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 the years 1994-1995 (Wave I). Every pupil attending the sampled schools on the interview day is asked to compile a questionnaire (in-school data) containing questions on respondents’ demographic and behavioral characteristics, education, family background and friendship. A subset of adolescents selected from the rosters of the sampled schools, about 20,000 individuals, is then asked to compile a longer questionnaire containing more sensitive individual and household information (in-home and parental data). Those subjects are interviewed again in 1995-1996 (Wave II), in 2001-2002 (Wave III), and in 2007-2008 (Wave IV).

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). This information is collected in Wave I and one year after, in Wave II. As a result, one can reconstruct the whole geometric structure of the friendship networks and their evolution, at least in the short run. Such detailed information on social interaction patterns allows us

---

8 The 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, both in Wave I and Wave II.
to measure the peer group more precisely than in previous studies by knowing exactly who
nominates whom in a network (i.e. who interacts with whom in a social group).

Moreover, and this has not been done before, one can distinguish between *long-lived* and
*short-lived* ties in the data. We define a *long-lived* friendship between two students if they
have nominated each other in both waves (i.e. in Wave I in 1994-1995 and in Wave II in
1995-1996) and a *short-lived* friendship if they have nominated each other in one wave only
(Wave I or Wave II). In Section 6.2.2 we check the robustness of our definition when links
may be erroneously observed as long or short-lived.\(^9\)

By matching the identification numbers of the friendship nominations to respondents’
identification numbers, one can also obtain information on the characteristics of nominated
friends. In addition, the longitudinal structure of the survey provides information on both
respondents and friends during adulthood. In particular, the questionnaire of Wave IV
contains detailed information on the highest education qualification achieved. We measure
*educational attainment* in completed years of full time education in Wave IV.\(^10\) Social con-
tacts (i.e. friendship nominations) are, instead, collected in Waves I and II.

Our final sample of in-home Wave I students (and friends) that are followed over time
and have non-missing information on our target variables both in Waves I, II and IV consists
of 1,819 individuals distributed over 116 networks. This large reduction in sample size with
respect to the original sample is mainly due to the network construction procedure - roughly
20 percent of the students do not nominate any friends and another 20 percent cannot be
correctly linked. In addition, we exclude networks consisting of 2-3 individuals, those with
more than 400 members and individuals who are not followed in Wave IV.\(^11\) In Wave I,
the mean and the standard deviation of network size are roughly 9.5 and 15, respectively.
Roughly 61\% of the nominations are not renewed in Wave II, and about 44\% new ones are
made. On average, these adolescents have roughly 30\% long-lived ties and 70\% short-lived
ties. Further details on nomination data can be found in Table A1 in Appendix A. Appendix

---

\(^9\)In principle, a short-lived tie observed only in Wave II can be a long-lived tie if it is not severed later,
while a tie observed in both Wave I and Wave II may be severed later, becoming a short-lived.

\(^10\)More precisely, the Wave IV questionnaire asks about the highest education qualification achieved (dis-
tinguishing between 8th grade or less, high school, vocational/technical training, bachelor’s degree, graduate
school, master’s degree, graduate training beyond a master’s degree, doctoral degree, post baccalaureate
professional education). Those with high school qualifications and higher are also asked to report the exact
year when the highest qualification was achieved. Such information allows us to construct a reliable measure
of each individual’s completed years of education.

\(^11\)We do not consider networks at the extremes of the network size distribution (i.e. consisting of 2-
3 individuals or more than 400) because peer effects can show extreme values (too high or too low) in
these edge networks (see Calvó-Armengol et al., 2009). The representativeness of the sample is preserved.
Summary statistics are available upon request.
A also gives a precise definition of the variables used in our study as well as their descriptive statistics (see Table A1). 12

3 Empirical model and identification strategy

3.1 Empirical model

Let \( \bar{r} \) be the total number of networks in the sample (i.e. \( \bar{r} = 116 \)), \( n_r \) 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. \( n = 1,819 \)). Let \( x_i, r = (x_{i,r}^1, \ldots, x_{i,r}^M)' \) denote the vector of individual observable characteristics of individual \( i \) belonging to network \( r \). Let us denote the adjacency matrix of the long-lived peers by \( G^L = \{g_{ij}^L \} \), where \( g_{ij}^L = 1 \) if \( i \) and \( j \) are long-lived friends (i.e. students \( i \) and \( j \) have nominated each other in Wave I and in Wave II). Similarly, let the adjacency matrix of the short-lived peers be \( G^S = \{g_{ij}^S \} \), where \( g_{ij}^S = 1 \) if \( i \) and \( j \) are short-lived friends (i.e. students \( i \) and \( j \) have nominated each other in one wave only). Our empirical model of agent \( i \) belonging to network \( r \) can then be written as: 13

\[
y_{i,r,t+1} = \phi^L \sum_{j=1}^{n_r} g_{ij}^L \gamma_{j,r,t+1} + \phi^S \sum_{j=1}^{n_r} g_{ij}^S \gamma_{j,r,t+1} + x_{i,r}' \delta + \frac{1}{g_{i,r,t}} \sum_{j=1}^{n_r} g_{ij}^L x_{i,r,t+1} + \frac{1}{g_{i,r,t}} \sum_{j=1}^{n_r} g_{ij}^S x_{j,r,t+1} + \eta_{r,t} + \epsilon_{i,r,t+1},
\]

where \( y_{i,r,t+1} \) is the highest education level reached by individual \( i \) at time \( t+1 \) who belonged to network \( r \) at time \( t \), where time \( t+1 \) refers to Wave IV in 2007-2008 while time \( t \) refers to Wave I in 1994-1995 and/or Wave II in 1995-1996 (depending on whether we consider short-lived or long-lived ties). Similarly, \( y_{j,r,t+1} \) is the highest education level reached by individual \( j \) at time \( t+1 \) who has been nominated as his/her friend by individual \( i \) at time \( t \) in network \( r \). Furthermore, \( x_{i,r,t+1} = (x_{i,r,t+1}^1, \ldots, x_{i,r,t+1}^M)' \) indicates the \( M \) variables accounting for observable differences in individual characteristics of individual \( i \) both at times \( t \) (e.g. self esteem, mathematics score, quality of the neighborhood, etc.) and \( t+1 \) (marital status, age, children, etc.) of individual \( i \). Some characteristics are clearly the same at times \( t \) and \( t+1 \), such as race, parents’ education, gender, etc. Also \( g_{i,r,t} = \sum_{j=1}^{n} g_{ij,r,t}^L \) and \( g_{i,r,t}^S = \sum_{j=1}^{n} g_{ij,r,t}^S \).

12 Information at the school level, such as school quality and the teacher/pupil ratio, is also available but we do not need to use it since our sample of networks is within schools and we use fixed network effects in our estimation strategy.

13 Model (1) stems from the first order conditions of the network game described in Appendix B.
are the total number of long-lived and short-lived friends each individual \( i \) has in network \( r \) at time \( t \). Finally, \( \epsilon_{i,r} \)'s are i.i.d. innovations with zero mean and variance \( \sigma^2 \) for all \( i \) and \( r \).

Let \( Y_r = (y_{1,r,t+1}, \ldots, y_{n_r,r,t+1})', X_r = (x_{1,r,t,t+1}, \ldots, x_{n_r,r,t,t+1})' \), and \( \epsilon_r = (\epsilon_{1,r}, \ldots, \epsilon_{n_r,r})' \). Denote the \( n_r \times n_r \) adjacency matrix by \( G_r = [g_{ij,r}] \), the row-normalized of \( G_r \) by \( G^*_r \), and the \( n_r \)-dimensional vector of ones by \( I_{n_r} \). As above, let us split the adjacency matrix into two submatrices \( G_r^L \) and \( G_r^S \), which keep trace of long-lived and short-lived friends, respectively.

Then, model (1) can be written in matrix form as:

\[
Y_r = \phi^L G_r^L Y_r + \phi^S G_r^S Y_r + X_r^* \beta + \eta_r I_{n_r} + \epsilon_r, \tag{2}
\]

where \( X_r^* = (X_r + G_r^s X_r + G_r^s I_{X_r}) \) and \( \beta = (\delta', \gamma^L, \gamma^S)' \).

For a sample with \( \bar{r} \) networks, stack up the data by defining \( Y = (Y_1', \ldots, Y_{\bar{r}}')', X^* = (X_1'^*, \ldots, X_{\bar{r}}'^*)', \epsilon = (\epsilon_1', \ldots, \epsilon_{\bar{r}}')', G = D(G_1, \ldots, G_{\bar{r}}), G^* = D(G_1^*, \ldots, G_{\bar{r}}^*), \iota = D(I_{n_1}, \ldots, I_{n_{\bar{r}}}) \) and \( \eta = (\eta_1, \ldots, \eta_{\bar{r}})' \), where \( D(A_1, \ldots, A_{\bar{r}}) \) is a block diagonal matrix in which the diagonal blocks are \( n_k \times n_k \) matrices \( A_k \)'s. For the entire sample, the model is thus:

\[
Y = \phi^L G^L Y + \phi^S G^S Y + X^* \beta + \iota \cdot \eta + \epsilon. \tag{3}
\]

In this model, \( \phi^L \) and \( \phi^S \) represent the endogenous effects, i.e. the agent’s outcome depends on that of his/her friends, while \( \gamma^L \) and \( \gamma^S \) represent the contextual effect, i.e. the agent’s choice/outcome depends on the exogenous characteristics of his/her friends. The vector of network fixed effects \( \eta \) captures the correlated effect where agents in the same network may behave similarly as they have similar unobserved individual characteristics or they face a similar (e.g. institutional) environment.\(^{14}\)

### 3.2 Identification and estimation

A number of papers have dealt with the identification and estimation of peer effects with network data (e.g. Bramoullé et al., 2009; Liu and Lee, 2010, Calvó-Armengol et al., 2009; Lin, 2010; Lee et al., 2010; Liu et al., 2012). Below, we review the crucial issues, while explaining how we tackle them.

**Reflection problem** In linear-in-means models, simultaneity in the behavior of interacting agents introduces a perfect collinearity between the expected mean outcome of the group and its mean characteristics. Therefore, it is difficult to differentiate between the effect of

\(^{14}\)As an analogy with time series models, the model in (3) can be referred to as a SARARMA(\( p, q \)) with \( p = 0 \) and \( q = 2 \), where \( p \) and \( q \) are the maximum number of spatial lags for the error and the outcome, respectively.
peers’ choice of effort (endogenous effects) and peers’ characteristics (contextual effects) that do have an impact on their effort choice (the so-called reflection problem; Manski, 1993). Basically, the reflection problem arises because, in the standard approach, individuals interact in groups, that is individuals are affected by all individuals belonging to their group and by nobody outside the group. In the case of social networks, instead, this is nearly never true since the reference group is individual specific. For example, take individuals $i$ and $k$ such that $g_{ik} = 1$. Then, individual $i$ is directly influenced by $g_i = \sum_{j=1}^{n_i} g_{ij} y_j$ while individual $k$ is directly influenced by $g_k = \sum_{j=1}^{n_k} g_{kj} y_j$, and there is little chance for these two values to be the same unless the network is complete (i.e. everybody is linked with everybody).\(^{15}\)

**Correlated effects** While a network approach allows us to distinguish endogenous effects from contextual effects, it does not necessarily estimate the causal effect of peers’ influence on individual behavior. The estimation results might be flawed because of the presence of peer-group specific unobservable factors affecting both individual and peer behavior. For example, a correlation between the individual and the peer-school performance may be due to an exposure to common factors (e.g. having good teachers) rather than to social interactions. The way in which this has been addressed in the literature is to exploit the architecture of network contacts to construct valid IVs for the endogenous effect. Since peer groups are individual specific in social networks, the characteristics of indirect friends are natural candidates. For example, consider a star network where individual $j$ is the star and is linked to individuals $i$ and $k$. In that case, individual $k$ affects the behavior of individual $i$ only through their common friend $j$, and she/he is not exposed to the factors affecting the peer group consisting of individual $i$ and individual $j$. As a result, the characteristics $x_k$ of individual $k$ are valid instruments for $y_j$, the endogenous outcome of $j$.

**Sorting** If the variables that drive the choice of peers are not fully observable, potential correlations between (unobserved) peer-group-specific factors and the target regressors are major sources of bias. We deal with this problems in two ways.

First, we follow the standard way out in the literature (e.g. Bramoullé et al., 2009) that consists in using network fixed effects. Network fixed effects are a remedy for the selection bias that originates from the possible sorting of individuals with similar unobserved characteristics into a network. The underlying assumption is that such unobserved characteristics are common to all the individuals within each network. This is reasonable in our case study where the networks are quite small (see Section 2).\(^{16}\) In our case, this assumption further

\(^{15}\)Formally, social effects are identified (i.e. no reflection problem) if $G^2 \neq 0$, where $G^2$ keeps track of indirect connections of length 2 in the network. This means that we need at least two individuals in the networks that have different links. This condition is generally satisfied in every real-world network.

\(^{16}\)93% of our networks have a size below 35.
implies that such unobserved characteristics are common to both weak and strong ties, which means that there should not be much difference between the friends who are long-lived and short-lived. We collect some evidence in line with this idea. Indeed, in Section 5.2 below (Tables 5 and 6), we provide evidence showing that there are no differences between peers in Waves I and II in terms of observable characteristics, that is that the link formation between these two waves is not significantly different. As a consequence, it seems reasonable to also assume that the influence of unobservable factors is the same for short and long-lived ties.

Second, as a robustness check, in Section 6.1 below, we will consider an explicit model of network formation and estimate simultaneously the outcome equation (1) and the bilateral choice of links. This approach allows for the presence of unobserved factors that vary by link-type, that is that are different for short and long-lived ties.

4 Estimation results

The aim of our empirical analysis is twofold, (i) to assess the presence of long-run peer effects in education and, (ii) to differentiate between the impact of short-lived and long-lived friends on education.

We consider 2SLS estimators (Liu and Lee, 2010) with network fixed effects and propose two innovations. First, we use two interaction matrices, one for long-lived ties and one for short-lived ties. Second, we take advantage of the longitudinal structure of our data and only include values lagged in time in the different instrumental matrices (i.e. observed in Wave I). Appendix C reviews the approach proposed by Liu and Lee (2010) and highlights the modification that is implemented in this paper.\(^{17}\)

4.1 Long-run peer effects

Table 1 collects the estimation results of model (1), without distinguishing between long-lived and short-lived ties. In other words, students \(i\) and \(j\) are friends, i.e. \(g_{ij} = 1\), if they have nominated each other in Wave I. We then look at the impact of friends from Wave I on own educational attainment in Wave IV. The first three columns show the results when using the

\(^{17}\)Observe that most of the traditional problems in the identification of peer effects arises when outcome and linking decisions are taken at the same time. In our analysis, we do not have this problem since there is a time lag between when friends are chosen (Waves I and II in 1994-1996) and when the outcome (education) is observed (Wave IV in 2007-2008). In addition, the longitudinal aspect of our analysis provides IVs at different points in time, i.e. the characteristics of indirect peers when they are at school and when they are adults. It is thus possible to use only variables lagged in time as instruments to ensure that the instruments are not correlated with the contemporaneous error term.
traditional set of instruments whereas, in the last three columns, the instrumental set only contains variables lagged in time (see Appendix C). The first-stage partial F-statistics (Stock et al., 2002 and Stock and Yogo, 2005) reveals that our instruments are quite informative and the OIR test provides evidence in line with their validity. Table 2 shows the results of the first stage.

[Insert Table 1 and Table 2 here]

The results in Table 1 do not change to any considerable extent across columns and reveal that the effect of friends’ education on own education is always significant and positive, i.e., there are long-lived and persistent peer effects in education. This shows that the “quality” of friends (in terms of future educational achievement) from high school has a positive and significant impact on the own future educational level, even though it might be that individuals who were close friends in 1994-1995 (Wave I) might no longer be friends in 2007-2008 (Wave IV). According to the bias-corrected 2SLS estimator,\(^\text{18}\) in a group of two friends, a standard deviation increase in the years of education of the friend translates into a roughly 5.4 percent increase of a standard deviation in the individual years of education (roughly two more months of education). If we consider an average group of four best friends (linked to each other in a network), a standard deviation increase in the level of education of each of the peers translates into a roughly 16 percent increase of a standard deviation in the individual’s educational attainment (roughly seven more months of education). This is a non-negligible effect, especially given our long list of controls and the fact that friendship networks might have changed over time. The influence of peers at school seems to be carried over time.

4.2 The role of long-lived ties

We would now like to determine how long- and short-lived ties affect educational choices by estimating the magnitude of \(\phi^L\) and \(\phi^S\) in equation (1). Table 3 shows the estimation results of equation (1).\(^\text{19}\) We find that the educational choices of short-lived friends have no significant impact on individual educational outcomes (years of schooling) while the educational choices of long-lived friends do have a positive and significant effect on own educational outcome. In terms of magnitude, a standard deviation increase in aggregate years of education of peers nominated both in Waves I and II (long-lived friends) translates

\(^{18}\)The bias-corrected 2SLS estimator is our preferred one since we have relatively small networks (see Appendix C).

\(^{19}\)We show the results for the bias-corrected 2SLS estimator, with the traditional set of instruments and when the instrumental set only contains variables lagged in time. The qualitative results when using the alternative estimators in Appendix C remain qualitatively unchanged. The latter are available upon request.
into roughly a 21 percent increase of a standard deviation in the individual’s educational attainment (roughly 8.3 more months of education). In an average group of four best friends (linked to each other in a network), a standard deviation increase of each of the peers translates into two more years of education. This is quite an important effect. It suggests that long-lived friends rather than short-lived friends matter for educational outcomes in the long run.\textsuperscript{20}

[Insert Table 3 here]

5 Understanding the mechanisms

Our empirical results displayed in Table 3 suggest that the distinction between long-lived and short-lived friends is important for understanding long-run peer effects in education. In Section 5.1, we propose a simple theoretical model that may explain this evidence. The idea underlying the theoretical mechanism is that convergence of preferences and formation of social norms need long-term relationships between peers. In Section 5.2, we provide additional empirical evidence ruling out alternative explanations.

5.1 Theoretical framework

In order to understand how long-lived and short-lived ties influence long-run educational outcomes, we extend the DeGroot (1974) model as follows.\textsuperscript{21} Consider a society consisting of a finite set of individuals $N = \{1, 2, \ldots, n\}$ who are linked in a directed network and who would like to gather information about an unknown parameter $\theta$. In our context, assume that there are two states of the world so that $\theta$ can be equal to either: \{It is worth continuing studying\} or \{It is not worth continuing studying\}. What is key in the DeGroot model is that agents update their beliefs by repeatedly taking \textit{weighted averages} of their neighbors’ beliefs (where neighbors are the people directly linked to each individual) with $p_{ij}$ being the weight that agent $i$ places on the current belief of agent $j$ in forming his or her belief for the next period. If the network is strongly connected and at least some individuals listen to themselves, then, in the limit, everybody belonging to the same network will converge to a consensus and the influence of each person will depend on their position in the network. This means that, if there are \textit{repeated interactions} between students from the same network, then they will all adopt the same social norm, which could be either \{It is worth continuing

\textsuperscript{20}When estimating equation (1) including only long-lived ties (i.e. $G^S = 0$), we obtain comparable results.

\textsuperscript{21}Appendix D contains the technical details of the DeGroot model.
studying} or \{It is not worth continuing studying\} depending on what the “influential” students believe about the social norm.\(^{22}\)

We use this theoretical framework to understand why, in our empirical results, short-lived friends have no impact on own education decision while long-lived friends have. We define short-lived friends as students who interact with each other only once while long-lived friends where students who interact for a longer time. There are two types of friend relationships between students in a network \(G\): short-lived friendships \((l = S)\) and long-lived friendships \((l = L)\). Quite naturally, we assume that each agent has a long-lived relationship with him/herself, i.e. \(g_{ii}^{L} = 1\) and \(g_{ii}^{S} = 0\). As in Section 3, this implies two types of adjacency matrices: \(G^{L} = \{g_{ij}^{L}\}\), where \(g_{ij}^{L} = 1\) if \(i\) and \(j\) are long-lived friends and \(G^{S} = \{g_{ij}^{S}\}\), where \(g_{ij}^{S} = 1\) if \(i\) and \(j\) are short-lived friends, with \(G^{L} + G^{S} = G\). Denote by \(\tilde{G}^{L}\) and \(\tilde{G}^{S}\) the row-normalized matrices of \(G^{L}\) and \(G^{S}\), respectively. Because short-lived friendships only interact once, they will influence the beliefs of each other only in the initial period. The updating stops there since short-lived friends do not meet anymore and thus students only update their beliefs once. On the contrary, long-lived friends interact repeatedly and thus update their beliefs all the time as in the standard DeGroot model. Then all students in the network will therefore reach a consensus after many interactions, even though only long-lived friends will matter in the long run.

How do we solve this model? In the first period, both short-lived and long-lived students influence each other so that

\[
\mathbf{b}^{(1)} = \tilde{G} \mathbf{b}^{(0)}
\]

where \(\mathbf{b}^{(t)}\) is the vector of beliefs of all students at time \(t\) (i.e. both short-lived and long-lived students) and \(\tilde{G}\) is the row-normalized matrix of \(G\). Now relabel \(\mathbf{b}^{(1)}\) as \(\tilde{\mathbf{b}}^{(0)}\), i.e. \(\tilde{\mathbf{b}}^{(0)} := \mathbf{b}^{(1)}\). We are now in the framework of the DeGroot model where the initial beliefs are given by \(\tilde{\mathbf{b}}^{(0)}\). As a result, we can apply Proposition 8 in Appendix D. If the network \(\tilde{G}^{L}\) is strongly connected and if at least some agents pay attention to themselves, i.e. \(g_{ii} > 0\) for some \(i\), then all students will reach a consensus in the long run, which is determined by:

\[
(\mathbf{b})^\infty = \lim_{t \to \infty} \left(\tilde{G}^{L}\right)^{t} \tilde{\mathbf{b}}^{(0)} = \lim_{t \to \infty} \left(\tilde{G}^{L}\right)^{t} \tilde{G} \mathbf{b}^{(0)}
\]

(4)

In equation (4), we see clearly the distinct influence of short-lived and long-lived friends.\(^{22}\)

\(^{22}\)Because students have parents with different incomes or students have different costs of studying (some like to study while others don’t), we can explain why, within a network with the same social norm, students take different decisions concerning the number of years they will spend in college. In particular, the students whose parents have low income or the students who have a high disutility of studying will end up not going to college even if the social norm in their network of friends says that it is worth continuing studying.

13
The updating matrix $\tilde{G}^L$ is only depending on long-lived friends because they are the ones who interact over time and help reach a consensus. However, the initial beliefs are a function of the beliefs of both short-lived and long-lived friends since $\tilde{G}$ includes both $G^L$ and $G^S$. In equation (4), we assume that all students have both long-lived and short-lived friends\textsuperscript{23} so that they are all included in the convergence process $(b)\infty$.

To illustrate this result, consider the example given at the end of Appendix D where the network is displayed in Figure A1. The adjacency matrix $G$ and the row-normalized one $\tilde{G}$ are given by:

$$G = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

and

$$\tilde{G} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}$$

Assume that both agents 1 and 2 and agents 1 and 3 have a long-lived friendship while agents agents 2 and 3 have a short-lived friendship. As stated above, we assume that each agent has a long-lived relationship with him/herself, i.e. $g^L_{ii} = 1$ and $g^S_{ii} = 0$. We have:

$$G^L = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

and

$$G^S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

so that $G^L + G^S = G$

Let us row-normalize these matrices so that

$$\tilde{G}^L = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

and

$$\tilde{G}^S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Observe that $\tilde{G}^L$ has been chosen so that it is equal to $P$ in the example in Appendix D where we don’t differentiate between short-lived and long-lived friends. As in the example in Appendix D, assume that the initial beliefs are given by: $b^{(0)} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$.

\textsuperscript{23}Observe that if some students have only short-lived friends, then there will be no consensus. Indeed, even if just one student $i$ has only short-lived friends, then $g^L_{ii} = 1$ and $g^S_{ij} = 0$ for all other $j$. Hence network $\tilde{G}^L$ will not be strongly connected, and the baseline deGroot model will not work properly (i.e. depending on the network structure there will either be more than one component with a different “consensus” each, or no consensus at all).
Let us first determine the initial beliefs. We have:

$$\tilde{b}^{(0)} := b^{(1)} = G b^{(0)} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/2 \\ 0 \end{pmatrix}$$

Now we can determine the consensus among all the students where the updates is only on the matrix $G^L$ for the long-lived students. It is easily shown that

$$\left(G^L \right)^t = \begin{pmatrix} 3/7 & 2/7 & 2/7 \\ 3/7 & 2/7 & 2/7 \\ 3/7 & 2/7 & 2/7 \end{pmatrix}$$

so that there is convergence to the following consensus:

$$(b)^\infty = \lim_{t \to \infty} \left(G^L \right)^t \tilde{b}^{(0)} = \begin{pmatrix} 3/7 \\ 2/7 \\ 2/7 \end{pmatrix}$$

In other words, no matter what beliefs $\tilde{b}^{(0)}$ the agents start with, they all end up with limiting beliefs corresponding to the entries of $(b)^\infty = \lim_{t \to \infty} \left(G^L \right)^t \tilde{b}^{(0)}$ where

$$(b_1)^\infty = (b_2)^\infty = (b_3)^\infty = \frac{3}{7} b_1^{(0)} + \frac{2}{7} b_2^{(0)} + \frac{2}{7} b_3^{(0)}$$

From these limiting beliefs, using the initial beliefs $\tilde{b}^{(0)}$, we can then calculate the consensus reached by all agents in the network. We have:

$$\frac{3}{7} \tilde{b}_1^{(0)} + \frac{2}{7} \tilde{b}_2^{(0)} + \frac{2}{7} \tilde{b}_3^{(0)} = \frac{3}{7} + \frac{2}{7} + \frac{2}{7} = 0.286$$

If the consensus is on the state \{It is worth continuing studying\}, this means that the three students reach a consensus for which there will agree that it is worth continuing studying with probability 0.286. This example shows that the beliefs converge over time for all the students and that they reach a consensus but it also shows that agent 1 has more influence than agents 2 and 3 over the limiting beliefs. Observe that, compared to the example in Appendix D where we don’t differentiate between short-lived and long-lived friends, the consensus of continuing studying here leads to a lower probability since $0.286 < 0.333$, even though we update on the same matrix $G^L = P$. This is because of the influence of the
short-lived friends who change the initial beliefs from \( b^{(0)} \) to \( \tilde{b}^{(0)} = G b^{(0)} \).

As a result, a possible interpretation of our evidence is that the *strength of interactions* between two students may affect how much they learn, the human capital accumulation and how much they value achievement. It also shapes social norms that accumulate over time, which affect years of schooling both directly and indirectly. This idea is related to Akerlof’s and Kranton’s (2002) concept of identity in economics, where learning at school can be viewed within a process of identity formation, resource allocation, and social interaction. In other words, following the sociology literature, Akerlof and Kranton (2002) postulate that students often care less about their studies than about what their friends think.\(^{24}\)

Observe that the aim of the model is to give some economic intuition of why different types of links (and thus friends) have different impacts on long-run outcomes (which is the steady-state). We are not directly testing this model in the data, i.e. equation (1) is not a reduced form of the model presented in this section. For example, in our data, short-lived links live only one period while long-lived links live two periods. In our model, long-lived links live an infinite number of periods while short-lived links live one period.\(^{25}\)

Given that there are no datasets with infinite (or very high number of) network observations for students, we can make some inference using our data. The observed long-lived ties have a higher probability to be the real long-lived ties, because they are observed more times than the others. Our sensitivity analysis in Section 6.2.2 explores this aspect, changing the link statuses and checking whether the main results still hold under link length misspecification.

### 5.2 Additional evidence

Our analysis so far suggests that the distinction between long-lived and short-lived ties is important for understanding long-run peer effects in education. The mechanism for the social effects is based on the idea that the convergence of preferences and the emergence of social norms among peers need long-term interactions. In this section, we aim at ruling out alternative explanations.

\(^{24}\)This is also related to the empirical study of De Giorgi et al. (2010) which shows that students from Bocconi University in Italy are more likely to choose a major if many of their peers make the same choice. They also show that peers can divert students from majors in which they have a relative ability advantage, with adverse consequences on academic performance.

\(^{25}\)One way to make the model closer to the data is to assume that long-lived friends (i.e. those nominated in Waves I and II) are still friends in Wave IV while short-lived ones are not. This seems quite reasonable and would then imply that short-lived links live only one period while long-lived links live an infinite number of periods.
In our analysis, we identified long-lived ties as peers nominated in both Wave I and Wave II. This definition implies that long-lived ties are friends who are more likely to be peers at the time of college decisions. One could thus put forward another explanation of why friends at school may influence education decisions: it could be the timing of friendship or decision proximity so that friends in the last grades (grades 10 to 12) are likely to have an impact on college decision, regardless of whether these are long-lived or short-lived ties. In other words, is it really the frequency and strength of social interactions or is it the timing of friendship formation that is crucial for future educational outcomes? We would therefore like to disentangle between the decision proximity effect and the strength of interaction effect. For this purpose, we select students in the last grades (grades 10 to 12) and distinguish between short-lived and long-lived ties and examine the effect on college choices. We estimate a modified version of model (1), that is

\[
y_{i,r,t+1} = \phi^L \sum_{j=1}^{n_r} g^{L}_{i,j,r,t} y_{j,r,t+1} + \phi^{S_1} \sum_{j=1}^{n_r} g^{S_1}_{i,j,r,t} y_{j,r,t+1} + \phi^{S_2} \sum_{j=1}^{n_r} g^{S_2}_{i,j,r,t} y_{j,r,t+1} + \frac{1}{g^{L}_{i,r,t}} \sum_{j=1}^{n_r} g^{L}_{i,j,r,t} \gamma_{j,r,t+1}^{L} \\
+ \frac{1}{g^{S_1}_{i,r,t}} \sum_{j=1}^{n_r} g^{S_1}_{i,j,r,t} \gamma_{j,r,t+1}^{S_1} + \frac{1}{g^{S_2}_{i,r,t}} \sum_{j=1}^{n_r} g^{S_2}_{i,j,r,t} \gamma_{j,r,t+1}^{S_2} + \eta_{r,t} + \epsilon_{i,r,t+1}.
\]

We here distinguish between short-lived ties where best friends have only been nominated in Wave I (lower grades) and not in Wave II (later grades), i.e. \( \phi^S = \phi^{S_1} \), from short-lived ties where best friends have only been nominated in Wave II and not in Wave I, i.e. \( \phi^S = \phi^{S_2} \). If the decision proximity matters, then coefficient \( \phi^{S_2} \) should be significant while \( \phi^{S_1} \) should not.

Table 4 contains the estimation results. The empirical results reveal that the education decision of short-lived ties continues to show a non-significant effect on individual education outcomes, regardless of whether peers are interacting in lower or higher grades, highlighting the crucial role of long-lived ties in college decision.

Table 4

Another concern is that peers nominated in different time periods may have a different long-run effect because students value peer characteristics differently in friendship decisions made over time. Do students select peers differently between the first and the second wave or is it really that distinct types of peers (short-lived versus long-lived ties) are of different importance? To disentangle these effects, we check whether students select peers differently between the first and the second wave. Table 5 compares the observable characteristics of
peers who only appear in Wave I, those who only appear in Wave II, and those who appear in both waves. One can see that, in fact, there are no differences between these peers in terms of observable characteristics.

To further investigate this issue, we test whether link formation differs between different waves. Let us consider a standard network formation model where the variables that explain friendship formation between students $i$ and $j$ belonging to network $r$ are the distances between them in terms of observed characteristics (see e.g. Currarini et al., 2009, 2010), and pool the data for Wave I ($\tau = 1$) and Wave II ($\tau = 2$).

$$g_{ij,r,t} = \alpha + \sum_{m=1}^{M} \beta_m|x_{i,r,t}^m - x_{j,r,t}^m| + \sum_{m=1}^{M} \gamma^m|x_{i,r,t}^m - x_{j,r,t}^m| \times d_{ij,r} + \epsilon_{ij,r,t}, \quad t = 1, 2.$$ (5)

In this model, $g_{ij,r,t} = 1$ if there is a link between $i$ and $j$ belonging to network $r$ at time $t$ (where $t =$ Wave I, Wave II), $x_{i,r,t}^m$ indicates the individual characteristic $m$ of individual $i$ in network $r$ at time $t$ and $d_{ij,r}$ is a dummy variable, which is equal to 1 if a link $g_{ij,t}$ exists in Wave II, and zero otherwise. The parameter in front of the dummy variable captures the differences between the importance of these characteristics in link formation between Wave I and Wave II. Estimating equation (5), Table 6 shows that most coefficients are not significant and that there are no observable differences in the link formation process between Waves I and II. We have also performed an F test that tests the joint significance of the $\gamma$ parameters.\footnote{The idea is similar to the Chow test in time series analysis to investigate the existence of a structural break (see e.g. Chow, 1960; Hansen, 2000, 2001).} Table 6 reports the $p$ value of this test. It reveals that, controlling for network fixed effects, we cannot reject the null hypothesis of $\gamma^m = 0, \forall m = 1, \ldots, M$. In summary, Tables 5 and 6 provide evidence showing that there are no differences between peers in Waves I and II in terms of observable characteristics and that the link formation between the different waves is not different.
agents (Wasserman and Faust, 1994). We present those indicators in Appendix E where we define the density and the assortativity of a network and, at the node and network level, the betweenness centrality, the closeness centrality and the clustering coefficient. When applied to our Wave I and Wave II networks, we obtain the results collected in Table 7. It appears that the two networks are topologically very similar.

[Insert Table 7 here]

6 Robustness checks

In this section, we check the robustness of our results with respect to two different issues: (i) the presence of unobserved factors different from network fixed effects (Section 6.1); (ii) mispecification of the network structure (Section 6.2).

6.1 Endogenous network formation

Our identification strategy hinges on the use of network fixed effects to control for unobserved factors driving both network formation and behavior on networks. If there are student-level unobservables that drive both peer choice and outcome choice, this strategy fails. A possible way to tackle this issue is to simultaneously estimate network formation and outcomes. This strategy can be pursued by using parametric modeling assumptions and Bayesian inferential methods that allow to integrate a network formation with the study of behavior over the formed networks. Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2015) propose two slightly different ways to implement this approach. In Goldsmith-Pinkham and Imbens (2013) unobservables are dichotomous and only one network is considered. As we have multiple networks in our data, we follow Hsieh and Lee (2015). They present a model with one peer type. We implement an extension of their method for heterogeneous peer effects. If there is an unobservable characteristic that drives the choice of, say, long-lived ties and is correlated with \( \epsilon_{i,r} \) then \( g_{ij}^r \) is endogenous and estimates of Model (1) are biased. By failing to account for similarities in (unobserved) characteristics, similar behaviors might mistakenly be attributed to peer influence when they simply result from similar characteristics. Let \( z_{i,r} \) denote such an unobserved characteristic which influence the link formation process. Let us also assume that \( z_{i,r} \) is correlated with \( \epsilon_{i,r} \) in Model (1) according to a bivariate normal

---

27 Another difference between those two procedures is that Goldsmith-Pinkham and Imbens (2013) set the same unobservable in both link formation and outcome equation while Hsieh and Lee (2015) use different unobservables for those equations and let them to be correlated.
distribution
\[(z_{i,r}, \epsilon_{i,r}) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_x & \sigma_{xz} \\ \sigma_{xz} & \sigma^2_z \end{pmatrix}\right)\].

Agents choose social contacts at two points in time, \(t-1\) and \(t\). At each time, agent \(i\) chooses to be friends with \(j\) according to a vector of observed and unobserved characteristics in a standard link formation probabilistic model (as in model (5))

\[P(g_{ij,r,t-1} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}) = \Lambda(\gamma_{0,t-1} + \sum_k |x_{i,r} - x_{j,r}| \gamma_{k,t-1} + |z_{i,r} - z_{j,r}| \theta_{t-1}), \quad (6)\]

and

\[P(g_{ij,r,t} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, g_{ij,r,t-1}, \gamma_{t}, \theta_{t}, \lambda) = \Lambda(\gamma_{0,t} + \lambda g_{ij,r,t-1} + \sum_k |x_{i,r} - x_{j,r}| \gamma_{k,t} + |z_{i,r} - z_{j,r}| \theta_{t}), \quad (7)\]

where \(\Lambda(\cdot)\) is a logistic function. Homophily behavior in the unobserved characteristics implies that \(\tau < 0\), where \(\tau = t - 1\), this meaning that the closer two individuals are in terms of unobservable characteristics, the higher is the probability that they are friends. The same argument holds for observables. If \(\sigma_{xz}\) and \(\theta_{r}\) are different from zero, then networks \(g_{ij,r}^{L}\) and \(g_{ij,r}^{S}\) in model (1) are endogenous. From Model (6) - (7), the probability of observing a short-lived tie is

\[P(g_{ij,r}^{S} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t}, \theta_{t}, \lambda, \gamma_{t-1}, \theta_{t-1}) = P(g_{ij,r,t} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t}, \theta_{t}, \lambda, g_{ij,r,t-1} = 0) \times P(g_{ij,r,t-1} = 0|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}) + P(g_{ij,r,t} = 0|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t}, \theta_{t}, \lambda, g_{ij,r,t-1} = 1) \times P(g_{ij,r,t-1} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1})\]

whereas the probability of observing a long-lived tie is

\[P(g_{ij,r}^{L} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t}, \theta_{t}, \lambda, \gamma_{t-1}, \theta_{t-1}) = P(g_{ij,r,t} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t}, \theta_{t}, \lambda, g_{ij,r,t-1} = 1) \times P(g_{ij,r,t-1} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}).\]

In this way, we have modeled the probability of being a long-lived or short-lived tie including unobservables that are allowed to be correlated with the error term in the outcome equation.28

Joint normality implies \(E(\epsilon_{i,r}|z_{i,r}) = \frac{\sigma_{xz}}{\sigma^2_z}z_{i,r}\), when conditioning on \(z_{i,r}\). Hence, the outcome equation is

\[28\text{The procedure can be easily extended to include more than one unobservable factor.}\]

20
where $\eta_l, r \sim N(0, \sigma^2 - \frac{\sigma^2}{\sigma^2_z})$. Note that if no correlation is at work ($\sigma_{zz} = 0$), then estimating equation (8) or (1) is equivalent. Given the complexity of this framework, it is convenient to simultaneously estimate the parameters of equations (6), (7) and (8) with a Bayesian approach. Bayesian inference requires the computation of marginal distribution for all parameters. However, since this requires 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 unobservable variable ($z_{i,r}$) is thus generated according to the joint likelihood of link formation and outcome; it is drawn in each MCMC step together with the parameters of the model. The Gibbs sampling algorithm allows us to draw random values for each parameter from their posterior marginal distribution, given previous values of other parameters. Once stationarity of the Markov Chain has been achieved, the random draws can be used to study the empirical distributions of the posterior.29

Table 8 panel (b) collects the results that are obtained when estimating simultaneously equations (8) and (6) - (7). Panel (a) shows the estimation result of the model, without distinguishing between short-lived and long lived ties (homogeneous peer effects). The first column in both panels reports the 2SLS results for comparison. Table 8 reveals that $\sigma_{zz}$ is not significantly different from zero for both the models with homogeneous and heterogeneous peer effects (columns (3) and (6) of Table 8). The Bayesian estimates are close to the 2SLS estimates.30 This evidence is thus in support of our identification strategy. Indeed, our list of controls and network fixed effects, together with the temporal lag between when friends are chosen and when education levels are attained, seem to account for unobserved factors driving both network formation and behavior over networks.

[Insert Table 8 here]
6.2 Measurement errors in network links

In this section, we perform two different robustness checks.

6.2.1 Directed networks

First, our empirical investigation has assumed that friendship relationships are symmetric, i.e. $g_{ij} = g_{ji}$. We check here how sensitive our results are to such an assumption, i.e. to a possible measurement error in the definition of the peer group. Indeed, our data make it possible to know exactly who nominates whom in a network and we find that 12 percent of the relationships in our dataset are not reciprocal. Instead of constructing an undirected network, in this section, we perform our analysis using directed networks. We focus on the choices made (outdegrees) and we denote a link from $i$ to $j$ as $g_{ij,r} = 1$ if $i$ has nominated $j$ as his/her friend in network $r$, and $g_{ij,r} = 0$, otherwise.\(^\text{31}\) Table 9 shows the estimation results of model (1) for directed networks. The results remain qualitatively unchanged and only slightly higher in magnitude.

\[\text{[Insert Table 9 here]}\]

6.2.2 A simulation experiment

Second, our identification and estimation strategies depend on the correct identification of long-lived and short-lived ties. In this section, we test the robustness of our results with respect to misspecification of long-lived and short-lived network topologies. Indeed, in our theoretical model, we assume that $\phi^L > \phi^S$ and our empirical analysis confirms this assumption by finding a significant effect of long-lived ties (but not short-lived ties) on educational outcomes. These results clearly depend on the definition of a long-lived and a short-lived tie. In the present robustness check, we want to check whether our results are robust even if we fail in exactly identifying long-lived and short-lived ties. To be more precise, we use simulated data to answer such questions as: Do our results change if some links are not assigned in the right category (short-lived or long-lived ties)? Do our results change if some links are not reported? To what extent? How many ties need to be misspecified before our results disappear?

In our analysis, we have defined a long-lived tie as a friend nominated twice, and a short-lived tie as a friend nominated just once. We can imagine that a student may be more likely

\(^{31}\)As highlighted by Wasserman and Faust (1994), centrality indices for directional relationships generally focus on choices made (outdegrees). The estimation results, however, remain qualitatively unchanged if we define the link using the nominations received (indegrees).
to report a long-lived tie than a short-lived tie. Let us suppose that individual $i$ reports a long-lived tie ($l = L$) with probability $p$, a short-lived tie ($l = S$) with probability $q$, with $q < p$, and another individual (neither a short- nor a long-lived tie, $l = N$) with probability $r$, with $r < q < p$.

This probabilistic scheme translates into the following transition table between observed and true types:

<table>
<thead>
<tr>
<th>Observed</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

For example, a long-lived tie appears as a long-lived tie with probability $p^2$, as a short-lived tie with probability $2p(1 - p)$, and may be missed with probability $(1 - p)^2$. In the table, $s = p^2 + q^2 + r^2$ denotes the probability of observing a long-lived tie, $w = 2p(1 - p) + 2q(1 - q) + 2r(1 - r)$ denotes the probability of observing a short-lived tie and $n = (1 - p)^2 + (1 - q)^2 + (1 - r)^2$ is the probability of not observing a tie.

Our empirical analysis assumes $s = p^2$, $w = 2q(1 - q)$, $n = (1 - r)^2$ and that the off diagonal elements are equal to zero. A misspecification of the network topology implies that the off diagonal elements are different from zero. Let us denote these off diagonal elements as $P_{LM}$, which are the probabilities of moving from state $L$ to state $M$; $L, M = \{S, L\}$. In our numerical exercise, we gradually change those elements from 0 to 1 at a pace of 0.005, i.e. $P_{LM} = [0, 0.005, 0.010, ..., 1]$.

Our misspecification experiment can be summarized by the following table:

<table>
<thead>
<tr>
<th>Observed</th>
<th>True</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
For ease of computation, we proceed in two steps. First, we change ties from long-lived to short-lived and vice versa, i.e. we change 
\[ P_{LS} = \frac{2p(1-p)}{p^2+2p(1-p)} \quad \text{and} \quad P_{SL} = \frac{q^2}{q^2+2q(1-q)}. \]
Second, for each combination of \( P_{LS} \) and \( P_{SL} \), we change ties to non-ties and vice versa, i.e. we change 
\[ P_{LN} = \frac{(1-p)^2}{n^2+2n(1-n)+(1-n)^2}, \quad P_{NL} = \frac{n^2}{n^2+2n(1-n)+(1-n)^2}, \quad P_{SN} = \frac{(1-q)^2}{q^2+2q(1-q)+(1-q)^2} \quad \text{and} \quad P_{NS} = \frac{(1-n)^2}{n^2+2n(1-n)+(1-n)^2}. \]

In this framework, the higher are these probabilities, the further away we are from our observed network topology. For example, the combination \( P_{LS} = 0.1 \) and \( P_{LN} = 0 \) means that 10% of the long-lived ties are replaced by short-lived ties; the combination \( P_{LS} = 0.3 \) and \( P_{LN} = 0.2 \) means that 30% of the long-lived ties are replaced by short-lived ties and 20% of the short-lived ties are replaced by unconnected individuals. In other words, our experiment does not only allow for the fact that long-lived and short-lived ties are not equally likely to be interchanged, but also considers the possibility that they each have some probability of generating a misreport that violates the exclusion restrictions. For each combination of \( P_{LS} \), \( P_{SL} \), \( P_{LN} \), \( P_{SN} \) and \( P_{SW} \), we draw one hundred network structures (samples) of a size equal to the real one \((n = 1,819)\). Then, we estimate model (3) replacing the real \( G_r^L \) and \( G_r^S \) matrices with the simulated ones in turn so that, in total, we estimate model (3) eighty thousand times for each type of estimator described in Appendix C.

Note that this exercise is quite similar to directly changing \( p, q \) and \( r \). The advantage of our approach is that it does not need to specify \( p, q \) and \( r \). Indeed, \( p, q \) and \( r \) are not known by the econometrician. They can be estimated imposing that observed and true numerosity are the same for each type of tie, but there is not any clear theoretical reason why this should be the case. An exploration of the entire space spanned by \((p, q, r)\) would imply a change in the observed (or true) network density which, in turn, would render our peer effect estimates non-comparable among combinations.

Simulated evidence

Figure 1 shows the results of our simulation experiment for the 2SLS bias-corrected lagged estimator.\(^{32}\) It depicts the estimates of long-lived and short-lived tie effects with 90% confidence bands, in the upper and lower panel, respectively.\(^{33}\)

\(^{32}\)The simulation results for the other estimators are similar. They are available upon request.

\(^{33}\)Standard errors have been calculated assuming drawing independence and taking into account the variation between estimates for each replacement rate. Specifically, the standard error at each replacement rate, say \( i \), is computed as follows:

\[ \sigma_i = \sqrt{W_i + B_i} \]

where \( W_i = \frac{1}{n} \sum_{j=1}^{n} \sigma_{ij}^2 \), \( B_i = \frac{1}{n} \sum_{j=1}^{n} (\hat{\phi}_{ij} - \bar{\phi}_i)^2 \), \( \sigma_{ij}^2 \) is the estimated variance of the \( j \)th estimator at the \( i \)th replacement rate, \( \hat{\phi}_{ij} \) is the \( j \)th estimate at the \( i \)th replacement rate and \( \bar{\phi}_i \) is the mean across the \( n \) estimates. In this experiment, \( n = 100 \).
The first important question concerns the percentage of network-structure misspecifications needed for the long-lived tie effects on college choice to disappear. The upper panel of Figure 1 shows the estimates for each combination of replacement rates – between long-lived and short-lived ties (\(P_{LS}\)) and between long-lived ties and no ties (\(P_{LN}\)). The graph shows that long-lived tie effects remain statistically significant for levels of \(P_{LS}\) and \(P_{LN}\) in the range of 0.005 and 0.35. Figure 2 depicts the conditional results (i.e. \(P_{LS}\) conditional on \(P_{LN} = 0\) in the upper panel and \(P_{LN}\) conditional on \(P_{LS} = 0\) in the lower panel). The upper panel shows that long-lived-tie effects remain statistically significant up to a percentage of randomly replaced links with short-lived ties of about 35%. The lower panel shows a similar result when increasing the percentage of links randomly replaced by zeros. This evidence implies that even if we do not observe or we imprecisely observe a portion of each individual’s long-lived ties, our results on the existence of this effect still hold.

The second question is what is the percentages of replacement needed in order to have a significant effect of short-lived ties. The lower panel of Figure 1 shows the estimates of short-lived tie effects for each combination of replacement rates – between short-lived and long-lived ties (\(P_{SL}\)) and between short-lived ties and no ties (\(P_{SN}\)). The graph shows that we need to replace almost 70% of the short-lived ties with long-lived ties before finding an effect which is statistically different from 0. Naturally, when replacing short-lived ties with no ties, we continue to detect no effect and the standard error increases with the percentage of replaced links. The lower panel of Figure 2 shows this evidence more clearly by depicting the conditional results (i.e. \(P_{SL}\) conditional on \(P_{SN} = 0\) in the upper panel and \(P_{SN}\) conditional on \(P_{SL} = 0\) in the lower panel). These results show that the effects of short-lived ties are found to be important only when the large majority of long-lived ties is labeled as short-lived ties.

Finally, we show in Figure 3 the rejection rates\(^{34}\) when using the \(2SLS\) bias-corrected estimator and the \(2SLS\) bias-corrected lagged estimator. This graph indicates that the \(2SLS\) bias-corrected lagged estimator tends to be more robust to a possible misspecification of long-lived and short-lived ties. Indeed, it appears that this estimator needs, on average, a

\(^{34}\)Rejection refers to the null hypothesis of having \(\phi^L = 0\) or \(\phi^S = 0\), respectively, for long-lived and short-lived ties effects. Each rate represents the frequency of rejection for the corresponding percentage of randomly replaced links.
higher percentage of misspecified ties to accept the hypothesis of no effects for long-lived ties and to reject it for short-lived ties.

Insert Figure 3 here

To wrap up, in this section, we have shown that the strength of long-lived ties $\phi^L$ is reduced and becomes insignificant when we have converted more than 35% of the long-lived ties into short-lived ties while the strength of short-lived ties $\phi^S$ is increasing and becomes significant after having replaced more than 60% of the short-lived ties with long-lived ties. To illustrate this result, consider a student $i$ who has twenty friends, ten long-lived ties \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} and ten short-lived ties \{11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}. Even if we incorrectly assign three friends from one category (long-lived tie) to the other (short-lived tie), our results will still hold. For instance, if we instead observe \{11, 12, 3, 4, 5, 6, 7, 8, 9, 10\} as long-lived ties (labeling 11 and 12 as long-lived when they are short-lived ties) and \{1, 2, 13, 14, 15, 16, 17, 18, 19, 20\} as short-lived ties (labeling 1 and 2 as short-lived when they are long-lived ties), we would still have a significant effect of long-lived ties on education and a non-significant effect of short-lived ties since we have “only” converted 30% of the links. As a result, from 3 to 8 incorrect assignments (which correspond to 30% to 80% conversion of long-lived ties into short-lived ties or the contrary), both effects will still be insignificant. It is only after having converted seven out of ten ties (i.e. more than 60% of the long-lived ties have been converted into short-lived ties, or the contrary) that we find that short-lived ties have a significant effect on education while long-lived ties do not.

7 Short-run versus long-run effects

So far, we have found that students nominate other students as their best friends but only their long-lived ties (i.e. students who are friends in both waves) influence them in their educational choices. Using Addhealth data for Wave I only, Calvó-Armengol et al. (2009) have studied the current effect of peers on education, finding that peers do affect the current education activity (i.e. grades) of students. They did not differentiate between different types of peers.

To further investigate this issue, we would now like to oppose the long-run effects to the short-run effects of peers on education by differentiating between the effect of long-lived ties and short-lived ties on school performance. For this purpose, we estimate the short-run
counterpart of equation (1):

\[ y_{i,r,t} = \phi^L \sum_{j=1}^{n_r} g^L_{ij,r,t} y_{j,r,t} + \phi^S \sum_{j=1}^{n_r} g^S_{ij,r,t} y_{j,r,t} + x'_{i,r,t} \delta 
+ \frac{1}{g^L_{i,r,t}} \sum_{j=1}^{n_r} g^L_{ij,r,t} x'_{j,r,t} \gamma^L + \frac{1}{g^S_{i,r,t}} \sum_{j=1}^{n_r} g^S_{ij,r,t} x'_{j,r,t} \gamma^S + \eta_{r,t} + \epsilon_{i,r,t}, \]  

(9)

where \( y_{i,r,t} \) is now the grade of student \( i \) who belongs to network \( r \) at time \( t \) where \( t \) refers to Wave II. The rest of the notation remains unchanged, which implies that we now deal with a traditional peer effects model where all individual and peer group characteristics are contemporaneous (i.e. in Wave II in 1995-1996). As in our investigation on long-run effects, we exploit variations in link formation in Waves I and II to differentiate between long-lived ties and short-lived ties. We then look at how these different types of peers affect each student’s grade obtained in Wave II. The identification and estimation strategy remains unchanged with the difference that now, we cannot use IV variables lagged in time (see Appendix C).

School performance is measured using the respondent’s scores received in Wave II in several subjects, namely English or language arts, history or social science, mathematics and science. The scores are coded as 1=D or lower, 2=C, 3=B, 4=A. For each individual, we calculate an index of school performance using a standard principal component analysis. The final composite index (labeled as GPA index or grade point average index) is the first principal component.\(^{35}\) It ranges between 0 and 6.09, with a mean equal to 2.29 and a standard deviation equal to 1.49.

The estimation results of model (9) are contained in Table 10. It appears that while in the long run, only long-lived ties matter, in the short run, both short and long-lived ties are important in determining a student’s performance at school. A standard deviation increase in aggregate GPA of peers translates, respectively, into a 8.1 (for long-lived ties) and a 4.8 (for short-lived ties) percent increase of a standard deviation in the individual’s GPA.

\[ \text{Insert Table 10 here} \]

Taking our analysis as a whole, our results suggest that, in the short run, all peers matter for education (i.e. grades) while, in the long run, only long-lived ties matter for future

\(^{35}\)The index explains roughly 56 percent of the total variance and captures a general performance at school since it is positively and highly correlated with the scores in all subjects. Further details on this procedure are available upon request.
educational choices (i.e. years of schooling). This is consistent with the model developed in Section 5.1 where, in the very short-run, both long-lived and short-lived friends have an impact on current beliefs while, in the long run, only long-lived friends have an impact on educational decisions since the latter are influenced by the social norm emerging from the iterations of beliefs. This is also consistent with the fact that, in the short-run analysis, the outcome is the students’ grades while, in the long-term, the outcome is the number years of education, where social norms matter more.

8 Concluding remarks

In this paper, we consider the estimation of heterogeneous spillover effects in a network model by looking at the impact of different types of friends made at school on education decisions. We find that a long-lived relationship has a positive impact on own education outcomes while a short-lived relationship does not.

Using the Moving to Opportunity (MTO) programs, Chetty et al. (2015) compare the long-run outcomes of children who moved to a better neighborhood as toddlers with those who moved in their late teens. They show that the exposure to low-poverty neighborhoods has a positive and significant impact on college education and adult earnings only for children who moved when they were less than 13 years old. That is, what matters for long-run outcomes are the years of exposure to a good neighborhood. If we think of friendship networks as “neighborhoods”, then this result is similar to the one obtained in our paper. More generally, our evidence is in line with several studies in sociology and economics (e.g. Coleman, 1988, Wellman and Wortley, 1990, Akerlof and Kranton, 2002), where long-lasting social interactions affect how much students value achievement, their human capital accumulation, and how social norms are formed.

References


Appendix A: Data Appendix

Table A1 provides a detailed description of the variables used in our study as well as the summary statistics for our sample. Among the individuals selected in our sample, 53 percent are female and 19 percent are blacks. The average parental education is high-school graduate. Roughly 10 percent have parents working in a managerial occupation, another 10 percent in the office or sales sector, 20 percent in a professional/technical occupation, and roughly 30 percent have parents in manual occupations. More than 70 percent of our individuals come from households with two married parents and from households of about four people on average. In Wave IV, 42 percent of our adolescents are now married and nearly half of them (43 percent) have at least a son or a daughter. The mean intensity in religion practice slightly decreases during the transition from adolescence to adulthood. On average, during their teenage years, our individuals felt that adults care about them and they had a good relationship with their teachers. Roughly, 30 percent of our adolescents were high-performing individuals at school, i.e. had the highest mark in mathematics. On average, these adolescents declare having the same number of best friends both in Wave I and Wave II (about 2.50 friends), although the composition of the friends changes.
Table A1: Data description and summary statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Description</th>
<th>Average (Std.Dev.)</th>
<th>Min - Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wave II (grade 7-12)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Individual socio-demographic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>Dummy variable taking value one if the respondent is female.</td>
<td>0.53 (0.50)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Black or African American</td>
<td>Race dummies. “White” is the reference group</td>
<td>0.19 (0.39)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Other races</td>
<td>Grade of student in the current year.</td>
<td>0.10 (0.30)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Student grade</td>
<td>“In the past 12 months, how often did you attend religious services?” coded as 2= never, 3= less than once a month, 4= once a month or more, but less than once a week, 5= once a week or more. Coded as 1 if the previous is skipped because of response “none” to the question: “What is your religion?” Mathematics score dummies. Score in mathematics in the most recent grading period. D is the reference category, coded (A, B, C, D, missing).</td>
<td>3.79 (1.83)</td>
<td>1 - 5</td>
</tr>
<tr>
<td>Religion practice</td>
<td></td>
<td>0.29 (0.45)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Mathematics score A</td>
<td></td>
<td>0.34 (0.48)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Mathematics score B</td>
<td></td>
<td>0.21 (0.41)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Mathematics score C</td>
<td></td>
<td>0.05 (0.21)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Mathematics score Missing</td>
<td>The school performance is measured using the respondent’s scores received in wave II in several subjects, namely English or language arts, history or social science, mathematics, and science. The scores are coded as 1=D or lower, 2=C, 3=B, 4=A. The final composite index is the first principal component score.</td>
<td>2.29 (1.49)</td>
<td>0 - 6.09</td>
</tr>
<tr>
<td>GPA</td>
<td>Sum of GPA attained by respondent’s peers</td>
<td>11.36 (8.85)</td>
<td>0 - 53.06</td>
</tr>
<tr>
<td>GPA of peers</td>
<td>Response to the question: “Compared with other people your age, how intelligent are you?”, coded as 1= moderately below average, 2= slightly below average, 3= about average, 4= slightly above average, 5= moderately above average, 6= extremely above average.</td>
<td>4.00 (1.09)</td>
<td>1 - 6</td>
</tr>
<tr>
<td>Self esteem</td>
<td>Response to the question: “How advanced is your physical development compared to other boys/girls your age?”, coded as 1= I look younger than most, 2= I look younger than some, 3= I look about average, 4= I look older than some, 5= I look older than most</td>
<td>3.31 (1.11)</td>
<td>1 - 5</td>
</tr>
<tr>
<td>Physical development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household size</td>
<td>Number of people living in the household</td>
<td>3.40 (1.34)</td>
<td>1 - 11</td>
</tr>
<tr>
<td>Two married parent family</td>
<td>Dummy taking value one if the respondent lives in a household with two parents (both biological and non biological) that are married. Schooling level of the 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 consider only the education of the father if both parents are in the household.</td>
<td>0.73 (0.44)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Parent education</td>
<td>Response to the question: “How much do you feel that adults care about you”, coded as 1= strongly agree, 2= agree, 3= neither agree nor disagree, 4= disagree, 5= strongly disagree.</td>
<td>3.25 (0.97)</td>
<td>1 - 5</td>
</tr>
<tr>
<td>Parent occupation manager</td>
<td>Response to the question: “You feel like you are part of your school coded as 1= strongly agree, 2= agree, 3= neither agree nor disagree, 4= disagree, 5= strongly disagree.</td>
<td>0.11 (0.31)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Parent occupation professional/technical</td>
<td>“none” is the reference group</td>
<td>0.21 (0.41)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Parent occupation office or sales worker</td>
<td></td>
<td>0.10 (0.33)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Parent occupation manual</td>
<td></td>
<td>0.30 (0.46)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Parent occupation other</td>
<td></td>
<td>0.14 (0.55)</td>
<td>0 - 1</td>
</tr>
<tr>
<td><strong>Protective factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School attachment</td>
<td>Response to the question: “How often have you had trouble getting along with your teachers?” coded as 0= never, 1= just a few times, 2= about once a week, 3= almost everyday, 4= everyday.</td>
<td>0.91 (0.94)</td>
<td>0 - 4</td>
</tr>
<tr>
<td>Relationship with teachers</td>
<td>Response to the question: “How advanced is your physical development compared to other boys/girls your age?”, coded as 1= I look younger than most, 2= I look younger than some, 3= I look about average, 4= I look older than some, 5= I look older than most</td>
<td>4.47 (0.73)</td>
<td>1 - 5</td>
</tr>
<tr>
<td>Social inclusion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Residential neighborhood</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residential building quality</td>
<td>Interviewer response to the question: “How well kept is the building in which the respondent lives”, coded as 1= very poorly kept, 2= poorly kept , 3= fairly well kept, 4= very well kept.</td>
<td>1.52 (0.80)</td>
<td>1 - 4</td>
</tr>
<tr>
<td><strong>Contextual effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave IV (aged 25 - 31)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years of education</td>
<td>Years of education attained by the individual.</td>
<td>14.42 (3.21)</td>
<td>7 - 24</td>
</tr>
<tr>
<td>Years of education of peers</td>
<td></td>
<td>35.73 (29.48)</td>
<td>7 - 326</td>
</tr>
<tr>
<td>Children</td>
<td>Dummy variable taking value one if the respondent has a child.</td>
<td>0.43 (0.50)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Married</td>
<td>Variable taking value one if the respondent is married</td>
<td>0.42 (0.49)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Religion practice</td>
<td>Response to the question: “How often have you attended religious services in the past 12 months?”, coded as 0= never, 1= a few times , 2= several times, 3= once a month, 4=2 or 3 times a month, 5= once a week, 6= more than once a week.</td>
<td>1.75 (1.64)</td>
<td>0 - 5</td>
</tr>
<tr>
<td><strong>Networks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Links in Wave I</td>
<td>Number of individual links in Wave I.</td>
<td>2.60 (2.57)</td>
<td>1 - 21</td>
</tr>
<tr>
<td>Links in Wave II</td>
<td>Number of individual links in Wave II.</td>
<td>2.49 (2.50)</td>
<td>1 - 26</td>
</tr>
<tr>
<td>Deleted links</td>
<td>Percentage of nominations in Wave I not renewed in Wave II.</td>
<td>0.61 (0.37)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>New links</td>
<td>Percentage of new nominations in Wave II.</td>
<td>0.44 (0.36)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Long-lived Ties</td>
<td>Percentage of Long-lived ties on total individual links.</td>
<td>0.28 (0.28)</td>
<td>0 - 1</td>
</tr>
<tr>
<td>Short-lived Ties</td>
<td>Percentage of Short-lived ties on total individual links.</td>
<td>0.72 (0.29)</td>
<td>0 - 1</td>
</tr>
</tbody>
</table>
Appendix B: A network game with ex-ante heterogenous agents

Preferences. We denote the educational effort level of individual $i$ by $y_i$ and the population effort profile by $y = (y_1, ..., y_n)'$. We characterize the short-lived and long-lived-tie relationships between two individuals by the strength of their relationship, denoted by $\phi$. This means that $\phi^L > \phi^S$. Each agent $i$ selects an effort $y_i \geq 0$, and obtains a payoff $u_i(y, g)$ that depends on the effort profile $y$ and the underlying network $g$, in the following way:

$$u_i(y, g) = (a_i + \eta + \varepsilon_i) y_i - \frac{1}{2}y_i^2 + \left( \phi^L \sum_{j \in N^L_i(g)} y_j + \phi^S \sum_{j \in N^S_i(g)} y_j \right) y_i$$

(10)

where $\phi^L, \phi^S > 0$, with $\phi^L > \phi^S$. 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; 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 $\phi^L \sum_{j \in N^L_i(g)} y_j y_j + \phi^S \sum_{j \in N^S_i(g)} y_j y_j$. The main difference to the standard approach is that here, we have heterogenous peer effects since long-lived and short-lived ties have different impacts on own utility. Indeed, we have:

$$\frac{\partial u_i(y, g)}{\partial y_i} = g^L_{ij} \phi^L + g^S_{ij} \phi^S \geq 0$$

where $g^L_{ij} = 1$ ($g^S_{ij} = 1$) if there exists a long-lived-tie (short-lived-tie) relationship between $i$ and $j$ and zero otherwise. To the best of our knowledge, this is the first paper that introduces different $\phi$ in a network model with strategic complementarities.

Observe that $\eta$ denotes the unobservable network characteristics, $\varepsilon_i$ 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:

$$a_i = \sum_{m=1}^{M} \beta_m x_i^m + \frac{1}{g^L_i} \sum_{m=1}^{M} \sum_{j=1}^{n^L} g^L_{ij} x_j^m x_i^m + \frac{1}{g^S_i} \sum_{m=1}^{M} \sum_{j=1}^{n^S} g^S_{ij} x_j^m x_i^m$$

(11)
where \( x_i^m \) is a set of \( M \) variables accounting for observable differences in the individual characteristics of individual \( i \), \( \beta_m, \gamma^L_m, \gamma^S_m \) are parameters and \( g_i^L = \sum_{j=1}^n g_{ij}^L \) and \( g_i^S = \sum_{j=1}^n g_{ij}^S \) constitute the total number of long-lived-tie and short-lived-tie friends of individual \( i \).

To summarize, when individual \( i \) exerts some effort in education, the benefits of the activity depend on own effects (i.e. on individual characteristics \( a_i \), some network characteristics \( \eta \) and some random element \( \varepsilon_i \), which is specific to individual \( i \) and non-observable by the researcher) and on peer effects, where the strength of interactions differs between long-lived and short-lived ties. Note that the utility (10) is concave in own decisions, and displays decreasing marginal returns in own effort levels. In sum,

\[
\begin{align*}
\sum_{i=1}^{\infty} \left( \frac{a_i + \eta + \varepsilon_i}{2} \right) g_{ii} + \sum_{j=1}^{n} \sum_{i=1}^{n} g_{ij} y_i y_j + \sum_{j=1}^{n} g_{ij} y_i y_j + a_i + \eta + \varepsilon_i
\end{align*}
\]

Nash equilibrium  We now characterize the Nash equilibrium of the game where agents choose their effort level \( y_i \geq 0 \) simultaneously. In equilibrium, each agent maximizes her utility (10) and we obtain the following best-reply function for each \( i = 1, \ldots, n \):

\[
y_i = \phi^L \sum_{j=1}^{n} g_{ij} y_j + \phi^S \sum_{j=1}^{n} g_{ij} y_j + a_i + \eta + \varepsilon_i
\]

where \( a_i \) is given by (11). This equation corresponds to (1). Denote \( \alpha_i = a_i + \eta + \varepsilon_i \) and the corresponding \((1 \times n)\) vector by \( \alpha \). The matrix form equivalent of (12) is:

\[
y = (I - \phi^L G^L - \phi^S G^S)^{-1} \alpha.
\]

Denote by \( \mu_1(G) \) the spectral radius of \( G \). We have:

**Proposition 1** If \( \mu_1 \left( \phi^L G^L + \phi^S G^S \right) < 1 \), the peer effect game with payoffs (10) has a unique interior Nash equilibrium in pure strategies given by (12) or by (13).

**Proof:** We need to show that \( I - A \) is non-singular (i.e. invertible), where \( A = \phi^L G^L + \phi^S G^S \). We know that \( I - A \) is non-singular if \( \mu_1 \left( \phi^L G^L + \phi^S G^S \right) < 1 \) (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.
Corollary 2 Assume that $G$ is symmetric. A sufficient condition for the Nash equilibrium given by (12) or (13) to exist, to be unique and to be interior is: \((\phi^L + \phi^S) \mu_1(G) < 1\).

**Proof:** Let us start with two lemmas. Denote by $\mu_1(A)$ the spectral radius of $A$.

**Lemma 3** If $A$ is an $n \times n$ Hermitian matrix, then

$$
\mu_1(A) = \sup_{\|v\|=1} v^T Av
$$

(14)

**Proof:** The Rayleigh–Ritz quotient is defined as:

$$
R(A, x) = \frac{x^T Ax}{x^T x} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij}}{x^T x}
$$

That is

$$
R(A, x) = \frac{x^T Ax}{x^T x} = \frac{\sum_{i=1}^{n} x_i a_{ii}}{x^T x} = \left( \frac{x}{\|x\|_2} \right)^T A \left( \frac{x}{\|x\|_2} \right)
$$

where $\|x\|_2 \equiv \sqrt{\sum_{i=1}^{n} |x_i|^2}$ is the Euclidian norm (or vector 2-norm). We want to compute

$$
\sup_{x \in \mathbb{R}^n \setminus \{0\}} R(A, x) = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \left( \frac{x}{\|x\|_2} \right)^T A \left( \frac{x}{\|x\|_2} \right)
$$

Define $y \equiv \frac{x}{\|x\|_2}$ so that $y^T y = 1$, then

$$
\sup_{x \in \mathbb{R}^n \setminus \{0\}} R(A, x) = \sup_{y \in \mathbb{R}^n \setminus \{0\}} \left\{ y^T A y \mid \|y\|_2 = 1 \right\}
$$

We want to show that $\sup_{y \in \mathbb{R}^n \setminus \{0\}} \left\{ y^T A y \mid \|y\|_2 = 1 \right\} = \mu_1(A)$. Observe that the function $y \rightarrow y^T A y$ is continuous with compact domain ($n-1$ dimensional sphere). Every continuous function attains a maximum and a minimum on a compact set. There is thus a $y$ so that the sup is a maximum. The problem is to find the critical points of the function $y \rightarrow y^T A y$ subject to the constraint $\|y\|_2 = y^T y = 1$, i.e. to find the critical points of the following
lagrangian

\[ \mathcal{L}(y) = y^T Ay - \mu (y^T y - 1) \]

where \( \mu \) is a Lagrange multiplier. The stationary points of \( y \rightarrow \mathcal{L}(y) \) are given by: \( \frac{d\mathcal{L}(y)}{dy} = 0 \).

We obtain:

\[ \frac{d\mathcal{L}(y)}{dy} = (A + A^T) y - \mu (I + I^T) y = 0 \]

Using the fact that \( A \) is symmetric, i.e. \( A = A^T \), then

\[ \frac{d\mathcal{L}(y)}{dy} = 2Ay - 2\mu y = 0 \]

which is equivalent to

\[ Ay = \mu y \]

This means that \( y \) is an eigenvector and \( \mu \) the associated eigenvalue. Therefore, the eigenvectors \( y_1, ..., y_n \) of \( A \) are the critical points of the Rayleigh Quotient and their corresponding eigenvalues \( \mu_1, ..., \mu_n \), with \( \mu_1 \geq ... \geq \mu_n \), are the stationary values of \( R(A, x) \).

At the stationary points, we have

\[ y^T Ay = y^T \mu y = \mu y^T y = \mu \]

This implies that

\[ \sup_{x \in \mathbb{R}^n \setminus \{0\}} R(A, x) = \sup_{y \in \mathbb{R}^n \setminus \{0\}} \{ y^T Ay \mid \|y\|_2 = 1 \} = \mu_1(A) \]

We have thus shown that \( \mu_1(A) = \sup_{\|v\|=1} v^T Av \). \( \square \)

**Lemma 4** If \( A \) and \( B \) are two \( n \times n \) symmetric matrices, then

\[ \mu_1(A + B) \leq \mu_1(A) + \mu_1(B) \] (15)

**Proof:** In Lemma 3, we have shown that

\[ \mu_1(A) = \sup_{\|v\|=1} v^T Av \]

which implies that

\[ \mu_1(B) = \sup_{\|v\|=1} v^T Bv \]

38
and

\[ \mu_1 (A+B) = \sup_{|v|=1} v^T (A+B) v \]

We need now to show that

\[ \mu_1 (A+B) \leq \mu_1 (A) + \mu_1 (B) \]

Using the sub-additivity of the sup function, we have

\[
\begin{align*}
\mu_1 (A+B) &= \sup_{|v|=1} v^T (A+B) v \\
&= \sup_{|v|=1} \left\{ v^T Av + v^T Bv \right\} \\
&\leq \sup_{|v|=1} v^T Av + \sup_{|v|=1} v^T Bv \\
&= \mu_1 (A) + \mu_1 (B)
\end{align*}
\]

which is the statement of the lemma. \[\blacksquare\]

In Proposition 8, we have shown that \( I - A \) is non-singular if \( \mu_1 (\phi^L G^L + \phi^S G^S) < 1 \). Assume now that \( G \) is symmetric so that both \( G^L \) and \( G^S \) also are symmetric. We can use Lemma 4, which states that (given \( \phi^L > 0 \) and \( \phi^S > 0 \)):

\[
\mu_1 (\phi^L G^L + \phi^S G^S) \leq \mu_1 (\phi^L G^L) + \mu_1 (\phi^S G^S) = \phi^L \mu_1 (G^L) + \phi^S \mu_1 (G^S)
\]

For \( A = \{a_{ij}\} \) and \( B = \{b_{ij}\} \), we say \( A \leq B \) if \( a_{ij} \leq b_{ij} \) for all \( i, j \). In our context, this means that \( 0 \leq G^L \leq G \) and \( 0 \leq G^S \leq G \) (since \( g^L_{ij} \leq g_{ij} \) and \( g^S_{ij} \leq g_{ij} \) for all \( i, j \)). As a result, using Theorem I*, page 600 in Debreu and Herstein (1953), we have: \( \mu_1 (G^L) \leq \mu_1 (G) \) and \( \mu_1 (G^S) \leq \mu_1 (G) \). This implies that: \( \phi^L \mu_1 (G^L) + \phi^S \mu_1 (G^S) \leq (\phi^L + \phi^S) \mu_1 (G) \) and the condition for \( I - \phi^L G^L - \phi^S G^S \) to be non-singular is given by:

\[ (\phi^L + \phi^S) \mu_1 (G) < 1 \]

which is the condition given in Corollary 2. \[\blacksquare\]

Corollary 2 gives an interesting result because it connects the adjacency matrix \( G \) to the split structure of peer effects \( \phi^L \) and \( \phi^S \) and it is directly comparable to the condition given in Ballester et al. (2006), i.e. \( \phi \mu_1 (G) < 1 \), where the peer effects were assumed to be the
same across all agents. The condition given in Proposition 8, \( \mu_1 (\phi^L G^L + \phi^S G^S) < 1 \), is less restrictive than \((\phi^L + \phi^S) \mu_1 (G) < 1 \) since \( \mu_1 (\phi^L G^L + \phi^S G^S) \leq \phi^L \mu_1 (G^L) + \phi^S \mu_1 (G^S) \) (see the proof of Corollary 2). However, it imposes a restriction on the exact topology of both short-lived and long-lived tie networks while \((\phi^L + \phi^S) \mu_1 (G) < 1 \) allows \( G^L \) and \( G^S \) to be less constrained in terms of the topology of the network but, more importantly, it does not require the matrix \( G \) (and thus \( G^L \) and \( G^S \)) to be symmetric.

To illustrate this result, consider a star network where individual A is the star and individuals B and C are in the periphery. Its adjacency matrix \( G \) (where individual A corresponds to the first row, individual B to the second and individual C to the third row) is given by:

\[
G = \begin{pmatrix}
0 & 1 & 1 \\
1 & 0 & 0 \\
1 & 0 & 0
\end{pmatrix}.
\]

Assume that individual A has a short-lived-tie relationship with B but a long-lived-tie relationship with C. We easily obtain:

\[
G^L = \begin{pmatrix}
0 & 0 & 1 \\
0 & 0 & 0 \\
1 & 0 & 0
\end{pmatrix} \quad \text{and} \quad G^S = \begin{pmatrix}
0 & 1 & 0 \\
1 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}.
\]

The largest eigenvalue of \( G \) is \( \sqrt{2} \) while the largest eigenvalue of \( G^L \) and of \( G^S \) is 1. Thus, the sufficient condition given in Corollary 2 is: \( \phi^L + \phi^S < 0.707 \). In that case, there exists a unique Nash equilibrium given by:

\[
y = \left( I - \phi^L G^L - \phi^S G^S \right)^{-1} \alpha
\]

\[
= \frac{1}{1 - (\phi^L)^2 - (\phi^S)^2} \begin{pmatrix}
\alpha_A + \alpha_B \phi^S + \alpha_C \phi^L \\
\alpha_A \phi^S + \alpha_B \left[ 1 - (\phi^L)^2 \right] + \alpha_C \phi^S \phi^L \\
\alpha_A \phi^S + \alpha_B \phi^S \phi^L + \alpha_C \left[ 1 - (\phi^S)^2 \right]
\end{pmatrix}.
\]

Interestingly, even though individuals B and C have one link each, it is easily verified that C’s educational effort is much higher than that of B and much closer to A, who has two links. This shows the importance of short-lived and long-lived ties in the relationship between different individuals.

Take, for example, \( \alpha_1 = \alpha_2 = \alpha_3 = 1 \). If, for example, \( \phi^L = 0.8 \) and \( \phi^S = 0.2 \) (which implies that \( (\phi^L)^2 + (\phi^S)^2 = 0.68 < 1 \)), then the condition given in Proposition 8 is satisfied
since $\mu_1 \left( \phi^L G^L + \phi^S G^S \right) = 0.825 < 1$ but the one given in Corollary 9 is not since $\phi^L + \phi^S = 1 > 0.707$. Since $\mu_1 \left( \phi^L G^L + \phi^S G^S \right) = 0.825 < 1$, we can still derive the equilibrium and obtain:

$$y_A^* = 6.25, \quad y_B^* = 4.125 \quad \text{and} \quad y_C^* = 6.$$
Appendix C: Estimation - Technical details -

Our econometric methodology extends Liu and Lee’s (2010) 2SLS estimation strategy to a social interaction model with two different network structures. Let us expose this approach and highlight the modification that is implemented in this paper. Model (3) can be written as:

\[ Y = Z\theta + \iota \cdot \eta + \epsilon, \]  

(16)

where \( Z = (G^L Y, G^S Y, X^*) \), \( \theta = (\phi^L, \phi^S, \beta')' \) and \( \iota = D(l_{n_1}, \ldots, l_{n_r}) \).

We treat \( \eta \) as a vector of unknown parameters. When the number of networks \( \bar{r} \) is large, we have the incidental parameter problem. Let \( J = D(J_1, \ldots, J_{\bar{r}}) \), where \( J_r = I_{n_r} - \frac{1}{n_r} Y_{r1} l_{n_r} \).

The network fixed effect can be eliminated by a transformation with \( J \) such that:

\[ JY = JZ\theta + J\epsilon. \]  

(17)

Let \( M = (I - \phi^L G^L - \phi^S G^S)^{-1} \). The equilibrium outcome vector \( Y \) in (16) is then given by the reduced form equation:

\[ Y = M(X^*\beta + \iota \cdot \eta) + M\epsilon. \]  

(18)

It follows that \( G^L Y = G^L MX^*\beta + G^L M\iota \eta + G^L M\epsilon \) and \( G^S Y = G^S MX^*\beta + G^S M\iota \eta + G^S M\epsilon \). \( G^L Y \) and \( G^S Y \) are correlated with \( \epsilon \) because \( E[(G^L M\epsilon)'\epsilon] = \sigma^2 \text{tr}(G^L M) \neq 0 \) and \( E[(G^S M\epsilon)'\epsilon] = \sigma^2 \text{tr}(G^S M) \neq 0 \). Hence, in general, (17) cannot be consistently estimated by OLS.\(^{36}\) If \( G \) is row-normalized such that \( G \cdot l_n = l_n \), where \( l_n \) is an \( n \)-dimensional vector of ones, the endogenous social interaction effect can be interpreted as an average effect.

Liu and Lee (2010) use an instrumental variable approach and propose different estimators based on different instrumental matrices, here denoted by \( Q_1 \) and \( Q_2 \). They first consider the 2SLS estimator based on the conventional instrumental matrix for the estimation of (17): \( Q_1 = J(GX^*, X^*) \) (finite-IVs 2SLS). Then, they propose to use additional instruments (IVs) \( JG_l \) and enlarge the instrumental matrix: \( Q_2 = (Q_1, JG_l) \) (many-IVs 2SLS). The additional IVs of \( JG_l \) are based on the row sums of \( G \) and are indicators of centrality in the networks.

Liu and Lee (2010) show that those additional IVs could help model identification when the conventional IVs are short-lived and improve on the estimation efficiency of the conventional 2SLS estimator based on \( Q_1 \). However, the number of such instruments depends on the number of networks. If the number of networks grows with the sample size, so does the

\(^{36}\)Lee (2002) has shown that the OLS estimator can be consistent in the spatial scenario where each spatial unit is influenced by many neighbors whose influences are uniformly small. However, in the current data, the number of neighbors are limited, and hence that result does not apply.
number of IVs. The 2SLS could be asymptotic biased when the number of IVs increases too fast relative to the sample size (see, e.g., Bekker, 1994; Bekker and van der Ploeg, 2005; Hansen et al., 2008). As detailed in Section 2, in this empirical study, we have a number of small networks. Liu and Lee (2010) also propose a bias-correction procedure based on the estimated leading-order many-IV bias (bias-corrected 2SLS). The bias-corrected many-IV 2SLS estimator is properly centered, asymptotically normally distributed, and efficient when the average group size is sufficiently large. Thus, it is the more appropriate estimator in our case study.\footnote{Liu and Lee (2010) also generalize this 2SLS approach to the GMM using additional quadratic moment conditions.}

Let us now derive the best 2SLS estimator for equation (17). From the reduced form equation (16), we have \( E(Z) = [G^L M (X^\beta + \eta \cdot \eta), G^S M (X^\beta + \eta \cdot \eta), X^\ast] \). The best IV matrix for \( JZ \) is given by

\[
Jf = JE(Z) = J[G^L M (X^\beta + \eta \cdot \eta), G^S M (X^\beta + \eta \cdot \eta), X^\ast]
\]

which is an \( n \times (3m + 2) \) matrix. However, this matrix is infeasible as it involves unknown parameters. Note that \( f \) can be considered as a linear combination of the vectors in \( Q_0 = J[G^L M (X^\ast + \omega), G^S M (X^\ast + \omega), X^\ast] \). As \( \omega \) has \( \bar{r} \) columns the number of IVs in \( Q_0 \) increases as the number of groups increases. Furthermore, as \( M = (I - \phi^L G^L - \phi^S G^S)^{-1} = \sum_{j=0}^{\infty} (\phi^L G^L + \phi^S G^S)^j \) when sup ||\( \phi^L G^L + \phi^S G^S ||_\infty < 1 \), \( MX^\ast \) and \( M\omega \), can be approximated by linear combinations of

\[
(G^L X^\ast, G^S X^\ast, G^S G^L X^\ast, (G^L)^2 X^\ast, (G^S)^2 X^\ast, (G^S)^2 G^L X^\ast, (G^L)^2 X^\ast, \ldots)
\]

and

\[
(G^L \omega, G^S \omega, G^S G^L \omega, (G^L)^2 \omega, (G^S)^2 \omega, (G^S)^2 G^L \omega, (G^L)^2 \omega, \ldots),
\]

respectively. Hence, \( Q_0 \) can be approximated by a linear combination of

\[
Q_\infty = J(G^L (G^L X^\ast, G^S X^\ast, G^S G^L X^\ast, \ldots, G^L \omega, G^S \omega, G^S G^L \omega, \ldots), G^S (G^L X^\ast, G^S X^\ast, G^S G^L X^\ast, \ldots, G^L \omega, G^S \omega, G^S G^L \omega, \ldots), X^\ast).
\]

Let \( Q_K \) be an \( n \times K \) submatrix of \( Q_\infty \) (with \( K \geq 3m + 2 \)) including \( X^\ast \). Let \( Q_S \) be an \( n \times K_S \) submatrix of \( Q_{\infty} = G^L (G^L X^\ast, G^S X^\ast, G^S G^L X^\ast, \ldots, G^L \omega, G^S \omega, G^S G^L \omega, \ldots) \) and \( Q_S \) an \( n \times K_S \) submatrix of \( Q_{W_\infty} = G^S (G^L X^\ast, G^S X^\ast, G^S G^L X^\ast, \ldots, G^L \omega, G^S \omega, G^S G^L \omega, \ldots) \). We assume that \( \frac{K_i}{K_S} = 1 \). Let \( P_K = Q_K (Q_K' Q_K)^{-1} Q_K' \) be the projector of \( Q_K \). The resulting
2SLS estimator is given by:

$$\hat{\theta}_{2sls} = (Z'P_KZ)^{-1}Z'P_Ky.$$  (20)

**Asymptotic properties of 2SLS estimator**

As shown by Liu and Lee (2010), the 2SLS with a fixed number of IVs would be consistent but not efficient. Asymptotic efficiency can be achieved using a sequence of IVs in which the number of IVs grows slow enough relative to the sample size. In general, $K$ may be seen as an increasing function of $n$. Following Liu and Lee (2010), we assume the following regularity conditions:

**Assumption C1:** The elements of $\epsilon$ are i.i.d with zero mean, variance $\sigma^2$ and a moment of order higher than four exists.

**Assumption C2:** The elements of $X^*$ are uniformly bounded constants, $X^*$ has the full rank $K$ and $\lim_{n \to \infty} \frac{1}{n}f'f$ is a finite non singular matrix.

**Assumption C3:** The sequences of matrices $\{G^L\}, \{G^S\}, \{M\}$ are uniformly bounded.

**Assumption C4:** $H = \lim_{n \to \infty} \frac{1}{n}f'f$ is a finite non singular matrix.

**Assumption C5:** There exists a $K \times (3m+2)$ matrix $\pi_K$ such that $\|f - Q_K \pi_K\|_\infty \to 0$ as $n, K \to \infty$.

The 2SLS estimator with an increasing number of IVs approximating $f$ can be asymptotically efficient under some conditions. However, when the number of instruments increases too fast, such an estimator could be asymptotically biased, which is known as the many-instrument problem. Let $\Psi_{K,S} = P_K G^L M$ and $\Psi_{K,W} = P_K G^S M$. The following proposition shows consistency and asymptotic normality of the 2SLS estimator (20).

**Proposition 5** Under assumptions C1-C5, if $K/n \to 0$, then $\sqrt{n} \left( \hat{\theta} - \theta - b_{2sls} \right) \overset{d}{\to} N \left( 0, \sigma^2 H^{-1} \right)$, where $b_{2sls} = \sigma^2 (Z'P_KZ)^{-1} \left[ \text{tr} (\Psi_{K,L}) , \text{tr} (\Psi_{K,S}) , 0_{3m \times 1} \right]' = O_p \left( K/n \right)$.

**Proof:** Let $JZ = J(f+\nu)$, where $\nu = [G^L \epsilon, G^S \epsilon, 0_{n \times 3m}]$. Assuming Lemma B.1-3 in Liu and Lee (2010) and Lemma A.3 in Donald and Newey (2001), we have

$$\frac{1}{n}Z'P_KZ = H - e_f + \frac{1}{n}v'P_K f + \frac{1}{n}f'P_K v + \frac{1}{n}v'P_K v = H + O(\text{tr}(e_f)) + O_p(\sqrt{K/n}) + O_p(K/n) = H + o_p(1)$$
where $H = \frac{1}{n}f'f$ and $e_f = \frac{1}{n}f'(I - P_K)f$, because $e_f = O(tr(e_f))$, $\frac{1}{n}v'P_Kv = O_p(K/n)$ and $\frac{1}{n}v'P_Kf = O_p(\sqrt{K/n})$. Furthermore, we have

$$
(Z'P_K\epsilon - \sigma^2[tr(\Psi_{KL}), tr(\Psi_{KS}), 0_{3m\times 1}])/\sqrt{n}
$$

$$= h - f'(I - P_K)\epsilon/\sqrt{n} + (\frac{1}{n}v'P_K\epsilon - \sigma^2[tr(\Psi_{KL}), tr(\Psi_{KS}), 0_{3m\times 1}])/\sqrt{n}
$$

$$= h + O_p(\sqrt{tr(e_f)}) + O_p(\sqrt{K/n})
$$

$$= h + o_p(1) \xrightarrow{d} N(0, \sigma^2H^{-1})
$$

where $h = f'\epsilon/\sqrt{n}$. Then, applying the Slutzky theorem, the proposition follows.

Due to the increasing number of IVs $\sqrt{n}(\hat{\theta} - \theta)$ has the bias $\sqrt{n}b_{2SLS}$, when $K^2/n \to 0$ the bias term converges to zero and the sequence of IV matrices $Q_K$ gives the best IV estimator as $\sigma^2H^{-1}$ reaches the efficiency lower bound for the IV estimators.

**Corollary 6** Under assumptions C1-C5, if $K^2/n \to 0$, then $\sqrt{n}(\hat{\theta} - \theta) N\left(0, \sigma^2H^{-1}\right)$.

In summary, having a sequence of IV matrices $\{Q_K\}$, condition $K/n \to 0$ is fundamental for the estimator to be consistent, while having $K^2/n \to 0$ provides the asymptotically best estimator because $\sigma^2H^{-1}$ brings the lower bound for the IV estimators.

In this paper, we use 2SLS estimators and propose two innovations. First, we use two centralities, one for long-lived ties and one for short-lived ties in $Q_2$ (many-IVs 2SLS). Second, we take advantage of the longitudinal structure of our data and only include in the different instrumental matrices values lagged in time (i.e. observed in wave I). Let $Q_{1L}$ and $Q_{2L}$ denote the set of instruments $Q_1$ and $Q_2$ which only include variables in Wave I (i.e. lagged in time).

Note that $[G^L, G^S]$ has $2\bar{r}$ columns, so if we include Bonacich centralities for both short-lived and long-lived ties from each of the $\bar{r}$ groups in $Q_K$, then $2\bar{r}/K \to 0$. Hence, $K/n \to 0$ implies that $2\bar{r}/n = 2/\bar{s} \to 0$ where $\bar{s}$ is the average group size. Then, as shown by Liu and Lee (2010) for the case of a single endogenous variable (i.e. coming from one interaction matrix), the average group size needs to be large enough, it should also be large relative to the number of groups because for the asymptotic efficiency it must be $K^2/n \to 0$ and it implies $(2\bar{r})^2/n = 2\bar{r}/\bar{s} \to 0$. If the network is not characterized by these properties, a bias correction should be used. Given the topology of the Add Health network, which is composed by quite a large number of relatively small networks, the best (feasible) estimator
is the bias-corrected one

$$\tilde{\theta}_{2sls} = (Z'P_{Q_k}Z)^{-1} \left[ Z'P_{Q_k}y - \sigma^2 \left[ \text{tr} \left( \tilde{\Psi}_{KL} \right), \text{tr} \left( \tilde{\Psi}_{KS} \right), 0_{3m \times 1} \right] \right], \quad (21)$$

where $\tilde{\Psi}_{KS} = P_KG^L\tilde{M}(\tilde{\phi}^L, \tilde{\phi}^S)$ and $\tilde{\Psi}_{KS} = P_KG^S\tilde{M}(\tilde{\phi}^L, \tilde{\phi}^S)$ are estimated with initial $\sqrt{n}$-consistent estimators of $\tilde{\sigma}, \tilde{\phi}^L$ and $\tilde{\phi}^S$. This estimator adjusts the 2SLS estimator by the estimated leading order bias $b_{2sls}$, which is presented in Proposition 5.

**Proposition 7** Under assumptions C1-C5, if $K/n \to 0$ and $\tilde{\sigma}, \tilde{\phi}^L$ and $\tilde{\phi}^S$ are $\sqrt{n}$-consistent initial estimators of $\sigma, \phi^L$ and $\phi^S$, then $\sqrt{n}(\tilde{\theta}_{2sls} - \theta) \xrightarrow{d} N \left( 0, \sigma^2\Pi^{-1} \right)$.

**Proof:** We need to show that

$$\{\sigma^2 \left[ \text{tr} \left( P_KG^L\tilde{M} \right), \text{tr} \left( P_KG^S\tilde{M} \right) \right]' - \sigma^2 \left[ \text{tr} \left( \Psi_{KL} \right), \text{tr} \left( \Psi_{KS} \right) \right] \}/\sqrt{n} = o_p(1)$$

where $\tilde{M} = M(\tilde{\phi}^L, \tilde{\phi}^S)$. Given Proposition 5, this is quite straightforward since

$$\{\sigma^2 \left[ \text{tr} \left( P_KG^L\tilde{M} \right), \text{tr} \left( P_KG^S\tilde{M} \right) \right]' - \sigma^2 \left[ \text{tr} \left( \Psi_{KL} \right), \text{tr} \left( \Psi_{KS} \right) \right] \}/\sqrt{n} = \sqrt{n}(\sigma^2 - \sigma^2) \left[ \text{tr} \left( P_KG^L\tilde{M} \right), \text{tr} \left( P_KG^S\tilde{M} \right) \right]' / n + \sqrt{n}\sigma^2 \left[ \text{tr} \left( P_KG^L(M - \tilde{M}) \right), \text{tr} \left( P_KG^S(M - \tilde{M}) \right) \right]' / n$$

$$= \sqrt{n}(\sigma^2 - \sigma^2) \left[ \text{tr} \left( P_KG^L\tilde{M} \right), \text{tr} \left( P_KG^S\tilde{M} \right) \right]' / n + \sqrt{n}\sigma^2 \left[ \text{tr} \left( P_KG^L\tilde{M} \right), \text{tr} \left( P_KG^S\tilde{M} \right) \right]' / n$$

$$+ \sqrt{n}\sigma^2 \left[ \text{tr} \left( P_KG^L\tilde{M} \right), \text{tr} \left( P_KG^S\tilde{M} \right) \right]' / n$$

$$= O_p(\sqrt{K/n}) = o_p(1)$$

because $\tilde{M} = M(\tilde{\phi}^L, \tilde{\phi}^S)$ considered in this paper are:

(i) *Finite-IV* : $\tilde{\theta}_{2sls} = (Z'P_1Z)^{-1}Z'P_1y$, where $P_1 = Q_1Q_1^{-1}Q_1'$ and $Q_1$ contains the linearly independent columns of $J[X^*, GX^*, GGX^*]$. 46
(ii) Many-IV: \( \hat{\theta}_{2ls2} = (Z'P_zZ)^{-1}Z'P_zy \), where \( P_2 = Q_2(Q_2'Q_2)^{-1}Q_2' \) and \( Q_2 \) contains the linearly independent columns of \([Q_1, JG^L, JG^S] \).

(iii) Bias-corrected: \( \hat{\theta}_{2ls} = (Z'P_zZ)^{-1}\{Z'P_zy - \tilde{\sigma}^2_{2ls1}[\text{tr}(P_2G^L\tilde{M}), \text{tr}(P_2G^S\tilde{M}), 0_{3m \times 1}]\} \), where \( \tilde{M} = (I - \phi^L_{2ls1}G^L - \phi^S_{2ls1}G^S)^{-1} \), and \( \tilde{\sigma}^2_{2ls1}, \phi^L_{2ls1} \) and \( \phi^S_{2ls1} \) are \( \sqrt{n} \)-consistent initial estimators of \( \sigma^2, \phi^L \) and \( \phi^S \) obtained by Finite-IV.

(iv) Finite-IV lagged: \( \hat{\theta}_{2ls1L} = (Z'P_3Z)^{-1}Z'P_3y \), where \( P_3 = Q_{1L}(Q_{1L}'Q_{1L})^{-1}Q_{1L}' \) and \( Q_{1L} \) contains the linearly independent and lagged in time columns of \( J[X^*, GX^*, GGX^*] \).

(v) Many-IV lagged: \( \hat{\theta}_{2ls2L} = (Z'P_4Z)^{-1}Z'P_4y \), where \( P_4 = Q_{2L}(Q_{2L}'Q_{2L})^{-1}Q_{2L}' \) and \( Q_{2L} \) contains the linearly independent columns of \([Q_{1L}, JG^L, JG^S] \)

(vi) Bias-corrected lagged:

\[
\hat{\theta}_{2ls1L} = (Z'P_4Z)^{-1}\{Z'P_4y - \tilde{\sigma}^2_{2ls1}[\text{tr}(P_4G^L\tilde{M}), \text{tr}(P_4G^S\tilde{M}), 0_{3m \times 1}]\}
\]

where \( \tilde{M} = (I - \phi^L_{2ls1L}G^L - \phi^S_{2ls1L}G^S)^{-1} \), and \( \tilde{\sigma}^2_{2ls1L}, \phi^L_{2ls1L} \) and \( \phi^S_{2ls1L} \) are \( \sqrt{n} \)-consistent initial estimators of \( \sigma^2, \phi^L \) and \( \phi^S \) obtained by Finite-IV lagged.
Appendix D: The DeGroot model

The network  Consider a society consisting of a finite set of individuals $N = \{1, 2, \ldots, n\}$ who would like to gather information about an unknown parameter $\theta$ of interest or form an opinion concerning an issue they need to make a decision on. Agents only concern is to estimate the true value of the unknown parameter $\theta$ by following an updating process; they do not seek to maximize their social influence, nor can they gain something by trying to propagate a particular belief.

The pattern of communication among the agents is captured by a directed network $G$. In other words, the adjacency matrix $G$ captures the pattern of communication of opinions across the network. A link from agent $i$ to agent $j$, $g_{ij} = 1$, has the interpretation that agent $i$ can observe agent $j$'s belief, i.e. $j$ is a neighbor (or more precisely an out-neighbor) of $i$. All out-neighbors of $i$ form his/her out-neighborhood $D(i) \subseteq N$. The network is directed because $g_{ij} = 1$ means that agent $i$ observes, pays attention to or listens to agent $j$ but not necessarily the reverse, i.e. attention may not be reciprocal. It is also reasonable to assume that every agent can observe (or pay attention to) him/herself. The diagonal of $G$ will thus consist of ones, that is $g_{ii} = 1$, for all $i \in N$.

Definition 1 A network $G$ is strongly connected if there exists a directed path from any node to any other node in $G$.

The belief updating process  In the DeGroot model, agents start with some initial beliefs, which they update through communication with their neighbors. The initial belief of agent $i$ is denoted by $b_i^{(0)}$. Here $b_i^{(0)}$ is a probability and lies in the interval $[0, 1]$. Hence the $n$–dimensional vector $\mathbf{b}^{(0)}$ denotes the agents’ initial beliefs. In our context, $b_i^{(0)}$ can be thought of as the probability that the statement \{It is worth continuing studying\} or the statement \{It is not worth continuing studying\} is true. In each round, agents ask their neighbors for their beliefs, as well as an assessment of how precise or accurate these beliefs are. Then they update their belief by weighing the reports they get based on the precision assessment reported. The belief of agent $i$ after $t$ rounds of communication, where $t \in \{0, 1, 2, \ldots\}$, will be denoted with $b_i^{(t)} \in \mathbb{R}$. The beliefs of all agents in $G$ after $t$ rounds of communication can be stacked into a vector $\mathbf{b}^{(t)} \in \mathbb{R}^n$.

DeGroot (1974) assumes that agents update their beliefs by repeatedly taking weighted averages of their neighbors’ beliefs with $p_{ij}$ being the weight that agent $i$ places on the
current belief of agent \(j\) in forming his or her belief for the next period. To be more precise, in the beginning, each agent \(i \in N\) assigns a (relative) weight \(p_{ij}\) on each of his/her neighbors \(j\) such that \(\sum_{j=1}^{n} p_{ij} = 1\). Observe that \(p_{ij}\) is the direct influence of agent \(j\) on \(i\) so that \(p_{ij} = 0\) for \(j \not\in D(i)\).

One way to interpret this model is as follows. Each agent \(i\) assigns an initial precision of \(\pi_{i} \in \mathbb{R}_{+}\) to his/her signal. This precision can be based on some arbitrary assessment of the agent, or it could be an objectively defined statistic (DeMarzo et al., 2003; Golub and Jackson, 2010, 2012). If, for example, the initial beliefs \(b^{(0)}\) are some noisy signals for the true value of the parameter \(\theta\), then \(\pi_{i}\) can be a sufficient statistic for the variance of \(i\)'s signal-generating distribution. Indeed, assume that each agent \(i\) receives a noisy signal \(s_{i}\) on the state of the world \(\theta\), which is given by: \(s_{i} = \theta + \varepsilon_{i}\), with \(\varepsilon_{i} \sim \mathcal{N}(0, \sigma_{i}^{2})\) and \(\varepsilon_{i} \perp \varepsilon_{j}\), \(\forall (i, j) \in N^{2}\). In that case, the initial belief is: \(b^{(0)}_{i} = s_{i}\) and the precision is \(\pi_{i} = 1/\sigma_{i}^{2}\). The initial precisions of all agents can be stacked into a vector \(\pi \in \mathbb{R}_{+}\). It will be assumed that \(\pi_{i} > 0\) for at least some agent \(i \in N\) in order for the communication process described below to be meaningful. In this framework, we can assume that:

\[
p_{ij} = \frac{g_{ij} \pi_{j}}{\sum_{k=1}^{n} g_{ik} \pi_{k}}
\]

(22)

In this interpretation, the DeGroot model can be thought as a boundedly rational version of this Bayesian process, where the agents do not adjust their weightings over time.

The corresponding matrix \(P = (p_{ij})\) is the interaction matrix of relative weights. It is a row stochastic matrix, so that its entries across each row sum to 1. Observe that the adjacency matrix \(G = (g_{ij})\) of the network is a \((0-1)\) matrix while the interaction matrix \(P = (p_{ij})\), corresponding to \(G\), is such that each element \(p_{ij}\) is between 0 and 1. In other words, \(P\) is the row-normalized version of \(G\). As a result, a network can be defined by either \(G\) or \(P\). At each period \(t \in \{0, 1, 2, \ldots\}\), agents revise their beliefs to a weighted average of the previous-period beliefs of their neighbors so that

\[
b^{(t)}_{i} = \sum_{j=1}^{n} p_{ij} b^{(t-1)}_{j}
\]

(23)

or, in matrix form

\[
b^{(t)} = P b^{(t-1)}
\]

(24)

The limiting beliefs can be calculated as a function of the initial beliefs and weights. They
are given by:
\[ b^\infty = \lim_{t \to \infty} P^t b^{(0)} \] (25)
where \( P^t \) is the matrix of cumulative influences in period \( t \).

**Definition 2** A matrix \( P \) is convergent if \( \lim_{t \to \infty} P^t b \) exists for all vectors \( b \in [0, 1]^n \).

This definition of convergence requires that beliefs converge for all initial vectors of beliefs. Indeed, if convergence fails for some initial vector, then there will be oscillations or cycles in the updating of beliefs and convergence will fail.

**Definition 3** Agents in network \( G \) reach a consensus if for any \( b(0) \in \mathbb{R}^n \), we have:
\[ \lim_{t \to +\infty} [b_i(t) - b_j(t)] = 0 \quad \text{for all } (i, j) \in \mathbb{N}^2 \]

We have the following main result, which states under which condition the updating process described below converge to a well-defined limit:

**Proposition 8 (DeGroot (1974))** Assume that network \( G \) is strongly connected and agents follow the average-based updating process described by expression (24). Then all agents in \( G \) reach a consensus, which, for each agent \( i \in N \), is given in the limit by:
\[ \left( \lim_{t \to +\infty} P^t b \right)_i = e b \quad \text{for every } b \in [0, 1]^n \]
where \( e \) is the unique left eigenvector of matrix \( P \).

The convergence result comes from standard Markov chain theory. Indeed, the matrix \( P \) is irreducible because the associated network \( G \) is assumed to be strongly connected. Moreover, the matrix \( P \) is aperiodic because every agent listens to him or herself, i.e. \( g_{ii} > 0 \), \( \forall i \). This guarantees that the process converges to a unique steady state because \( P \) is ergodic. Finally, since the matrix \( P \) is row stochastic, its largest eigenvalue is 1, and therefore, there is a unique left eigenvector \( e \) with positive components such that \( e = eP \). The eigenvector property is just saying that \( e_i = \sum_{j \in N} p_{ji} e_j \) for all \( i \), so that the opinions of agents with greater influence have a greater weight in the final convergence belief.

**Corollary 9** Assume that the network \( G \) is undirected so that \( g_{ij} = 1 \) means that \( i \) and \( j \) pay attention to each other and define \( p_{ij} = g_{ij} / d_i(G) \), where \( d_i(G) = \sum_{j \in N} g_{ij} \) is the degree
of i. Then, under the same assumptions as in Proposition 8, the same convergence result holds where $e_i$ is now given by:

$$e_i = \frac{d_i(G)}{\sum_{j \in N} d_j(G)}$$

This is a particular case of Proposition 8 where we assume that each agent equally split attention among his or her neighbors in an undirected network. In that case, Corollary 9 shows that social influence is proportional to the agent's degree. An interesting feature of Corollary 9 is that social influence is only determined by the degree distribution of the network and not by any structural properties of the network such as the average path length or the centrality of the network. However, Golub and Jackson (2010) shows that the speed of convergence is affected by a structural property of the network, namely homophily. They show that homophily does not alter agents' social influence and therefore has no effect on the long-term learning but significantly reduces the speed of convergence.

**An example of convergence with the DeGroot model** To illustrate the notations and the results in Proposition 8, let us consider the following network:

![Directed network](image)

**Figure A1: Directed network**
where the adjacency matrix is given by:

\[
G = \begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 0 \\
1 & 1 & 1
\end{pmatrix}
\]

This means, for example, that agent 2 pays attention to agent 3 but agent 3 does not pay attention to agent 2. It is easily verified that this directed network is strongly connected. The weights are arbitrarily determined and given by

\[
P = \begin{pmatrix}
1 & 1/3 & 1/3 \\
1/2 & 1/2 & 0 \\
0 & 1/2 & 1/2
\end{pmatrix}
\]

This means, for example, that all agents put equal weight of all their neighbors. Assuming that the initial beliefs are given by: \( b^{(0)} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T \), it is easily verified that:

\[
b^{(1)} = P b^{(0)} = \begin{pmatrix}
1/3 & 1/3 & 1/3 \\
1/2 & 1/2 & 0 \\
0 & 1/2 & 1/2
\end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}
\]

\[
b^{(2)} = P b^{(1)} = \begin{pmatrix}
1/3 & 1/3 & 1/3 \\
1/2 & 1/2 & 0 \\
0 & 1/2 & 1/2
\end{pmatrix} \begin{pmatrix} 1/3 \\ 1/2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.278 \\ 0.417 \\ 0.25 \end{pmatrix}
\]

By continuing iterating, there is convergence to the following consensus:

\[
b^\infty = \lim_{t \to \infty} P^t b^{(0)} = \begin{pmatrix}
1/3 = 0.333 \\
4/9 = 0.444 \\
2/9 = 0.222
\end{pmatrix}
\]

(26)

This means that

\[
P^t = \begin{pmatrix}
1/3 & 4/9 & 2/9 \\
1/3 & 4/9 & 2/9 \\
1/3 & 4/9 & 2/9
\end{pmatrix}
\]

which implies that a consensus is reached. In other words, no matter what are the initial beliefs \( b^{(0)} \), all agents end up with limiting beliefs corresponding to the entries of \( b^\infty = \)
\[ \lim_{t \to \infty} P^t b^{(0)} \] where
\[ b_1^\infty = b_2^\infty = b_3^\infty = \frac{1}{3} b_1^{(0)} + \frac{4}{9} b_2^{(0)} + \frac{2}{9} b_3^{(0)} \] (27)

This example shows that the beliefs converge over time and that agents reach a consensus but it also shows that agent 2 is the most influential individual in the network over the limiting beliefs. Since \( b_1^{(0)} = 1 \), \( b_2^{(0)} = 0 \) and \( b_3^{(0)} = 0 \), we have:
\[ b_1^\infty = b_2^\infty = b_3^\infty = \frac{1}{3} = 0.333 \]

If there are two states of the world and if the consensus is on the state \{It is worth continuing to study\}, then this means that all students in this network agree that with 33.3 percent that it is worth continuing to study.
Appendix E: Network Structure Indicators

Let $N$ be a set of nodes with cardinality $n$. Let $G$ be the adjacency matrix, whose generic element $g_{ij}$ is equal to one if an edge (link) from $j$ to $i$ exists (here we consider indirect networks, so $g_{ij} = g_{ji}$). We consider the following network structure measures (Wasserman and Faust, 1994).

**Density**

$$D_s(g) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij}}{n(n-1)}.$$

**Betweenness centrality**

Let $\delta_{jk}$ be the number of shortest paths between node $j$ and node $k$ and $\delta_{jk}^i$ be the number of shortest paths between node $j$ and node $k$ through $i$. For each node, betweenness centrality is:

$$B_i = \frac{1}{(n-1)(n-2)} \sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\delta_{jk}^i}{\delta_{jk}}.$$

It assumes values in $[0, 1]$. At the network level:

$$B(g) = \frac{\sum_{i=1}^{n} |B_i^* - B_i|}{n-1}$$

where $B_i^*$ is the maximum value of betweenness centrality among the nodes. This index is equal to one when a node has centrality equal to 1 and all others have zero centrality.

**Closeness centrality**

Let $d(i, j)$ be the shortest path between two nodes. For each node, closeness centrality is:

$$C_{2i} = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d(i, j)}.$$

At the network level:

$$C_2(g) = \frac{\sum_{i=1}^{n} C_{2i}}{n}.$$
Assortativity

Assortativity measures the correlation pattern in the degree distribution. If highly degree connected nodes are often linked with similar ones, it shows a positive sign. Let \( d_i \) be \( i \)'s number of links and \( m = \frac{1}{n} \sum d_i \) the average number of links among nodes. Assortativity is defined as:

\[
A(g) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (d_i - m)(d_j - m)g_{ij}}{\sum_{i=1}^{n} (d_i - m)^2}.
\]

Clustering coefficient

For all \( i \) such that \( i \in N' := \{i \in N | n_i(g) \geq 2 \} \), where \( n_i(g) \) is the cardinality of \( N_i(g) \) and \( N_i(g) \) is the set of direct links of node \( i \), the clustering coefficient is:

\[
Cl_i = \frac{\sum_{i \in N_i(g)} \sum_{k \in N_i(g)} g_{ik}}{n_i(g)[n_i(g) - 1]}
\]

For the other nodes (singleton and with only one link), the value is imposed to be equal to zero. At the network level:

\[
C(g) = \sum_{i \in N'} \frac{n_i(g)[n_i(g) - 1]}{\sum_{j \in N'} n_j(g)[n_j(g) - 1]} Cl_i.
\]

This is a simple weighted mean of the clustering coefficients in which each node has a weight proportional to the number of possible connections among its direct links.
Appendix F: Bayesian Estimation

Prior and Posteriors Distributions  In order to draw random values from the marginal posterior distributions of parameters we need to set prior distributions of those parameters. Once priors and likelihoods are specified, we can derive marginal posterior distributions of parameters and draw values from them. Given the link formation, the probability of observing the short- and long-lasting networks, $G^L_r$ and $G^S_r$ is

$$P(G^L_r|x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}) = \prod_{i\neq j} P(g^L_{ij,t-1}|x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}),$$

$$P(G^S_r|x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}) = \prod_{i\neq j} P(g^S_{ij,t}|x_{i,j}, x_{j,r}, z_{i,r}, z_{j,r}).$$

Following Hsieh and Lee (2015) our prior distributions are

$$z_{i,r} \sim N(0, 1)$$

$$\omega \sim N_{2K+3}(\omega_0, \Omega_0)$$

$$\phi^L \sim U[-\kappa_L, \kappa_L]$$

$$\phi^S \sim U[-\kappa_S, \kappa_S]$$

$$\beta^* \sim N_{3K+1}(\beta_0, B_0)$$

$$(\sigma^2, \sigma_{\delta z}) \sim TN_2(\sigma_0, \Sigma_0)$$

$$\eta_r|\sigma_\eta \sim N(0, \sigma_\eta)$$

$$\sigma_\eta \sim IG(S_0, \Xi_0)$$

where $\omega = (\delta^L, \theta^L, \delta^S, \theta^S)$, $\kappa_L = \frac{1}{\kappa} - |\phi^L|$, $\kappa_S = \frac{1}{\kappa} - |\phi^S|$ and $\kappa = 1/\max(\min(\sum_{i} g^S_i), \max_{i}(\sum_{j} g^S_{ij}))$ from Gershgorin Theorem, $U[\cdot]$, $TN_2(\cdot)$ and $IG(\cdot)$ are respectively the uniform, bivariate truncated normal, and inverse gamma distributions. Those distributions depend on hyper-parameters (like $\beta_0$) that are set by the
econometrician. It follows that the marginal posteriors are

\[
P(Z_r | Y_r, G_r^S, G_r^L, \rho) \propto \prod_{r=1}^{n_r} \prod_i^n \phi(z_{i,r}) P(Y_r, G_r^S, G_r^L | Z_r, \rho)
\]

\[
P(\omega | Y_r, G_r^S, G_r^L) \propto \phi^{2K+3}(\omega, \omega_0, \Omega_0) \prod_{r=1}^{n_r} P(G_r^S, G_r^L | Z_r, \omega)
\]

\[
P(\phi^S, \phi^S | Y_r, G_r^S, G_r^L, Z_r, \beta, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) \propto \prod_{r=1}^{n_r} P(Y_r | G_r^S, G_r^L, Z_r, \beta, \sigma_\varepsilon^2, \sigma_{\varepsilon z})
\]

\[
P(\beta^* | Y_r, G_r^S, G_r^L, Z_r, \phi^S, \phi^L) \propto \phi^2((\sigma_\varepsilon^2, \sigma_{\varepsilon z}), \sigma_0, \Sigma_0) \prod_{r=1}^{n_r} P(Y_r | G_r^L, G_r^S, Z_r, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta)
\]

\[
P(\eta_r | Y_r, G_r^L, G_r^S, Z_r, \phi^S, \phi^L, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta) \propto \phi(\eta_r, \tilde{\eta}_r, \tilde{M}_r)
\]

\[
P(\sigma_\eta | Y_r, G_r^L, G_r^S, Z_r, \phi^S, \phi^L, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) \propto \nu(\frac{\zeta_0 + \tau}{2}, \frac{\zeta_0 + \sum_{r=1}^{n_r} \eta_r^2}{2})
\]

where \(\rho = (\omega, \phi^S, \phi^L, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta)\), \(\phi^i(\cdot)\) is the multivariate \(i\)-dimensional normal density function, \(\phi^*_i(\cdot)\) is the truncated counterpart, \(\nu(\cdot)\) is the inverse gamma density function.

**Sampling Algorithm** We start our algorithm by picking \((\omega^{(1)}, \phi^{L(1)}, \phi^{S(1)}, \beta^{(1)}, \sigma_{\varepsilon}^{(1)}, \sigma_{\varepsilon z}^{(1)}, \sigma_{\eta}^{(1)})\) as starting values. For \(\beta^{(1)}, \eta^{(1)}, \phi^{L(1)}, \phi^{S(1)}\) we use OLS estimates, while we set the variances-covariances \(\sigma_{\varepsilon}^{(1)}, \sigma_{\varepsilon z}^{(1)}, \sigma_{\eta}^{(1)}\) at 0.38. We ought to draw samples of \(z_{i,r}^t\) from \(P(z_{i,r} | Y_r, G_r^L, G_r^S, \rho)\), \(i = 1, \ldots, n\). To do this, we first draw a candidate \(z_{i,r}^t\) from a normal distribution with mean \(z_{i,r}^{(t-1)}\), then we rely on a M-H decision rule: if \(z_{i,r}^t\) is accepted we set \(z_{i,r}^t = z_{i,r}^t\), otherwise \(z_{i,r}^t = z_{i,r}^{(t-1)}\). Once all \(z_{i,r}\) are sampled, we move to the sampling of \(\beta^*\). By specifying a normal prior and a normal likelihood we can now easily sample \(\beta^*\) from a multivariate normal distribution. A diffuse prior for \(\sigma_{\varepsilon}^2\) allows us to sample it from an inverse chi-squared distribution. We follow the Bayesian spatial econometric literature by

---

The algorithm is robust to different starting values. However, speed of convergence may increase significantly.
sampling $\phi^L, \phi^S$ from uniform distributions with support $[-\kappa_L, \kappa_L]$ and $[-\kappa_S, \kappa_S]$, as defined above. A M-H step is then performed over a normal likelihood: if accepted, then $\phi^{st} = \tilde{\phi}^{st}$ and $\phi^{lt} = \tilde{\phi}^{lt}$. For network fixed effects we deal again with normal prior and normal likelihood, so $\eta$ is easily sampled from a multivariate normal. We sample $\sigma^2, \sigma_{\varepsilon \varepsilon}$ from a truncated bivariate normal over an admissible region $\Xi$ such that the variance-covariance matrix is positive definite. Acceptance or rejection is determined by the usual M-H decision rule. At each of the M-H steps, the algorithm accepts the new values if the likelihood is higher than the current one.
Figure 1. Misspecification of long-lived ties and short-lived ties

Numerical simulation

Bias corrected 2SLS lagged estimates of long-lived tie effect ($\phi^L$)

Bias corrected 2SLS lagged estimates of short-lived tie effect ($\phi^S$)

Notes: For each combination of replacement rates, we plot the average estimate of peer effects. Standard errors are derived assuming drawing independence and accounting for both within and between sample variation.
Figure 2. Simulation experiment
Single replacement effect

Bias corrected 2SLS lagged estimates of long-lived tie effect ($\phi^L$)

Bias corrected 2SLS lagged estimates of short-lived tie effect ($\phi^S$)

Notes: For each replacement rate, we plot the average estimate of peer effects. Standard errors are derived assuming drawing independence and accounting for both within and between sample variation.
Figure 3. Rejection rates of null hypothesis
Comparing estimators

Notes: the rejection rate of null hypothesis for long-lived and short-lived tie effects.
<table>
<thead>
<tr>
<th>Dep.Var.</th>
<th>Years of Education</th>
<th>2SLS Lagged</th>
<th>2SLS</th>
<th>Many IV</th>
<th>Bias Corrected</th>
<th>Finite IV</th>
<th>Many IV</th>
<th>Bias Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer effects (φ)</td>
<td>0.0057 ***</td>
<td>0.0052 ***</td>
<td>0.0052 ***</td>
<td>0.0064 ***</td>
<td>0.0058 ***</td>
<td>0.0059 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0020)</td>
<td>(0.0019)</td>
<td>(0.0019)</td>
<td>(0.0026)</td>
<td>(0.0020)</td>
<td>(0.0020)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.9702 ***</td>
<td>0.9718 ***</td>
<td>0.9718 ***</td>
<td>0.7383</td>
<td>1.0433 ***</td>
<td>1.0434 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2004)</td>
<td>(0.2004)</td>
<td>(0.2004)</td>
<td>(0.6123)</td>
<td>(0.2408)</td>
<td>(0.2408)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>-0.1346</td>
<td>-0.1341</td>
<td>-0.1341</td>
<td>-0.2933</td>
<td>-0.1833</td>
<td>-0.1830</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4088)</td>
<td>(0.4088)</td>
<td>(0.4088)</td>
<td>(0.7722)</td>
<td>(0.4464)</td>
<td>(0.4464)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other races</td>
<td>-0.3445</td>
<td>-0.3447</td>
<td>-0.3447</td>
<td>-0.3873</td>
<td>-0.2520</td>
<td>-0.2627</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2912)</td>
<td>(0.2911)</td>
<td>(0.2911)</td>
<td>(0.4561)</td>
<td>(0.3111)</td>
<td>(0.3111)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religion Practice</td>
<td>0.2515 ***</td>
<td>0.2521 ***</td>
<td>0.2521 ***</td>
<td>0.2515 ***</td>
<td>0.2521 ***</td>
<td>0.2521 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0540)</td>
<td>(0.0540)</td>
<td>(0.0540)</td>
<td>(0.1373)</td>
<td>(0.0599)</td>
<td>(0.0599)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Size</td>
<td>0.0325</td>
<td>0.0327</td>
<td>0.0327</td>
<td>0.0247</td>
<td>0.0355</td>
<td>0.0355</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0574)</td>
<td>(0.0574)</td>
<td>(0.0574)</td>
<td>(0.0900)</td>
<td>(0.0605)</td>
<td>(0.0605)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent education</td>
<td>0.2890 ***</td>
<td>0.2895 ***</td>
<td>0.2895 ***</td>
<td>0.2275 *</td>
<td>0.2962 ***</td>
<td>0.2862 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0969)</td>
<td>(0.0969)</td>
<td>(0.0969)</td>
<td>(0.1528)</td>
<td>(0.1051)</td>
<td>(0.1051)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics score A</td>
<td>1.3550 ***</td>
<td>1.3578 ***</td>
<td>1.3578 ***</td>
<td>0.8917 *</td>
<td>1.2388 ***</td>
<td>1.2385 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2551)</td>
<td>(0.2551)</td>
<td>(0.2551)</td>
<td>(0.5065)</td>
<td>(0.2816)</td>
<td>(0.2816)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics score B</td>
<td>0.9691 ***</td>
<td>0.9683 ***</td>
<td>0.9683 ***</td>
<td>0.6889 *</td>
<td>0.8941 ***</td>
<td>0.8941 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2399)</td>
<td>(0.2399)</td>
<td>(0.2399)</td>
<td>(0.3966)</td>
<td>(0.2532)</td>
<td>(0.2532)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics score C</td>
<td>0.5254 ***</td>
<td>0.5272 ***</td>
<td>0.5272 ***</td>
<td>0.3422</td>
<td>0.510 **</td>
<td>0.510 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2583)</td>
<td>(0.2583)</td>
<td>(0.2583)</td>
<td>(0.4019)</td>
<td>(0.2709)</td>
<td>(0.2709)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics score missing</td>
<td>0.6214</td>
<td>0.6210</td>
<td>0.6210</td>
<td>1.1487</td>
<td>0.6678</td>
<td>0.6677 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4286)</td>
<td>(0.4286)</td>
<td>(0.4286)</td>
<td>(0.8355)</td>
<td>(0.4471)</td>
<td>(0.4471)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resid. building qual.</td>
<td>0.1771 **</td>
<td>0.1798 **</td>
<td>0.1797 **</td>
<td>0.0950 **</td>
<td>0.1264 **</td>
<td>0.1262 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0966)</td>
<td>(0.0966)</td>
<td>(0.0966)</td>
<td>(0.1069)</td>
<td>(0.1053)</td>
<td>(0.1053)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student grade</td>
<td>0.4781 ***</td>
<td>0.4766 ***</td>
<td>0.4766 ***</td>
<td>0.4051 ***</td>
<td>0.5042 ***</td>
<td>0.5043 ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0851)</td>
<td>(0.0850)</td>
<td>(0.0850)</td>
<td>(0.1674)</td>
<td>(0.0914)</td>
<td>(0.0914)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children</td>
<td>-0.4171 ***</td>
<td>-0.4157 ***</td>
<td>-0.4157 ***</td>
<td>-1.5107</td>
<td>-1.2429 **</td>
<td>-1.2438 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1688)</td>
<td>(0.1668)</td>
<td>(0.1668)</td>
<td>(1.7631)</td>
<td>(0.6263)</td>
<td>(0.6263)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religion Practice (Wave 4)</td>
<td>0.1511 ***</td>
<td>0.1514 ***</td>
<td>0.1514 ***</td>
<td>1.0470</td>
<td>0.3429 **</td>
<td>0.3427 **</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0514)</td>
<td>(0.0514)</td>
<td>(0.0514)</td>
<td>(1.1264)</td>
<td>(0.1873)</td>
<td>(0.1873)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>-0.1168</td>
<td>-0.1166</td>
<td>-0.1166</td>
<td>-0.8750</td>
<td>-0.6920</td>
<td>-0.6924</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1655)</td>
<td>(0.1655)</td>
<td>(0.1655)</td>
<td>(2.113)</td>
<td>(0.6381)</td>
<td>(0.6381)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parental occupation dummies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contextual effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First stage F statistic</td>
<td>1210.64</td>
<td>700.35</td>
<td>1240.51</td>
<td>708.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR test p-value</td>
<td>0.503</td>
<td>0.461</td>
<td>0.554</td>
<td>0.503</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1819</td>
<td>1819</td>
<td>1819</td>
<td>1819</td>
<td>1819</td>
<td>1819</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Networks</td>
<td>116</td>
<td>116</td>
<td>116</td>
<td>116</td>
<td>116</td>
<td>116</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
### Table 2: 2SLS Lagged First Stage Results

<table>
<thead>
<tr>
<th>Variables: X</th>
<th>X</th>
<th>GX</th>
<th>G²X</th>
<th>G³X</th>
<th>G⁴X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.4516</td>
<td>0.7482**</td>
<td>-0.2222</td>
<td>-0.0209</td>
<td>0.0084</td>
</tr>
<tr>
<td>(0.4456)</td>
<td>(0.3483)</td>
<td>(0.1738)</td>
<td>(0.0422)</td>
<td>(0.0110)</td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>-1.1913*</td>
<td>-0.3352</td>
<td>0.1265</td>
<td>0.0248</td>
<td>-0.0061</td>
</tr>
<tr>
<td>(0.6487)</td>
<td>(0.4546)</td>
<td>(0.1264)</td>
<td>(0.0570)</td>
<td>(0.0071)</td>
<td></td>
</tr>
<tr>
<td>Other races</td>
<td>0.1140</td>
<td>0.7957*</td>
<td>0.0356</td>
<td>-0.1779***</td>
<td>0.0176</td>
</tr>
<tr>
<td>(0.6163)</td>
<td>(0.4818)</td>
<td>(0.2197)</td>
<td>(0.0611)</td>
<td>(0.0162)</td>
<td></td>
</tr>
<tr>
<td>Religion Practice</td>
<td>-0.1296</td>
<td>-0.0112</td>
<td>-0.0067</td>
<td>-0.0065</td>
<td>0.0012</td>
</tr>
<tr>
<td>(0.1922)</td>
<td>(0.0962)</td>
<td>(0.0431)</td>
<td>(0.0101)</td>
<td>(0.0024)</td>
<td></td>
</tr>
<tr>
<td>Household Size</td>
<td>-0.2091</td>
<td>0.1924**</td>
<td>0.0392</td>
<td>-0.0117</td>
<td>0.0013</td>
</tr>
<tr>
<td>(0.1355)</td>
<td>(0.0948)</td>
<td>(0.0458)</td>
<td>(0.0124)</td>
<td>(0.0029)</td>
<td></td>
</tr>
<tr>
<td>Parent education</td>
<td>-0.3902*</td>
<td>0.6201***</td>
<td>0.1962***</td>
<td>0.0218</td>
<td>-0.0062</td>
</tr>
<tr>
<td>(0.1966)</td>
<td>(0.1520)</td>
<td>(0.0625)</td>
<td>(0.0200)</td>
<td>(0.0043)</td>
<td></td>
</tr>
<tr>
<td>Mathematics score A</td>
<td>-1.5474*</td>
<td>2.7857***</td>
<td>1.2068***</td>
<td>0.0480</td>
<td>-0.0044***</td>
</tr>
<tr>
<td>(0.5663)</td>
<td>(0.4119)</td>
<td>(0.1749)</td>
<td>(0.0463)</td>
<td>(0.0093)</td>
<td></td>
</tr>
<tr>
<td>Mathematics score B</td>
<td>-1.9044*</td>
<td>2.4911***</td>
<td>0.9199***</td>
<td>-0.9298</td>
<td>-0.0219***</td>
</tr>
<tr>
<td>(0.5545)</td>
<td>(0.3843)</td>
<td>(0.1756)</td>
<td>(0.0461)</td>
<td>(0.0090)</td>
<td></td>
</tr>
<tr>
<td>Mathematics score missing</td>
<td>-1.0947*</td>
<td>1.8637***</td>
<td>0.9099***</td>
<td>0.0658</td>
<td>-0.0282***</td>
</tr>
<tr>
<td>(0.5913)</td>
<td>(0.4270)</td>
<td>(0.2005)</td>
<td>(0.0479)</td>
<td>(0.0114)</td>
<td></td>
</tr>
<tr>
<td>Resid. building qual.</td>
<td>-0.0857</td>
<td>1.8655***</td>
<td>0.4239</td>
<td>-1.1896**</td>
<td>0.0976</td>
</tr>
<tr>
<td>(0.9780)</td>
<td>(0.7666)</td>
<td>(0.3482)</td>
<td>(0.0891)</td>
<td>(0.0198)</td>
<td></td>
</tr>
<tr>
<td>Student grade</td>
<td>-0.0623</td>
<td>-0.2335</td>
<td>-0.0974</td>
<td>-0.0019</td>
<td>0.0058</td>
</tr>
<tr>
<td>(0.1558)</td>
<td>(0.1545)</td>
<td>(0.0650)</td>
<td>(0.0204)</td>
<td>(0.0042)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: OLS estimation results, standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The instrumental set also includes the individual number of connections. See Appendix B for further details on IV estimation of spatial models.

### Table 3: Long and short-lived ties: Long-run effects

<table>
<thead>
<tr>
<th>Dep.Var.</th>
<th>Years of Education</th>
<th>2SLS</th>
<th>2SLS Lagged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-lived ties (φ^L)</td>
<td>0.0317***</td>
<td>0.0345***</td>
<td></td>
</tr>
<tr>
<td>(0.0137)</td>
<td>(0.0155)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-lived ties (φ^S)</td>
<td>0.0062</td>
<td>0.0080</td>
<td></td>
</tr>
<tr>
<td>(0.0055)</td>
<td>(0.0063)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: We report bias-corrected 2SLS estimates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Table 4: Short-lived ties in different grades (10 - 12)

<table>
<thead>
<tr>
<th>Dep.Var.</th>
<th>Years of Education</th>
<th>2SLS</th>
<th>2SLS Lagged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-lived ties ($\phi^L$)</td>
<td></td>
<td>0.0419***</td>
<td>0.0485**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0187)</td>
<td>(0.0212)</td>
</tr>
<tr>
<td>Short-lived in lower grades ($\phi^S_1$)</td>
<td></td>
<td>0.0015</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0237)</td>
<td>(0.0331)</td>
</tr>
<tr>
<td>Short-lived in higher grades ($\phi^S_2$)</td>
<td></td>
<td>0.0011</td>
<td>0.0049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0207)</td>
<td>(0.0257)</td>
</tr>
</tbody>
</table>

- Individual socio-demographic: yes
- Family Background: yes
- Protective Factors: yes
- Residential neighborhood: yes
- Contextual Effects: yes
- Network Fixed Effects: yes
- Observations: 628
- Networks: 41

Notes: We report bias-corrected 2SLS estimates. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5: Peer characteristics

<table>
<thead>
<tr>
<th></th>
<th>Wave I</th>
<th>Wave II</th>
<th>Wave I and Wave II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>std</td>
<td>min</td>
</tr>
<tr>
<td>Years of education</td>
<td>14.96</td>
<td>3.48</td>
<td>9.00</td>
</tr>
<tr>
<td>GPA</td>
<td>2.89</td>
<td>0.87</td>
<td>0.00</td>
</tr>
<tr>
<td>Female</td>
<td>0.48</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0.06</td>
<td>0.24</td>
<td>0.00</td>
</tr>
<tr>
<td>Other races</td>
<td>0.09</td>
<td>0.29</td>
<td>0.00</td>
</tr>
<tr>
<td>Religion Practice</td>
<td>3.98</td>
<td>1.82</td>
<td>1.00</td>
</tr>
<tr>
<td>Household Size</td>
<td>3.34</td>
<td>1.37</td>
<td>1.00</td>
</tr>
<tr>
<td>Two married parents family</td>
<td>0.75</td>
<td>0.44</td>
<td>0.00</td>
</tr>
<tr>
<td>Parent education</td>
<td>3.33</td>
<td>0.90</td>
<td>1.00</td>
</tr>
<tr>
<td>Mathematics score A</td>
<td>0.24</td>
<td>0.42</td>
<td>0.00</td>
</tr>
<tr>
<td>Mathematics score B</td>
<td>0.28</td>
<td>0.45</td>
<td>0.00</td>
</tr>
<tr>
<td>Mathematics score C</td>
<td>0.25</td>
<td>0.43</td>
<td>0.00</td>
</tr>
<tr>
<td>Mathematics score missing</td>
<td>0.08</td>
<td>0.27</td>
<td>0.00</td>
</tr>
<tr>
<td>Residential building quality</td>
<td>1.56</td>
<td>0.78</td>
<td>1.00</td>
</tr>
<tr>
<td>Student grade</td>
<td>10.44</td>
<td>0.50</td>
<td>10.00</td>
</tr>
<tr>
<td>Children</td>
<td>0.48</td>
<td>0.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Religion Practice (Wave 4)</td>
<td>1.44</td>
<td>1.62</td>
<td>0.00</td>
</tr>
<tr>
<td>Married</td>
<td>0.50</td>
<td>0.50</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes: Differences between means are never statistical significant at conventional levels of significance.
Table 6: Network formation in Wave I and Wave II
Model (8) OLS estimation results

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>( \gamma ) Coefficient</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.0025</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>-0.0107</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Other races</td>
<td>-0.0265</td>
<td>(0.036)</td>
</tr>
<tr>
<td>Student grade</td>
<td>-0.0226</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Religion Practice</td>
<td>0.0032</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Mathematics score A</td>
<td>-0.0063</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Mathematics score B</td>
<td>0.0178</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Mathematics score C</td>
<td>0.0122</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Mathematics score missing</td>
<td>-0.0030</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Parent education</td>
<td>-0.0073</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Household Size</td>
<td>0.0125</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Parent occupation professional/technical</td>
<td>0.0103</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Parent occupation manual</td>
<td>0.0028</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Parent occupation office or sales worker</td>
<td>0.0162</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Parent occupation other</td>
<td>0.0460**</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Two married parents family</td>
<td>-0.0026</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Residential building quality</td>
<td>0.0085</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Network fixed effects</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Chow test p value</td>
<td>0.6083</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>6,932</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Wave I and Wave II network structure

<table>
<thead>
<tr>
<th></th>
<th>Wave I</th>
<th>Wave II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network structure indicators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>0.0010</td>
<td>0.0007</td>
</tr>
<tr>
<td>Betweeness</td>
<td>0.0040</td>
<td>0.0055</td>
</tr>
<tr>
<td>Closeness</td>
<td>0.0131</td>
<td>0.0090</td>
</tr>
<tr>
<td>Assortativity</td>
<td>2.4105</td>
<td>2.9041</td>
</tr>
<tr>
<td>Clustering coefficient</td>
<td>0.0784</td>
<td>0.1516</td>
</tr>
</tbody>
</table>

Notes: Network structure indicators are described in Appendix D.
### Table 8: Bayesian estimation

<table>
<thead>
<tr>
<th></th>
<th>(a) homogeneous peer effects</th>
<th>(b) heterogeneous peer effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2SLS</td>
<td>Bayesian estimation</td>
</tr>
<tr>
<td></td>
<td>outcome eq.</td>
<td>link form.</td>
</tr>
<tr>
<td></td>
<td>without link form.</td>
<td>with link form.</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Peer effects ($\phi$)</td>
<td>0.0045***</td>
<td>0.0040**</td>
</tr>
<tr>
<td></td>
<td>(0.0017)</td>
<td>(0.0016)</td>
</tr>
<tr>
<td>Long-lived ties ($\phi^L$)</td>
<td>0.0097**</td>
<td>0.0087**</td>
</tr>
<tr>
<td></td>
<td>(0.0042)</td>
<td>(0.0044)</td>
</tr>
<tr>
<td>Short-lived ties ($\phi^S$)</td>
<td>0.0020</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>(0.0021)</td>
<td>(0.0021)</td>
</tr>
<tr>
<td>Unobservables</td>
<td>0.2303</td>
<td>-0.1524</td>
</tr>
<tr>
<td>(\sigma_{\epsilon z})</td>
<td>(0.1847)</td>
<td>(0.0933)</td>
</tr>
<tr>
<td>(\theta)</td>
<td>-3.6833***</td>
<td>0.7942***</td>
</tr>
<tr>
<td>(\theta_{1-1=1})</td>
<td>(1.4227)</td>
<td>(0.1262)</td>
</tr>
<tr>
<td>(\theta_{1=2})</td>
<td>0.5214***</td>
<td>(0.1168)</td>
</tr>
<tr>
<td>Individual socio-demographic</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Family background</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Protective factors</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Residential neighborhood</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Contextual effects</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Network fixed effects</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Observations</td>
<td>932</td>
<td>433846</td>
</tr>
<tr>
<td>Networks</td>
<td>33</td>
<td>33</td>
</tr>
</tbody>
</table>

Notes: Columns (2),(3),(5), and (6) report the means and the standard deviations of the posterior distributions of the parameters. We let our chain run for 200,000 iterations, discarding the first 20,000 iterations. Ergodicity of the Markov Chain is achieved.
<table>
<thead>
<tr>
<th></th>
<th>2SLS</th>
<th>2SLS Lagged</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Long-lived ties ($\phi^L$)</strong></td>
<td>0.0393***</td>
<td>0.0474***</td>
</tr>
<tr>
<td></td>
<td>(0.0160)</td>
<td>(0.0183)</td>
</tr>
<tr>
<td><strong>Short-lived ties ($\phi^S$)</strong></td>
<td>0.0049</td>
<td>0.0052</td>
</tr>
<tr>
<td></td>
<td>(0.0063)</td>
<td>(0.0070)</td>
</tr>
<tr>
<td>Individual socio-demographic</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Family Background</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Protective Factors</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Residential neighborhood</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Contextual Effects</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Network Fixed Effects</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Observations</td>
<td>1819</td>
<td>1819</td>
</tr>
<tr>
<td>Networks</td>
<td>116</td>
<td>116</td>
</tr>
</tbody>
</table>

Notes: We report bias-corrected 2SLS estimates. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Table 10: Long and short-lived ties: Short-run effects

<table>
<thead>
<tr>
<th>Dep.Var. GPA</th>
<th>2SLS</th>
<th>Bayesian estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-lived ties ($\phi^L$)</td>
<td>0.0238*** (0.0097)</td>
<td>0.0253*** (0.0093)</td>
</tr>
<tr>
<td>Short-lived ties ($\phi^S$)</td>
<td>0.0079* (0.0046)</td>
<td>0.0121** (0.0051)</td>
</tr>
<tr>
<td>Unobservables ($\sigma_{\epsilon z}$)</td>
<td>0.0053 (0.1036)</td>
<td></td>
</tr>
</tbody>
</table>

Individual socio-demographic: yes, yes  
Family Background: yes, yes  
Protective Factors: yes, yes  
Residential neighborhood: yes, yes  
Contextual Effects: yes, yes  
Network Fixed Effects: yes, yes  
Observations: 1819, 932  
Networks: 116, 33

Notes: We report bias-corrected 2SLS estimates.  
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.