# DISCUSSION PAPER SERIES

No. 9804

## HETEROGENEOUS PEER EFFECTS IN EDUCATION

Eleonora Patacchini, Edoardo Rainone and Yves Zenou

***LABOUR ECONOMICS***

**Centre for Economic Policy Research**

**www.cepr.org**

**Available online at:** **www.cepr.org/pubs/dps/DP9804.php**

ISSN 0265-8003

# HETEROGENEOUS PEER EFFECTS IN EDUCATION

**Eleonora Patacchini, Syracuse University, EIEF, IZA and CEPR**
**Edoardo Rainone, Banca d'Italia and Sapienza University of Rome**
**Yves Zenou, Stockholm University, IFN, GAINS and CEPR**

Discussion Paper No. 9804
February 2014

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 9804

February 2014

# ABSTRACT

## Heterogeneous Peer Effects in Education*

We develop a network model looking at the role of different types of peers in education. The empirical salience of the model is tested using a very detailed longitudinal dataset of adolescent friendship networks. We find that there are strong and persistent peer effects in education but peers tend to be influential only when their friendships last more than a year and not a shorter period of time. In the short run, however, both types of ties have an impact on current grades.

JEL Classification: C31, D85, I21 and Z13
Keywords: education, efficient 2sls estimation, long-term peer effects, social networks and spatial autoregressive model

Eleonora Patacchini
Syracuse University
426 Eggers Hall
Syracuse, NY 13244-1020
USA

Email: epatacch@maxwell.syr.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=158584

Edoardo Rainone
University of Rome 'La Sapienza'
P.le Aldo Moro 5
00185 Roma
ITALY

Email: edoardo.rainone@uniroma1.it

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=173120

Yves Zenou
Department of Economics
Stockholm University
10691 Stockholm
SWEDEN

Email: yves.zenou@ne.su.se

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=126027

*We thank Orazio Attanasio, two anonymous referees, as well as Francine Blau, Larry Blume, Alfonso Flores-Lagunes, Xiaodong Liu, Lawrence Khan, Guido Kuersteiner, Francesca Molinari, Ingman Prucha, 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.

Submitted 17 December 2013

# 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.[1] 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 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 focus 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.

To be more precise, we extend the network model proposed by Calvó-Armengol et al. (2009) to incorporate heterogeneous friendship relationships. We then test this model using 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, (*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 a temporal interval between friends' nomination and educational outcomes.

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.1919

The empirical counterpart of our theoretical model has the specification of a spatial autoregressive (SAR) model with different spatial weight matrices. The specification and estimation of a traditional spatial autoregressive model with one weight matrix in the social space have been studied by Liu and Lee (2010). In this framework, the different positions of group members as measured by the Bonacich (1987) centrality provide additional information

[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).

for identification and estimation. In this paper, we extend the Liu and Lee (2010) 2SLS approach to a network model with two interaction matrices using appropriate IVs. First, we use two centralities, 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). The asymptotic consistency and efficiency of the proposed estimators are proved.

We use network fixed effects to control for the sorting of individuals into groups along unobserved characteristics. Even with this strategy, there may still be individual-level unobserved characteristics that influence both friendship formation and education decisions. Imbens and Goldsmith-Pinkham (2013) argue that endogeneity of this sort has testable implications. We check such implications in our data and show that, once we condition on our long list of controls and network fixed effects, no signs of network endogeneity are revealed. This suggests that, in our case where networks are quite small, network fixed effects are able to capture any troubling source of unobserved heterogeneity. We also consider possible measurement errors in peer groups using a simulation experiment. Our results are robust to various types of network topology misspecification.

Our results show that there are strong and persistent peer effects in education. In other words, the "quality" of friends (in terms of future educational achievement) from high school has a positive and significant impact on own future education level. 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 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.[2] Given that strong and weak ties are commonly defined in terms of the level of common friends, we prefer to use short-lived and short-lived ties, even though the two notions are relatively close. Interestingly, compared to the literature on the labor market, we

[2] 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.

find the opposite result for education outcomes.[3,4] 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).

We also look at the long versus short-run effects of peers on education. 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.

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. Contrary to these studies, we are able to highlight the mechanisms behind the long-run effects of peers.

The paper unfolds as follows. The theoretical model is developed in Section 2. Our data are described in Section 3, while the estimation and identification strategy is discussed in Section 4. Section 5 collects the empirical evidence and investigates the economic mechanisms behind our peer-effects results. Section 6 shows the robustness of our results with respect to network topology misspecification while Section 7 compares the short-run versus the long-run effects of peers on education. Finally, Section 8 concludes the paper.

## 2 Theoretical framework

Consider a population of $n$ individuals.

[3] 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.

[4] See also Patacchini and Zenou (2008) who find evidence of the strength of weak ties in crime.

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

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

There are two types of relationships in the network, *short-lived* and *long-lived* ties.

Denote the *long-lived* adjacency matrix by $\mathbf{G}^L$ and the *short-lived* adjacency matrix by $\mathbf{G}^S$, with $\mathbf{G}^L + \mathbf{G}^S = \mathbf{G}$, where superscripts $L$ and $S$ denote a long-lived and a short-lived-tie relationship, respectively. To construct the *long-lived-tie* adjacency matrix $\mathbf{G}^L$, we put a 1 in $\mathbf{G}$ only if there is a long-lived-tie relationship between $i$ and $j$, i.e. $g_{ij}^L = 1$ and a 0 otherwise. Similarly, to construct the *short-lived-tie* adjacency matrix $\mathbf{G}^S$, we put a 1 in $\mathbf{G}$ only if there is a short-lived-tie relationship between $i$ and $j$, i.e. $g_{ij}^S = 1$ and a 0 otherwise. This means that the intersection between $i$'s long-lived ties and $i$'s short-lived ties is empty $\forall i$.

To illustrate these matrices, consider the following network:

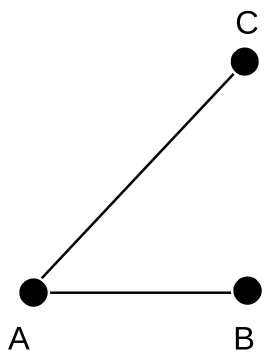



Figure 1: A star network

Its adjacency matrix $\mathbf{G}$ (where individual A corresponds to the first row, individual B to the second and individual C to the third row) is given by:

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

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

$$\mathbf{G}^L = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{G}^S = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Finally, we denote by $N_i^L(g)$ and $N_i^S(g)$ the set of long- and short-lived friends each individual $i$ has in network $g$ and the cardinality of these sets by $d_i^L(g)$ and $d_i^S(g)$, respectively. In the network of Figure 1, we have, for example: $N_A^L(g) = \{C\}$, $N_A^S(g) = \{B\}$, $d_A^L(g) = 1$ and $d_1^S(g) = 1$.

**Preferences** Individuals decide how much effort to exert in education (e.g. how many years to study). We denote the educational effort level of individual $i$ by $y_i$ (i.e. years of schooling) and the population effort profile by $\mathbf{y} = (y_1, ..., y_n)'$. 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 $\phi^L > \phi^S$. Even though this is not directly modeled, following our definition above, a long-lived-tie relationship means that individuals often interact with each other (i.e. they are friends for at least two periods in the AddHealth data) while a short-lived-tie relationship indicates that individuals interact less with each other (i.e. they are friends for at most one period in the AddHealth data). As stated above, friendship relationships are reciprocal so that if $i$ has a long-lived (short-lived) tie relationship with $j$, then $j$ also has a long-lived (short-lived) tie relationship with $i$. Each agent $i$ selects an effort $y_i \geq 0$, and obtains a payoff $u_i(\mathbf{y}, g)$ that depends on the effort profile $\mathbf{y}$ and the underlying network $g$, in the following way:

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

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, 2014) 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 $\phi^L \sum_{j \in N_i^L(g)}^{n} y_i y_j + \phi^S \sum_{j \in N_i^S(g)}^{n} y_i y_j$. 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:

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

where $g_{ij}^L = 1$ ($g_{ij}^S = 1$) 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:

$$a_i = \sum_{m=1}^{M} \beta_m x_i^m + \frac{1}{g_i^L} \sum_{m=1}^{M} \sum_{j=1}^{n^L} g_{ij}^L \, x_j^m \gamma_m^L + \frac{1}{g_i^S} \sum_{m=1}^{M} \sum_{j=1}^{n^S} g_{ij}^S \, x_j^m \gamma_m^S \tag{2}$$

where $x_i^m$ is a set of $M$ variables accounting for observable differences in the individual characteristics of individual $i$, $\beta_m, \gamma_m^L, \gamma_m^S$ 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 (1) is concave in own decisions, and displays decreasing marginal returns in own effort levels. In sum,

$$u_i(\mathbf{y},g) = \underbrace{(a_i + \eta + \varepsilon_i)\ y_i}_{\text{Benefits from own effort}} \underbrace{-\frac{1}{2}y_i^2}_{\text{Costs}} + \underbrace{\phi^L \sum_{j=1}^{n} g_{ij}^L \, y_i y_j}_{\text{Benefits from long-lived ties' effort}} + \underbrace{\phi^S \sum_{j=1}^{n} g_{ij}^S \, y_i y_j}_{\text{Benefits from short-lived ties' effort}}$$

**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 (1) and we obtain the following best-reply function for each $i = 1, ..., n$:

$$y_i = \phi^L \sum_{j=1}^{n} g_{ij}^L y_j + \phi^S \sum_{j=1}^{n} g_{ij}^S y_j + a_i + \eta + \varepsilon_i \tag{3}$$

where $a_i$ is given by (2). Denote $\alpha_i = a_i + \eta + \varepsilon_i$ and the corresponding $(1 \times n)$ vector by $\boldsymbol{\alpha}$. The matrix form equivalent of (3) is:

$$\mathbf{y} = \left(\mathbf{I} - \phi^L \mathbf{G}^L - \phi^S \mathbf{G}^S\right)^{-1} \boldsymbol{\alpha}. \tag{4}$$

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

**Proposition 1** *If $\mu_1\left(\phi^L \mathbf{G}^L + \phi^S \mathbf{G}^S\right) < 1$, the peer effect game with payoffs (1) has a unique interior Nash equilibrium in pure strategies given by (3) or by (4).*

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

**Corollary 1** *Assume that $\mathbf{G}$ is symmetric. A sufficient condition for the Nash equilibrium given by (3) or (4) to exist, to be unique and to be interior is: $\left(\phi^L + \phi^S\right)\mu_1(\mathbf{G}) < 1$.*

This is an interesting result because it connects the adjacency matrix $\mathbf{G}$ 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. $\phi\mu_1(\mathbf{G}) < 1$, where the peer effects were assumed to be the same across all agents. The condition given in Proposition 1, $\mu_1\left(\phi^L \mathbf{G}^L + \phi^S \mathbf{G}^S\right) < 1$, is less restrictive than $\left(\phi^L + \phi^S\right)\mu_1(\mathbf{G}) < 1$ since $\mu_1\left(\phi^L \mathbf{G}^L + \phi^S \mathbf{G}^S\right) \leq \phi^L \mu_1\left(\mathbf{G}^L\right) + \phi^S \mu_1\left(\mathbf{G}^S\right)$ (see the proof of Corollary 1). However, it imposes a restriction on the exact topology of both short-lived and long-lived tie networks while $\left(\phi^L + \phi^S\right)\mu_1(\mathbf{G}) < 1$ allows $\mathbf{G}^L$ and $\mathbf{G}^S$ to be less constrained in terms of the topology of the network but, more importantly, it does not require the matrix $\mathbf{G}$ (and thus $\mathbf{G}^L$ and $\mathbf{G}^S$) to be symmetric.

To illustrate this result, consider the network described in Figure 1. The largest eigenvalue of $\mathbf{G}$ is $\sqrt{2}$ while the largest eigenvalue of $\mathbf{G}^L$ and of $\mathbf{G}^S$ is 1. Thus, the sufficient condition given in Corollary 1 is: $\phi^L + \phi^S < 0.707$. In that case, there exists a unique Nash equilibrium given by:

$$\begin{aligned}
\mathbf{y} &= \left(\mathbf{I} - \phi^L \mathbf{G}^L - \phi^S \mathbf{G}^S\right)^{-1} \boldsymbol{\alpha} \\
&= \frac{1}{\left[1 - \left(\phi^L\right)^2 - \left(\phi^S\right)^2\right]} \begin{pmatrix} \alpha_A + \alpha_B \phi^S + \alpha_C \phi^L \\ \alpha_A \phi^S + \alpha_B \left[1 - \left(\phi^L\right)^2\right] + \alpha_C \phi^S \phi^L \\ \alpha_A \phi^S + \alpha_B \phi^S \phi^L + \alpha_C \left[1 - \left(\phi^S\right)^2\right] \end{pmatrix}.
\end{aligned}$$

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 $\left(\phi^L\right)^2 + \left(\phi^S\right)^2 = 0.68 < 1$), then the condition given in Proposition 1 is satisfied since $\mu_1\left(\phi^L \mathbf{G}^L + \phi^S \mathbf{G}^S\right) = 0.825 < 1$ but the one given in Corollary 1 is not since $\phi^L + \phi^S = 1 > 0.707$. Since $\mu_1\left(\phi^L \mathbf{G}^L + \phi^S \mathbf{G}^S\right) = 0.825 < 1$, we can still derive the equilibrium and obtain:

$$y_A^* = 6.25,\ y_B^* = 4.125 \text{ and } y_C^* = 6.$$

This once more shows the importance of the position in the network and the role of short-lived and long-lived ties in educational outcomes. Long-lived ties influence each other so that their joint efforts increase independently of the network position. Even though this is not directly modeled here, one can interpret a link in the network between two students $i$ and $j$, i.e. $g_{ij} = 1$ as an information exchange about education or even the establishment of a social norm in education. For example, if my friends think it is cool to study and go to college, then I am more likely to do so. This is even more true if these friends have a long-lived-tie relationship with each other. If they only have a short-lived-tie relationship, then the influence on each other is less important and we cannot speak about social norms. This is the main idea that we want to test in this paper.

# 3 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).[5] 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 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* tie or 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 or relationship if they have nominated each other in one wave only (Wave I *or* Wave II).

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.[6] Social contacts (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.[7] In addition, we exclude networks consisting of 2-3 individuals, those with

[5] 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.

[6] More precisely, the Wave IV questionnaire asks about the highest education qualification achieved (distinguishing 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.

[7] The representativeness of the sample is preserved. Summary statistics are available upon request.

more than 400 members and individuals who are not followed in Wave IV.[8] 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 B1 in Appendix B. Appendix B also gives a precise definition of the variables used in our study as well as their descriptive statistics (see Table B1).[9]

# 4 Empirical model and estimation strategy

## 4.1 Empirical model

Let $\bar{r}$ be the total number of networks in the sample, $n_r$ the number of individuals in the $r$th network $g_r$, and $n = \sum_{r=1}^{\bar{r}} n_r$ the total number of sample observations. Let $\boldsymbol{x}_{i,r} = (x_{i,r}^m, \cdots, x_{i,r}^M)'$. Using (2), the econometric model corresponding to the best-reply function (3) of agent $i$ in network $g_r$ can be written as:

$$\begin{aligned} y_{i,r,t+1} &= \phi^L \sum_{j=1}^{n_r} g_{ij,r,t}^L y_{j,r,t+1} + \phi^S \sum_{j=1}^{n_r} g_{ij,r,t}^S y_{j,r,t+1} + \boldsymbol{x}_{i,r}' \delta \\ &+ \frac{1}{g_{i,r,t}^L} \sum_{j=1}^{n_r} g_{ij,r,t}^L x_{j,r,t,t+1}' \gamma^L + \frac{1}{g_{i,r,t}^S} \sum_{j=1}^{n_r} g_{ij,r,t}^S x_{j,r,t,t+1}' \gamma^S + \eta_{r,t} + \epsilon_{i,r,t+1}, \end{aligned} \tag{5}$$

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, $x_{i,r,t,t+1}' = (x_{i,r,t,t+1}^1, \cdots, x_{i,r,t,t+1}^m)'$ 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 $g_{i,r,t}^L = \sum_{j=1}^n g_{ij,r,t}^L$ and $g_{i,r,t}^S = \sum_{j=1}^n g_{ij,r,t}^S$

[8] 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 Calvo-Armengol et al., 2009).

[9] 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.

are the total number of long-lived (long-lived-tie) and short-lived (short-lived-tie) 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 $\mathbf{Y}_r = (y_{1,r,t+1}, \cdots, y_{n_r,r,t+1})'$, $\mathbf{X}_r = (x_{1,r,t,t+1}, \cdots, x_{n_r,r,t,t+1})'$, and $\boldsymbol{\epsilon}_r = (\epsilon_{1,r}, \cdots, \epsilon_{n_r,r})'$. Denote the $n_r \times n_r$ adjacency matrix by $\mathbf{G}_r = [g_{ij,r}]$, the row-normalized of $\mathbf{G}_r$ by $\mathbf{G}_r^*$, and the $n_r$-dimensional vector of ones by $\mathbf{l}_{n_r}$. Let us split the adjacency matrix into two submatrices $\mathbf{G}_r^L$ and $\mathbf{G}_r^S$, which keep trace of long-lived and short-lived ties, respectively. Then, model (5) can be written in matrix form as:

$$\mathbf{Y}_r = \phi^L \mathbf{G}_r^L \mathbf{Y}_r + \phi^S \mathbf{G}_r^S \mathbf{Y}_r + \mathbf{X}_r^* \beta + \eta_r \mathbf{l}_{n_r} + \boldsymbol{\epsilon}_r, \tag{6}$$

where $\mathbf{X}_r^* = (\mathbf{X}_r + \mathbf{G}_r^{*S}\mathbf{X}_r + \mathbf{G}_r^{*W}\mathbf{X}_r)$ and $\beta = (\delta', \gamma^{L\prime}, \gamma^{S\prime})'$.

For a sample with $\bar{r}$ networks, stack up the data by defining $\mathbf{Y} = (\mathbf{Y}_1', \cdots, \mathbf{Y}_{\bar{r}}')'$, $\mathbf{X}^* = (\mathbf{X}_1^{*\prime}, \cdots, \mathbf{X}_{\bar{r}}^{*\prime})'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \cdots, \boldsymbol{\epsilon}_{\bar{r}}')'$, $\mathbf{G} = \mathrm{D}(\mathbf{G}_1, \cdots, \mathbf{G}_{\bar{r}})$, $\mathbf{G}^* = \mathrm{D}(\mathbf{G}_1^*, \cdots, \mathbf{G}_{\bar{r}}^*)$, $\boldsymbol{\iota} = \mathrm{D}(\mathbf{l}_{n_1}, \cdots, \mathbf{l}_{n_{\bar{r}}})$ and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_{\bar{r}})'$, where $\mathrm{D}(\mathbf{A}_1, \cdots, \mathbf{A}_K)$ is a block diagonal matrix in which the diagonal blocks are $n_k \times n_k$ matrices $\mathbf{A}_k$'s. For the entire sample, the model is thus:

$$\mathbf{Y} = \phi^L \mathbf{G}^L \mathbf{Y} + \phi^S \mathbf{G}^S \mathbf{Y} + \mathbf{X}^* \beta + \boldsymbol{\iota} \cdot \boldsymbol{\eta} + \boldsymbol{\epsilon}. \tag{7}$$

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 $\boldsymbol{\eta}$ 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.[10]

## 4.2 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

[10] As an analogy with time series models, the model in (7) 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.

and its mean characteristics. Therefore, it is difficult to differentiate between the effect of 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).[11]

**Correlated effects** While a network approach allows us to distinguish endogenous effects from correlated 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. Consider the network in Figure 2. 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 $j$.

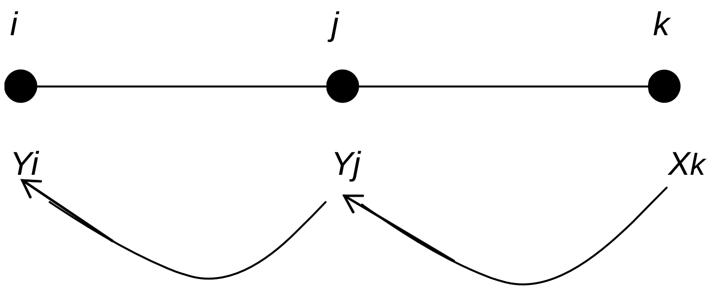



Figure 2: Identification through intransitive triads

[11] Formally, social effects are identified (i.e. no reflection problem) if $\mathbf{G}^2 \neq \mathbf{0}$, where $\mathbf{G}^2$ 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.

**Sorting** In most cases, individuals sort into groups non-randomly. For example, kids whose parents are low educated or worse than average in unmeasured ways would be more likely to sort with low human capital peers. If the variables that drive this process of selection are not fully observable, potential correlations between (unobserved) group-specific factors and the target regressors are major sources of bias. The richness of social network data (where we observe individuals over networks) provides a possible way out by the use of *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 the individuals within each network. This is reasonable in our case study where the networks are quite small (see Section 3).

However, if there are student-level unobservables that drive both network formation and outcome choice, this strategy fails. Recently, Imbens and Goldsmith-Pinkham (2013) highlight the fact that endogeneity of this sort can be tested. Signals of individual-level correlated unobservables would motivate the use of parametric modeling assumptions and Bayesian inferential methods to integrate a network formation with the study of behavior over the formed networks. The next section contains the results which are obtained by applying the approach proposed by Imbens and Goldsmith-Pinkham (2013) to our case.

## 4.3 Endogenous network formation

The basic idea of Imbens and Goldsmith-Pinkham (2013) is that, under homophily, linked individuals are likely to be similar not only in terms of *observed characteristics* but also in terms of *unobserved characteristics* that could influence their behavior.[12] By failing to account for similarities in (unobserved) characteristics, similar behaviors might mistakenly be attributed to peer influence when they simply result due to similar characteristics. In order to highlight the problem, let us write the model (6) as follows:

$$\mathbf{Y}_r = \phi^L \mathbf{G}_r^L \mathbf{Y}_r + \phi^S \mathbf{G}_r^S \mathbf{Y}_r + \mathbf{X}_r^* \beta + \eta_r \mathbf{l}_{n_r} + \underbrace{\boldsymbol{\zeta} \mathbf{v}_r + \mathbf{e}_r}_{\boldsymbol{\epsilon}_r}, \tag{8}$$

where $\mathbf{v}_r = (v_{1,r}, \cdots, v_{n_r,r})'$ denotes a vector of *unobserved characteristics* at the individual level and $\mathbf{e}_r = (e_{1,r}, \cdots, e_{n_r,r})'$ is a vector of random disturbances. Let us consider a network formation model where the variables that explain the short- ($l = S$) or long-lived ($l = L$) tie

[12] This is also true under dissortative matching (i.e. heterophily) but, in that case, there are links if people are dissimilar.

between students $i$ and $j$ belonging to network $r$, i.e. $g_{ij,r}^l$, are the distances between them in terms of observed and unobserved characteristics:

$$g_{ij,r}^l = \alpha^l + \sum_{m=1}^{M} \delta_m^l |x_{i,r}^m - x_{j,r}^m| + \theta^l |v_{i,r} - v_{j,r}| + \eta_r^l + u_{ij,r}^l. \qquad (9)$$

Homophily behavior in the unobserved characteristics implies that $\theta^l < 0$, i.e. the closer two individuals are in terms of unobservable characteristics, the higher is the probability that they are friends. If $\zeta$ is different from zero, then networks $G_r^L$ and $G_r^S$ in model (8) are endogenous.

A testable implication of this problem would be to find a negative correlation between the predicted probability of forming a link (based on observable characteristics), as measured by $\widehat{g_{ij,r}^l}$, and the unobserved similarity in pairs, as measured by the difference in residuals from equation (8), $|\widehat{\epsilon}_{i,r} - \widehat{\epsilon}_{j,r}|$.[13] Evidence against network endogeneity would be to find a zero correlation. This is what we want to test.

For this purpose, we first run a simple logit regression on (9) where the endogenous variable $g_{ij,r}^l$ is a dummy variable that takes a value of 1 if there is a link of type $l = S, L$ between $i$ and $j$ and 0 otherwise. The results are given in Table 1 where we see that there is homophily behavior in observable characteristics since most of the estimated coefficients have a negative sign and are significant. From this regression, we obtain $q_{ij,r}^l = \widehat{g_{ij,r}^l}$, the predicted probability of observing a link between $i$ and $j$.

[*Insert Table* 1 *here*]

Then, if we observe in the data that for two students who are friends (i.e. $g_{ij,r}^l = 1$), there is a low value of $q_{ij,r}^l$, this means that we are not explaining the link formation between students $i$ and $j$ by their observed characteristics but by their unobserved ones. As a result, we should find low values of $q_{ij,r}^l$ associated with low values of $|\widehat{\epsilon}_{i,r} - \widehat{\epsilon}_{j,r}|$, i.e. the friendship relationship between students $i$ and $j$ is explained by similarity in unobserved rather than observed characteristics. A similar argument can be applied for observed non-friend pairs $i$ and $j$ for which $g_{ij,r}^l = 0$. Testing for network endogeneity then implies regressing $q_{ij,r}^l$ on $|\widehat{\epsilon}_{i,r} - \widehat{\epsilon}_{j,r}|$ when a link is observed ($g_{ij,r}^l = 1$) and when it is not ($g_{ij,r}^l = 0$). The results of this regression are reported in the first two rows of each panel of Table 2. In Table 2, we display our results for $g_{ij,r}^l = 1$ (upper panel) for both long-lived ties (panel $a$) and short-lived ties (panel $b$) and for $g_{ij,r}^l = 0$ (lower panel) for both long-lived ties (panel $a$) and short-lived ties

[13] Under dissortative matching (i.e. heterophily), the correlation should be positive.

(panel $b$). We can see that there is no significant correlation between these two variables in all possible regressions both for short-lived and long-lived ties.

We measure *low values* of $q_{ij,r}^{l}$ using the empirical distributions of the predicted probabilities $q_{ij,r}^{l}$. More specifically, we calculate the distribution of $q_{ij,r}^{l}$ for all possible pairs of nodes $i$ and $j$ in network $r$ and consider some threshold values on this distribution. We consider three threshold values: 25%, 35% and 45%. So, for example, in the second column of the upper panel of Table 2, we look at the distribution of $q_{ij,r}^{l}$ whose values are below the 25th percentile (or first quartile) of this distribution. Then, we can compare the observed value of the link $g_{ij,r}^{l}$ (which is equal to 1 or 0) with $q_{ij,r}^{l}$. For example, in the first panel in Table 2, we have $\Pr\left(q_{ij,r}^{L} < 0.45 \mid g_{ij,r}^{L} = 1\right) = 6.06\%$, which means that given that we observe that a pair of nodes $i$ and $j$ has a long-lived link, the probability that the predicted value of $g_{ij,r}^{l}$, i.e. $q_{ij,r}^{L}$, lies below the median is very small (i.e. 6.06%). This conditional probability is even smaller for the 25th percentile since $\Pr\left(q_{ij,r}^{L} < 0.25 \mid g_{ij,r}^{L} = 1\right) = 1.30\%$. We can also calculate the opposite, $\Pr\left(q_{ij,r}^{L} > 0.45 \mid g_{ij,r}^{L} = 1\right) = 93.94\%$, which means that, given that we observe that a pair of nodes has a long-lived-tie link, the probability that the predicted value of $g_{ij,r}^{l}$ lies above the median is very high (i.e. 93.94%).[14]

Our results can be summarized as follows:

($i$) We fail to predict the existence of a link in less than 6% and 11% of the cases, for long-lived and short-lived ties, respectively.

($ii$) In those cases when we fail to predict the existence of a link, we do not find any sign of correlation of the sort discussed above.

($iii$) Those results are robust when considering different thresholds.

Therefore, conditional on the (unusually) large set of controls provided by the AddHealth data, on local effects and on network fixed effects, we find no evidence of endogeneity of the adjacency matrix $\mathbf{G}_{r}^{l}$.

[*Insert Table* 2 *here*]

In order to gain further confidence in our evidence, we also perform an additional exercise. From Table 1, we can see that the variable "residential building quality" is, as expected, important in the link formation process for both types of ties in the sense that students residing in similar residential neighborhoods tend to be friends with each other. Now, suppose that this variable is unobservable to the econometrician, who consequently estimates both link formation and peer effects ignoring it. This fact generates unobservables in the estimation

[14] Similar interpretations can be made when a link is not observed in the data as shown in the lower panels of Table 2.

of both link formation and outcome equations and should make the correlation between $q^l_{ij,r}$ and $|\widehat{\epsilon}_{i,r} - \widehat{\epsilon}_{j,r}|$ positive and significantly different from zero because $\theta^l < 0$ for $l = L, S$ and $\boldsymbol{\zeta}$ >0, when $g^l_{ij} = 1$. As we can see in the last three columns of Table 2, this is exactly what we find when introducing this source of correlation between error terms in equations (9) and (5). Note that in this case:

$(i)$ We fail to predict the existence of a link more frequently than before (roughly twice) under each threshold for both long-lived and short-lived ties.

$(ii)$ The correlation between $q^l_{ij}$ and $|\widehat{\epsilon}_{i,r} - \widehat{\epsilon}_{j,r}|$, when $g^l_{ij} = 1$, is positive and significantly different from zero for both long-lived and short-lived ties for different thresholds.

$(iii)$ Those correlations are decreasing in the threshold values, showing that this correlation is higher for low explained links.

We will thus proceed in our analysis with the assumption of (conditional) network exogeneity.

We also note that most of the traditional problems in the identification of peer effects arises when own and peers' 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.

In this paper, we consider 2SLS estimators (Liu and Lee, 2010) and propose two innovations. First, we use two centralities, 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.

# 5 Estimation results

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

## 5.1 Long-run peer effects

Table 3 collects the estimation results of model (5), without distinguishing between long-lived and short-lived ties. The first three columns show the results when using the 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.

[*Insert Table* 3 *here*]

The results in Table 3 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,[15] 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.

## 5.2 The role of long-lived ties

We would now like to test our model, i.e., Proposition 1, especially equation (3) and its econometric equivalent (5), and thus determine how long- and short-lived ties affect educational choices by estimating the magnitude of $\phi^L$ and $\phi^S$. Table 4 shows the estimation results of model (5).[16] Corollary 1 requires, as a sufficient condition, that $\phi_L + \phi_S$ is in absolute value

[15] The bias-corrected 2SLS estimator is our preferred one since we have relatively small networks (see Appendix C).

[16] 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

smaller than the inverse of the largest eigenvalue of the block-diagonal network matrix $\mathbf{G}_r$, i.e. $\phi_L + \phi_S < 1/\mu_1(\mathbf{G}_r)$. Given that the largest eigenvalue of $\mathbf{G}_r$ is 6.48, the existence of the equilibrium in our model requires values for $\phi_L + \phi_S$ within the range $[0, 0.154)$. Table 4 shows that our results are within this parameter space.

We find that the educational choices of short-lived ties have no significant impact on individual educational outcomes (years of schooling) while the educational choices of long-lived ties 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 ties) translates 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 ties* rather than *short-lived ties* matter for educational outcomes in the long run.[17]

[*Insert Table* 4 *here*]

If we now return to the theoretical model of Section 2, this result means that a long-lived relationship at school between two students (where $\phi = \phi^L$) has an important impact on their future educational choices. One possible interpretation 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 result 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. Our empirical result could confirm this intuition by showing that peers, especially long-lived ties, play an important role in a student's education decision.[18]

alternative estimators in Appendix C remain qualitatively unchanged. The latter are available upon request.

[17] When estimating model (5) including only long-lived ties (i.e. $\mathbf{G}^S = 0$), we obtain comparable results.

[18] 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.

### 5.3 Understanding the mechanisms

Our results suggest that the distinction between long-lived ties and short-lived ties is important for understanding long-run peer effects in education. In our analysis, we identified long-lived ties as peers nominated in both Wave I and Wave II, which means 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 (5), that is

$$\begin{aligned} y_{i,r,t+1} \quad = \quad & \phi^{L}\sum_{j=1}^{n_r} g^{L}_{ij,r,t}y_{j,r,t+1} + \phi^{S_1}\sum_{j=1}^{n_r} g^{S_1}_{ij,r,t}y_{j,r,t+1} + \phi^{S_2}\sum_{j=1}^{n_r} g^{S_2}_{ij,r,t}y_{j,r,t+1} + \frac{1}{g^{L}_{i,r,t}}\sum_{j=1}^{n_r} g^{L}_{ij,r,t}x'_{j,r,t,t+1}\gamma^{L} \\ & + \frac{1}{g^{S_1}_{i,r,t}}\sum_{j=1}^{n_r} g^{S_1}_{ij,r,t}x'_{j,r,t,t+1}\gamma^{S_1} + \boldsymbol{x}'_{i,r}\delta + \frac{1}{g^{S_2}_{i,r,t}}\sum_{j=1}^{n_r} g^{S_2}_{ij,r,t}x'_{j,r,t,t+1}\gamma^{S_2} + \eta_{r,t} + \epsilon_{i,r,t+1}. \end{aligned}$$

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 5 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.

[*Insert Table* 5 *here*]

A last 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 6 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.

[*Insert Table* 6 *here*]

To further investigate this issue, we test whether link formation differs between different waves. For this purpose, we use model (9) and pool the data for Wave I ($t = 1$) and Wave II ($t = 2$):

$$g_{ij,r,t} = \alpha + \sum_{m=1}^{M} \beta_m |x^m_{i,r,t} - x^m_{j,r,t}| + \sum_{m=1}^{M} \gamma^m |x^m_{i,r,t} - x^m_{j,r,t}| \times d_{ij,r} + \epsilon_{ij,r,t}, \quad t = 1, 2. \qquad (10)$$

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), $x^m_{i,r,t}$ 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. Table 7 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.[19] Table 7 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 6 and 7 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.

[*Insert Table* 7 *here*]

Finally, we investigate whether there are any structural differences across Wave I and Wave II in terms of the topology of the network. Over the past years, social network theorists have proposed a number of measures to account for the variability in network location across

[19] 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).

agents (Wasserman and Faust, 1994). We present those indicators in Appendix D 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 8. It appears that the two networks are topologically very similar.

[*Insert Table* 8 *here*]

# 6 Robustness checks

In this section, we perform two different robustness checks.

## 6.1 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.[20] Table 9 shows the estimation results of model (5) for directed networks. The results remain qualitatively unchanged and only slightly higher in magnitude.

[*Insert Table* 9 *here*]

## 6.2 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

[20] 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).

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 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 A

| | | Observed | | | |
|---|---|---|---|---|---|
| | | L | S | N | |
| True | L | $p^2$ | $2p(1-p)$ | $(1-p)^2$ | 1 |
| | S | $q^2$ | $2q(1-q)$ | $(1-q)^2$ | 1 |
| | N | $r^2$ | $2r(1-r)$ | $(1-r)^2$ | 1 |
| | | $s$ | $w$ | $n$ | |

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 B

| | | Observed | | |
|---|---|---|---|---|
| | | L | S | N |
| True | L | $\cdot$ | $P_{SW}$ | $P_{SN}$ |
| | S | $P_{WS}$ | $\cdot$ | $P_{WN}$ |
| | N | $P_{NS}$ | $P_{NW}$ | $\cdot$ |

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 $P_{LS} = \frac{2p(1-p)}{p^2+2p(1-p)}$ and $P_{SL} = \frac{q^2}{q^2+2q(1-q)}$. Second, for each combination of $P_{LS}$ and $P_{SL}$, we change ties to non-ties and vice versa, i.e. we change $P_{LN} = \frac{(1-p)^2}{p^2+2p(1-p)+(1-p)^2}$, $P_{NL} = \frac{n^2}{n^2+2n(1-n)+(1-n)^2}$, $P_{SN} = \frac{(1-q)^2}{q^2+2q(1-q)+(1-q)^2}$ and $P_{NS} = \frac{(1-n)^2}{n^2+2n(1-n)+(1-n)^2}$.

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 $P_{LS}$, $P_{SL}$, $P_{LN}$, $P_{NL}$, $P_{SN}$ 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 (7) replacing the real $G_r^L$ and $G_r^S$ matrices with the simulated ones in turn so that, in total, we estimate model (7) 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 3 shows the results of our simulation experiment for the *2SLS bias-corrected lagged* *estimator*.[21] It depicts the estimates of long-lived and short-lived tie effects with 90% confidence bands, in the upper and lower panel, respectively.[22]

[*Insert Figure* 3 *here*]

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 3 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 5 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.

[*Insert Figure* 4 *here*]

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 3 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 4 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

[21] The simulation results for the other estimators are similar. They are available upon request.

[22] 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 $W_i = \frac{1}{n}\sum_{j=1}^{n} \sigma_{ij}^2$ , $B_i = \frac{1}{n}\sum_{j=1}^{n}(\phi_{ij} - \bar{\phi}_i)^2$ , $\sigma_{ij}^2$ is the estimated variance of the *jth* estimator at the *ith* replacement rate, $\phi_{ij}$ is the *jth* estimate at the *ith* replacement rate and $\bar{\phi}_i$ is the mean across the $n$ estimates. In this experiment, $n = 100$.

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 5 the rejection rates[23] 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 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* 5 *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.

[23]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.

# 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 (5):

$$\begin{aligned} y_{i,r,t} &= \phi^L \sum_{j=1}^{n_r} g_{ij,r,t}^L y_{j,r,t} + \phi^S \sum_{j=1}^{n_r} g_{ij,r,t}^S y_{j,r,t} + \boldsymbol{x}_{i,r}' \delta \\ &+ \frac{1}{g_{i,r,t}^L} \sum_{j=1}^{n_r} g_{ij,r,t}^L x_{j,r,t}' \gamma^L + \frac{1}{g_{i,r,t}^S} \sum_{j=1}^{n_r} g_{ij,r,t}^S x_{j,r,t}' \gamma^S + \eta_{r,t} + \epsilon_{i,r,t}, \end{aligned} \tag{11}$$

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 will first estimate equation (11) for peers who are only friends in Wave I, then for peers who are only friends in Wave II and finally for peers who are friends in both Waves I and II. We will look at how peers (as defined above) 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.[24] It ranges between 0 and 6.09, with a mean equal to 2.29 and a standard

[24] 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.

deviation equal to 1.49.

The estimation results of model (11) 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.

[*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 educational choices (i.e. years of schooling).

# 8 Concluding remarks

In this paper, we investigate *heterogenous peer effects* by looking at the impact of friends made at school on education decisions. We find that there are long-term peer effects. We also find that a *long-lived* relationship (i.e. a student who has been nominated twice) has a positive impact on own education outcomes while a *short-lived* relationship (i.e. a student who has been nominated only once) does not. Our findings also reveal that the strength of interactions between students matters in a student's career. In line with several studies in sociology and economics (e.g. Coleman, 1988, Wellman and Wortley, 1990, Akerlof and Kranton, 2002), our results suggest that long-lasting social interactions affect how much students value achievement, their human capital accumulation, and how social norms are formed. On the other hand, we find that, in the short run, any relationship (whether it is a long or short lived) has an impact on current grades.

# References


[1] Akerlof, G.A. and R.E. Kranton (2002), "Identity and schooling: Some lessons for the economics of education," *Journal of Economic Literature* 40, 1167-1201.

[2] Ballester, C., Calvó-Armengol, A. and Y. Zenou (2006), "Who's who in networks. Wanted: the key player," *Econometrica* 74, 1403-1417.

[3] Bekker, P. (1994), "Alternative approximations to the distributions of instrumental variable estimators," *Econometrica* 62, 657-681.

[4] Bekker, P. and J. van der Ploeg (2005), "Instrumental variable estimation based on grouped data," *Statistica Neerlandica* 59, 239-267.

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

[6] Bramoullé, Y., Djebbari, H. and B. Fortin (2009), "Identification of peer effects through social networks," *Journal of Econometrics* 150, 41-55.

[7] Brooks-Gunn, J., Duncan, G.J., Kato Klebanov, P. and N. Sealand (1993), "Do neighborhoods influence child and adolescent development," *American Journal of Sociology* 99, 353-395.

[8] Calvó-Armengol, A., Patacchini, E. and Y. Zenou (2009), "Peer effects and social networks in education," *Review of Economic Studies* 76, 1239-1267.

[9] Chow, G.C. (1960), "Test of equality between sets of coefficients in two linear regressions," *Econometrica* 28, 591-603.

[10] Coleman, J. (1988), "Social capital in the creation of human capital," *American Journal of Sociology* 94, 95-120.

[11] Corcoran, M., Gordon, R., Laren, D. and G. Solon (1992), "The association between men's economic status and their family and community origins" *Journal of Human Resources* 27, 575-601.

[12] Datcher, L. (1982), "Effects of communuty and family background on achievement," *Review of Economics and Statistics* 64, 32-41.

[13] De Giorgi, G., Pellizzari, M. and S. Redaelli (2010), "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics* 2, 241-275.

[14] Donald, S.G., and W.K. Newey (2001), "Choosing the number of Instruments," *Econometrica* 1161-1191.

[15] Debreu, G. and I.N. Herstein (1953), "Nonnegative square matrices," *Econometrica* 21, 597-607.

[16] Durlauf, S.E. (2004), "Neighborhood effects," In: J.V. Henderson and J-F. Thisse (Eds.), *Handbook of Regional and Urban Economics Vol. 4*, Amsterdam: Elsevier Science, pp. 2173-2242.

[17] Evans, W.N, Oates, W.E and R.M. Schwab (1992), "Measuring peer group effects: A study of teenage behavior," *Journal of Political Economy* 100, 966-991.

[18] Goldsmith-Pinkham, P. and G.W. Imbens (2013), "Social networks and the identification of peer effects," *Journal of Business and Economic Statistics* 31, 253-264.

[19] Granovetter, M.S. (1973), "The strength of short-lived ties," *American Journal of Sociology* 78, 1360-1380.

[20] Granovetter, M.S. (1974), *Getting a Job: A Study of Contacts and Careers*, Cambridge, MA: Harvard University Press.

[21] Granovetter, M.S. (1983), "The strength of short-lived ties: A network theory revisited," *Sociological Theory* 1, 201-233.

[22] Hansen, B.E. (2000), "Testing for structural change in conditional models," *Journal of Econometrics* 97, 93-115.

[23] Hansen, B.E. (2001), "The new econometrics of structural change: Dating breaks in U.S. labor productivity," *Journal of Economic Perspectives* 15, 117-128.

[24] Hansen, C., Hausman, J. and W. Newey (2008), "Estimation with many instrumental variables," *Journal of Business and Economic Statistics* 26, 398-422.

[25] Ioannides, Y.M. (2011), "Neighborhood effects and housing," In: J. Benhabib, A. Bisin, and M.O. Jackson (Eds.), *Handbook of Social Economics, Vol. 1B*, Amsterdam: Elsevier Science, pp. 1281-1340.

[26] Ioannides, Y.M. (2012), *From Neighborhoods to Nations: The Economics of Social Interactions*, Princeton: Princeton University Press.

[27] Ioannides, Y.M. and G. Topa (2010), "Neighborhood effects: Accomplishments and looking beyond them," *Journal of Regional Science* 50, 343-362.

[28] Jackson, M.O. (2008), *Social and Economic Networks*, Princeton: Princeton University Press.

[29] Jackson, M.O. and Y. Zenou (2014), "Games on networks," In: P. Young and S. Zamir (Eds.), *Handbook of Game Theory, Vol. 4*, Amsterdam: Elsevier, forthcoming.

[30] Lai, G., N. Lin and S.-Y. Leung (1998), "Network resources, contact resources, and status attainment," *Social Networks* 20, 159-178.

[31] Lee, L-F. (2002), "Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models," *Econometric Theory* 18, 252-277.

[32] Lee, L-F. and Liu X. (2010), "Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances," *Econometric Theory* 26, 187-230.

[33] Lee, L-F., Liu, X. and X. Lin (2010), "Specification and estimation of social interaction models with network structures," *Econometrics Journal* 13, 145-176.

[34] Lin, X. (2010), "Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables," *Journal of Labor Economics* 28, 825-860.

[35] Lin, N., W.M. Ensel and J.C. Vaughn (1981), "Social resources and strength of ties: Structural factors in occupational status attainment," *American Sociological Review* 46, 393-405.

[36] Liu, X. and L-F. Lee (2010), "GMM estimation of social interaction models with centrality," *Journal of Econometrics* 159, 99-115.

[37] Liu, X., Patacchini, E., Zenou, Y. and L-F. Lee (2012), "Criminal networks: Who is the key player?" CEPR Discussion Paper No. 8772.

[38] Manski, C.F. (1993), "Identification of endogenous effects: The reflection problem," *Review of Economic Studies* 60, 531-542.

[39] Marsden, P.V. and J.S. Hurlbert (1988), "Social resources and mobility outcomes: A replication and extension," *Social Forces* 66, 1038-1059.

[40] Meyer, C.D. (2000), *Matrix Analysis and Applied Linear Algebra*, Philadelphia: SIAM.

[41] Patacchini, E. and Y. Zenou (2008), "The strength of short-lived ties in crime," *European Economic Review* 52, 209-236.

[42] Putnam, R. (2000), *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon and Schuster.

[43] Sacerdote, B. (2001), "Peer effects with random assignment: Results from Dartmouth roomates," *Quarterly Journal of Economics* 116, 681-704.

[44] Wasserman, S. and K. Faust (1994), *Social Network Analysis. Methods and Applications*, Cambridge: Cambridge University Press.

[45] Wellman, B. and S. Wortley (1990), "Different strokes from different folks: Community ties and social support," *American Journal of Sociology* 96, 558-588.

[46] Yakubovich, V. (2005), "Short-lived ties, information, and influence: How workers find jobs in a local Russian labor market," *American Sociological Review* 70, 3, 408-421.

[47] Zax, J.S. and D.I. Rees (2002), "IQ, academic performance, environment, and earnings," *Review of Economics and Statistics* 84, 600-616.

## Appendix A: Proofs of propositions and corollaries in the text

**Proof of Proposition 1:** We need to show that $\mathbf{I}-\mathbf{A}$ is non-singular (i.e. invertible), where $\mathbf{A}\equiv\phi^{L}\mathbf{G}^{L}+\phi^{S}\mathbf{G}^{S}$. We know that $\mathbf{I}-\mathbf{A}$ is non-singular if $\mu_1\left(\phi^{L}\mathbf{G}^{L}+\phi^{S}\mathbf{G}^{S}\right)<1$ (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). ▮

**Proof of Corollary 1:** Let us start with two lemmas. Denote the spectral radius of $\mathbf{A}$ **by** $\mu_1(\mathbf{A})$.

**Lemma 1** *If* $\boldsymbol{A}$ *is an* $n\times n$ *Hermitian matrix, then*

$$\mu_1(\boldsymbol{A})=\sup_{|\boldsymbol{v}|=\mathbf{1}}\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}. \tag{12}$$

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

$$R(\boldsymbol{A},\boldsymbol{x})=\frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}$$

where $\boldsymbol{x}^T$ is the transpose of $\boldsymbol{x}$. This is equal to

$$R(\boldsymbol{A},\boldsymbol{x})=\frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_ja_{ij}}{\boldsymbol{x}^T\boldsymbol{x}}.$$

That is

$$R(\boldsymbol{A},\boldsymbol{x})=\frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}=\frac{\frac{\boldsymbol{x}^T\mathbf{A}\boldsymbol{x}}{\|\boldsymbol{x}\|_2^2}}{\frac{\boldsymbol{x}^T\boldsymbol{x}}{\|\boldsymbol{x}\|_2^2}}=\frac{\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right)^T\boldsymbol{A}\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right)}{\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right)^T\left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right)}$$

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

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

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

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

We want to show that $\sup_{\boldsymbol{y}\in\mathbb{R}^n\setminus\{\mathbf{0}\}}\left\{\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y} \mid \|\boldsymbol{y}\|_2 = 1\right\} = \mu_1(\boldsymbol{A})$. Observe that the function $\boldsymbol{y} \to \boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y}$ is continuous with compact domain ($n-1$ dimensional sphere). Every continuous function attains a maximum and a minimum on a compact set. Thus, there is a $\boldsymbol{y}$ so that the sup is a maximum. The problem is to find the critical points of the function $\boldsymbol{y} \to \boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y}$ subject to the constraint $\|\boldsymbol{y}\|_2^2 = \boldsymbol{y}^T\boldsymbol{y} = 1$, i.e. to find the critical points of the following Lagrangian

$$\mathcal{L}(\boldsymbol{y}) = \boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y} - \mu\left(\boldsymbol{y}^T\boldsymbol{y} - 1\right)$$

where $\mu$ is a Lagrange multiplier. The stationary points of $\boldsymbol{y} \to \mathcal{L}(\boldsymbol{y})$ are given by: $\frac{d\mathcal{L}(\boldsymbol{y})}{d\boldsymbol{y}} = \mathbf{0}$. We obtain:

$$\frac{d\mathcal{L}(\boldsymbol{y})}{d\boldsymbol{y}} = \left(\boldsymbol{A}+\boldsymbol{A}^T\right)\boldsymbol{y} - \mu\left(\boldsymbol{I}+\boldsymbol{I}^T\right)\boldsymbol{y} = \mathbf{0}.$$

Using the fact that $\boldsymbol{A}$ is *symmetric*, i.e. $\mathbf{A}=\boldsymbol{A}^T$, then

$$\frac{d\mathcal{L}(\boldsymbol{y})}{d\boldsymbol{y}} = 2\boldsymbol{A}\boldsymbol{y} - 2\mu\boldsymbol{y} = \mathbf{0}$$

which is equivalent to

$$\boldsymbol{A}\boldsymbol{y} = \mu\,\boldsymbol{y}.$$

This means that $\boldsymbol{y}$ is an eigenvector and $\mu$ the associated eigenvalue. Therefore, the eigenvectors $\boldsymbol{y}_1,...,\boldsymbol{y}_n$ of $\boldsymbol{A}$ 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 $R(\boldsymbol{A},\boldsymbol{x})$.

At the stationary points, we have

$$\boldsymbol{y}^T\mathbf{A}\mathbf{y} = \boldsymbol{y}^T\mu\,\boldsymbol{y} = \boldsymbol{\mu}\,\boldsymbol{y}^T\boldsymbol{y} = \mu.$$

This implies that

$$\sup_{\boldsymbol{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} R(\boldsymbol{A},\boldsymbol{x}) = \sup_{\boldsymbol{y}\in\mathbb{R}^n\setminus\{\mathbf{0}\}}\left\{\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y} \mid \|\boldsymbol{y}\|_2 = 1\right\} = \mu_1(\boldsymbol{A}).$$

Thus, we have shown that $\mu_1(\boldsymbol{A}) = \sup_{|\boldsymbol{v}|=\mathbf{1}} \boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v}$. ▮

**Lemma 2** *If $\boldsymbol{A}$ and $\boldsymbol{B}$ are two $n \times n$ symmetric matrices, then*

$$\mu_1(\boldsymbol{A}+\boldsymbol{B}) \leq \mu_1(\boldsymbol{A}) + \mu_1(\boldsymbol{B}). \tag{13}$$

**Proof:** In Lemma 1, we have shown that

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

which implies that

$$\mu_1(\boldsymbol{B}) = \sup_{|\boldsymbol{v}|=\mathbf{1}} \boldsymbol{v}^T \boldsymbol{B}\boldsymbol{v}$$

and

$$\mu_1(\boldsymbol{A}+\boldsymbol{B}) = \sup_{|\boldsymbol{v}|=\mathbf{1}} \boldsymbol{v}^T(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{v}.$$

We now need to show that

$$\mu_1(\boldsymbol{A}+\boldsymbol{B}) \leq \mu_1(\boldsymbol{A}) + \mu_1(\boldsymbol{B}).$$

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

$$\begin{aligned}
\mu_1(\boldsymbol{A}+\boldsymbol{B}) &= \sup_{|\boldsymbol{v}|=\mathbf{1}} \boldsymbol{v}^T(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{v} \\
&= \sup_{|\boldsymbol{v}|=\mathbf{1}} \left\{\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^T\boldsymbol{B}\boldsymbol{v}\right\} \\
&\leq \sup_{|\boldsymbol{v}|=\mathbf{1}} \boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v} + \sup_{|\boldsymbol{v}|=\mathbf{1}} \boldsymbol{v}^T\boldsymbol{B}\boldsymbol{v} \\
&= \mu_1(\boldsymbol{A}) + \mu_1(\boldsymbol{B}),
\end{aligned}$$

which is the statement of the lemma. ▮

In Proposition 1, we have shown that $\mathbf{I}-\mathbf{A}$ is non-singular if $\mu_1\left(\phi^L\mathbf{G}^L + \phi^S\mathbf{G}^S\right) < 1$. Assume now that $\mathbf{G}$ is symmetric so that both $\mathbf{G}^L$ and $\mathbf{G}^S$ are also symmetric. We can use Lemma 2, which states that (given $\phi^L > 0$ and $\phi^S > 0$):

$$\begin{aligned}
\mu_1\left(\phi^L\mathbf{G}^L + \phi^S\mathbf{G}^S\right) &\leq \mu_1\left(\phi^L\mathbf{G}^L\right) + \mu_1\left(\phi^S\mathbf{G}^S\right) \\
&= \phi^L\mu_1\left(\mathbf{G}^L\right) + \phi^S\mu_1\left(\mathbf{G}^S\right).
\end{aligned}$$

For $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$, we say that $A \leq B$ 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 $g_{ij}^L \leq g_{ij}$ and $g_{ij}^S \leq g_{ij}$ for all $i, j$). As a result, using Theorem I* in Debreu and Herstein (1953, p 600), we have: $\mu_1\left(\mathbf{G}^L\right) \leq \mu_1(\mathbf{G})$ and $\mu_1\left(\mathbf{G}^S\right) \leq \mu_1(\mathbf{G})$. This implies that: $\phi^L\mu_1\left(\mathbf{G}^L\right) + \phi^S\mu_1\left(\mathbf{G}^S\right) \leq \left(\phi^L + \phi^S\right)\mu_1(\mathbf{G})$ and the condition for $\mathbf{I} - \phi^L \mathbf{G}^L - \phi^S \mathbf{G}^S$ to be non-singular is given by:

$$\left(\phi^L + \phi^S\right) \mu_1 \left(\mathbf{G}\right) < 1$$

which is the condition given in Corollary 1. ∎

## Appendix B: Data Appendix

Table B1 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.

[*Insert Table B1 here*]

Table B1: Data description and and summary statistics

| Variables | Description | Average (Std.Dev.) | Min - Max |
|---|---|---|---|
| **Wave II (grade 7-12 )** | | | |
| *Individual socio-demographic* | | | |
| Female | Dummy variable taking value one if the respondent is female. | 0.53 (0.50) | 0 - 1 |
| Black or African American | Race dummies. "White" is the reference group | 0.19 (0.39) | 0 - 1 |
| Other races | // | 0.10 (0.30) | 0 - 1 |
| Student grade | Grade of student in the current year. | 9.07 (1.65) | 7 - 12 |
| Religion practice | Response to the question: "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?" | 3.79 (1.83) | 1 - 5 |
| Mathematics score A | Mathematics score dummies. Score in mathematics at the most recent grading period. D is the reference category, coded (A, B, C, D, missing). | 0.29 (0.45) | 0 - 1 |
| Mathematics score B | // | 0.34 (0.48) | 0 - 1 |
| Mathematics score C | // | 0.21 (0.41) | 0 - 1 |
| Mathematics score Missing | // | 0.05 (0.21) | 0 - 1 |
| GPA | 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. | 2.29 (1.49) | 0 - 6.09 |
| GPA of peers | Sum of GPA attained by respondent's peers | 11.36 (8.85) | 0 - 53.06 |
| Self esteem | 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. | 4.00 (1.09) | 1 - 6 |
| Physical development | 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 | 3.31 (1.11) | 1 - 5 |
| *Family background* | | | |
| Household size | Number of people living in the household | 3.40 (1.34) | 1 - 11 |
| Two married parent family | Dummy taking value one if the respondent lives in a household with two parents (both biological and non biological) that are married. | 0.73 (0.44) | 0 - 1 |
| Parent education | 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. | 3.25 (0.97) | 1 - 5 |
| Parent occupation manager | Parent occupation dummies. Closest description of the job of (biological or nonbiological) parent that is living with the child is manager. If both parents are in the household, the occupation of the father is considered. "none" is the reference group | 0.11 (0.31) | 0 - 1 |
| Parent occupation professional/technical | // | 0.21 (0.41) | 0 - 1 |
| Parent occupation office or sales worker | // | 0.10 (0.33) | 0 - 1 |
| Parent occupation manual | // | 0.30 (0.46) | 0 - 1 |
| Parent occupation other | // | 0.14 (0.35) | 0 - 1 |
| *Protective factors* | | | |
| School attachment | 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. | 1.90 (0.90) | 1 - 5 |
| Relationship with teachers | 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 | 0.91 (0.94) | 0 - 4 |
| Social inclusion | Response to the question: "How much do you feel that adults care about you, coded as 5= very much, 4= quite a bit, 3= somewhat, 2= very little, 1= not at all | 4.47 (0.73) | 1 - 5 |
| *Residential neighborhood* | | | |
| Residential building quality | 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. | 1.52 (0.80) | 1 - 4 |
| *Contextual effects* | Average of peers' characteristics for all of listed variables. | | |
| **Wave IV (aged 25 - 31)** | | | |
| Years of education | Years of education attained by the individual. | 14.42 (3.21) | 7 - 24 |
| Years of education of peers | Sum of years of education attained by respondent's peers. | 35.73 (29.48) | 7 - 326 |
| Children | Dummy variable taking value one if the respondent has a child. | 0.43 (0.50) | 0 - 1 |
| Married | Variable taking value one if the respondent is married | 0.42 (0.49) | 0 - 1 |
| Religion practice | Response to the question: 'How often have you attended religioius 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. | 1.75 (1.64) | 0 - 5 |
| **Networks** | | | |
| Links in Wave I | Number of individual links in Wave I. | 2.60 (2.57) | 1 - 21 |
| Links in Wave II | Number of individual links in Wave II. | 2.49 (2.50) | 1 - 26 |
| Deleted links | Percentage of nominations in Wave I not renewed in Wave II. | 0.61 (0.37) | 0 - 1 |
| New links | Percentage of new nominations in Wave II. | 0.44 (0.36) | 0 - 1 |
| Long-lived Ties | Percentage of Long-lived ties on total individual links. | 0.28 (0.28) | 0 - 1 |
| Short-lived Ties | Percentage of Short-lived ties on total individual links. | 0.72 (0.29) | 0 - 1 |

## 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 (7) can be written as:

$$\mathbf{Y} = \mathbf{Z}\theta + \boldsymbol{\iota} \cdot \boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{14}$$

where $\mathbf{Z} = (\mathbf{G^L Y}, \mathbf{G^S Y}, \mathbf{X}^*)$, $\theta = (\phi^L, \phi^S, \beta')'$ and $\boldsymbol{\iota} = \mathrm{D}(\mathbf{l}_{n_1}, \cdots, \mathbf{l}_{n_{\bar{r}}})$.

We treat $\boldsymbol{\eta}$ as a vector of unknown parameters. When the number of networks $\bar{r}$ is large, we have the incidental parameter problem. Let $\mathbf{J} = \mathrm{D}(\mathbf{J}_1, \cdots, \mathbf{J}_{\bar{r}})$, where $\mathbf{J}_r = \mathbf{I}_{n_r} - \frac{1}{n_r}\mathbf{l}'_{n_r}\mathbf{l}_{n_r}$. The network fixed effect can be eliminated by a transformation with $\mathbf{J}$ such that:

$$\mathbf{JY} = \mathbf{JZ}\theta + \mathbf{J}\boldsymbol{\epsilon}. \tag{15}$$

Let $\mathbf{M} = (\mathbf{I} - \phi^L\mathbf{G}^L - \phi^S\mathbf{G}^S)^{-1}$. The equilibrium outcome vector $\mathbf{Y}$ in (14) is then given by the reduced form equation:

$$\mathbf{Y} = \mathbf{M}(\mathbf{X}^*\beta + \boldsymbol{\iota} \cdot \boldsymbol{\eta}) + \mathbf{M}\boldsymbol{\epsilon}. \tag{16}$$

It follows that $\mathbf{G}^L\mathbf{Y} = \mathbf{G}^L\mathbf{M}\mathbf{X}^*\beta + \mathbf{G}^L\mathbf{M}\boldsymbol{\iota}\boldsymbol{\eta} + \mathbf{G}^L\mathbf{M}\boldsymbol{\epsilon}$ and $\mathbf{G}^S\mathbf{Y} = \mathbf{G}^S\mathbf{M}\mathbf{X}^*\beta + \mathbf{G}^S\mathbf{M}\boldsymbol{\iota}\boldsymbol{\eta} + \mathbf{G}^S\mathbf{M}\boldsymbol{\epsilon}$. $\mathbf{G}^L\mathbf{Y}$ and $\mathbf{G}^S\mathbf{Y}$ are correlated with $\boldsymbol{\epsilon}$ because $\mathrm{E}[(\mathbf{G}^L\mathbf{M}\boldsymbol{\epsilon})'\boldsymbol{\epsilon}] = \sigma^2\mathrm{tr}(\mathbf{G}^L\mathbf{M}) \neq 0$ and $\mathrm{E}[(\mathbf{G}^S\mathbf{M}\boldsymbol{\epsilon})'\boldsymbol{\epsilon}] = \sigma^2\mathrm{tr}(\mathbf{G}^S\mathbf{M}) \neq 0$. Hence, in general, (15) cannot be consistently estimated by OLS.[25] If $\mathbf{G}$ is row-normalized such that $\mathbf{G} \cdot \mathbf{l}_n = \mathbf{l}_n$, where $\mathbf{l}_n$ is a $n$-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 $\mathbf{Q}_1$ and $\mathbf{Q}_2$. They first consider the 2SLS estimator based on the conventional instrumental matrix for the estimation of (15): $\mathbf{Q}_1 = \mathbf{J}(\mathbf{GX}^*, \mathbf{X}^*)$ (*finite-IVs 2SLS*). Then, they propose to use additional instruments (IVs) $JG\iota$ and enlarge the instrumental matrix: $Q_2 = (Q_1, JG\iota)$ (many-IVs 2SLS). The additional IVs of $JG\iota$ 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 $\mathbf{Q}_1$. 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

[25] 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 3, 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.[26]

Let us now derive the best 2SLS estimator for equation (15). From the reduced form equation (14), we have $E(\mathbf{Z}) = [\mathbf{G}^L\mathbf{M}(\mathbf{X}^*\beta + \boldsymbol{\iota}\cdot\boldsymbol{\eta}), \mathbf{G}^S\mathbf{M}(\mathbf{X}^*\beta + \boldsymbol{\iota}\cdot\boldsymbol{\eta}), \mathbf{X}^*]$. The best IV matrix for $\mathbf{JZ}$ is given by

$$\mathbf{J}f = \mathbf{JE}(\mathbf{Z}) = J[\mathbf{G}^L\mathbf{M}(\mathbf{X}^*\beta + \boldsymbol{\iota}\cdot\boldsymbol{\eta}), \mathbf{G}^S\mathbf{M}(\mathbf{X}^*\beta + \boldsymbol{\iota}\cdot\boldsymbol{\eta}), \mathbf{X}^*] \tag{17}$$

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 $\mathbf{Q_0} = J[\mathbf{G}^L\mathbf{M}(\mathbf{X}^*+\boldsymbol{\iota}), \mathbf{G}^S\mathbf{M}(\mathbf{X}^*+\boldsymbol{\iota}), \mathbf{X}^*]$. As $\boldsymbol{\iota}$ has $\bar{r}$ columns the number of IVs in $\mathbf{Q_0}$ increases as the number of groups increases. Furthermore, as $\mathbf{M} = (\mathbf{I} - \phi^L\mathbf{G}^L - \phi^S\mathbf{G}^S)^{-1} = \sum_{j=0}^{\infty}(\phi^L\mathbf{G}^L + \phi^S\mathbf{G}^S)^j$ when $\sup ||\phi^L\mathbf{G}^L + \phi^S\mathbf{G}^S||_\infty < 1$, $\mathbf{MX}^*$ and $\mathbf{M}\boldsymbol{\iota}$, can be approximated by linear combinations of

$$(\mathbf{G}^L\mathbf{X}^*, \mathbf{G}^S\mathbf{X}^*, \mathbf{G}^S\mathbf{G}^L\mathbf{X}^*, \left(\mathbf{G}^L\right)^2\mathbf{X}^*, \left(\mathbf{G}^S\right)^2\mathbf{X}^*, \left(\mathbf{G}^S\right)^2\mathbf{G}^L\mathbf{X}^*, \left(\mathbf{G}^S\right)^2\left(\mathbf{G}^L\right)^2\mathbf{X}^*, \cdots)$$

and

$$(\mathbf{G}^L\boldsymbol{\iota}, \mathbf{G}^S\boldsymbol{\iota}, \mathbf{G}^S\mathbf{G}^L\boldsymbol{\iota}, \left(\mathbf{G}^L\right)^2\boldsymbol{\iota}, \left(\mathbf{G}^S\right)^2\boldsymbol{\iota}, \left(\mathbf{G}^S\right)^2\mathbf{G}^L\boldsymbol{\iota}, \left(\mathbf{G}^S\right)^2\left(\mathbf{G}^L\right)^2\boldsymbol{\iota}, \cdots),$$

respectively. Hence, $\mathbf{Q}_0$ can be approximated by a linear combination of

$$\begin{aligned}\mathbf{Q}_\infty \;=\; & \mathbf{J}(\mathbf{G}^L(\mathbf{G}^L\mathbf{X}^*, \mathbf{G}^S\mathbf{X}^*, \mathbf{G}^S\mathbf{G}^L\mathbf{X}^*, \cdots, \mathbf{G}^L\boldsymbol{\iota}, \mathbf{G}^S\boldsymbol{\iota}, \mathbf{G}^S\mathbf{G}^L\boldsymbol{\iota}, \cdots), \\ & \mathbf{G}^S(\mathbf{G}^L\mathbf{X}^*, \mathbf{G}^S\mathbf{X}^*, \mathbf{G}^S\mathbf{G}^L\mathbf{X}^*, \cdots, \mathbf{G}^L\boldsymbol{\iota}, \mathbf{G}^S\boldsymbol{\iota}, \mathbf{G}^S\mathbf{G}^L\boldsymbol{\iota}, \cdots), \mathbf{X}^*).\end{aligned}$$

Let $\mathbf{Q_K}$ be an $n \times K$ submatrix of $\mathbf{Q}_\infty$ (with $K \geq 3m+2$) including $X^*$. Let $Q_S$ be an $n \times K_S$ submatrix of $Q_{S\infty} = G^L(G^LX^*, G^SX^*, G^SG^LX^*, \cdots, G^L\iota, G^S\iota, G^SG^L\iota, \cdots)$ and $Q_S$ an $n \times K_S$ submatrix of $Q_{W\infty} = G^S(G^LX^*, G^SX^*, G^SG^LX^*, \cdots, G^L\iota, G^S\iota, G^SG^L\iota, \cdots)$. We assume that $\frac{K_S}{K_S} = 1$. Let $\mathbf{P_K} = \mathbf{Q_K}(\mathbf{Q_K}'\mathbf{Q_K})^{-1}\mathbf{Q_K}'$ be the projector of $\mathbf{Q_K}$. The resulting

[26] Liu and Lee (2010) also generalize this 2SLS approach to the GMM using additional quadratic moment conditions.

2SLS estimator is given by:

$$\widehat{\theta}_{2sls} = \left(\mathbf{Z}'\mathbf{P_K}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{P_K}\mathbf{y}. \tag{18}$$

**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 $\mathbf{X}^*$ are uniformly bounded constants, $\mathbf{X}^*$ has the full rank $k$ and $\lim_{n\to\infty}\mathbf{X}^{*\prime}\mathbf{X}^*$ exists and is nonsingular.

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

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

**Assumption C5:** There exists a $\mathbf{K}\times(3m+2)$ matrix $\pi_K$ such that $\|f-\mathbf{Q_K}\pi_K\|_\infty \to 0$ 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 $\mathbf{\Psi}_{K,S} = \mathbf{P_K}\mathbf{G}^L\mathbf{M}$ and $\mathbf{\Psi}_{K,W} = \mathbf{P_K}\mathbf{G}^S\mathbf{M}$. The following proposition shows consistency and asymptotic normality of the 2SLS estimator (18).

**Proposition 2** *Under assumptions C1-C5, if* $K/n\to 0$*, then* $\sqrt{n}\left(\widehat{\theta}-\theta-b_{2sls}\right)\xrightarrow{d} N\left(0,\sigma^2\overline{\mathbf{H}}^{-1}\right)$*, where* $b_{2sls} = \sigma^2\left(\mathbf{Z}'\mathbf{P_K}\mathbf{Z}\right)^{-1}\left[\operatorname{tr}\left(\mathbf{\Psi}_{K,L}\right), \operatorname{tr}\left(\mathbf{\Psi}_{K,S}\right), \mathbf{0}_{3m\times 1}\right]' = O_p\left(K/n\right)$.

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

$$\begin{aligned}\frac{1}{n}\mathbf{Z}'\mathbf{P_K}\mathbf{Z} &= \mathbf{H} - e_f + \frac{1}{n}v'\mathbf{P_K}f + \frac{1}{n}f'\mathbf{P_K}v + \frac{1}{n}v'\mathbf{P_K}v \\ &= \mathbf{H} + \mathbf{O}(\operatorname{tr}(e_f)) + \mathbf{O}_p(\sqrt{K/n}) + \mathbf{O}_p(K/n) \\ &= \overline{\mathbf{H}} + o_p(1)\end{aligned}$$

where $\mathbf{H} = \frac{1}{n} f' f$ and $e_f = \frac{1}{n} f'(I - \mathbf{P_K}) f$, because $e_f = \mathbf{O}(\mathrm{tr}(e_f))$, $\frac{1}{n} v' \mathbf{P_K} v = \mathbf{O}_p(K/n)$ and $\frac{1}{n} v' \mathbf{P_K} f = \mathbf{O}_p(\sqrt{K/n})$. Furthermore, we have

$$
\begin{aligned}
& (\mathbf{Z}'\mathbf{P_K}\boldsymbol{\epsilon} - \sigma^2 \left[\mathrm{tr}\left(\boldsymbol{\Psi}_{K,L}\right), \quad \mathrm{tr}\left(\boldsymbol{\Psi}_{K,S}\right), \mathbf{0}_{3m\times 1}\right]')/\sqrt{n} \\
= \; & h - f'(I - \mathbf{P_K})\boldsymbol{\epsilon}/\sqrt{n} + (\frac{1}{n} v'\mathbf{P_K}\boldsymbol{\epsilon} - \sigma^2 \left[\mathrm{tr}\left(\boldsymbol{\Psi}_{K,L}\right), \mathrm{tr}\left(\boldsymbol{\Psi}_{K,S}\right), \mathbf{0}_{3m\times 1}\right]')/\sqrt{n} \\
= \; & h + \mathbf{O}_p(\sqrt{\mathrm{tr}(e_f)}) + \mathbf{O}_p(\sqrt{K/n}) \\
= \; & h + o_p(1) \xrightarrow{d} N\left(0, \sigma^2 \overline{\mathbf{H}}\right),
\end{aligned}
$$

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

Due to the increasing number of IVs $\sqrt{n}\left(\widehat{\theta} - \theta\right)$ 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 $\mathbf{Q_K}$ gives the best IV estimator as $\sigma^2 \overline{\mathbf{H}}^{-1}$ reaches the efficiency lower bound for the IV estimators.

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

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 $\sigma^2 \overline{\mathbf{H}}^{-1}$ 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 $[G^L \iota, G^S \iota]$ 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 $Q_{\mathbf{K}}$, 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

$$\widehat{\theta}_{c2sls} = (\mathbf{Z}'\mathbf{P}_{\mathbf{Q_K}}\mathbf{Z})^{-1}\left[\mathbf{Z}'\mathbf{P}_{\mathbf{Q_K}}\mathbf{y} - \widetilde{\sigma}^2\left[\operatorname{tr}\left(\widetilde{\mathbf{\Psi}}_{K,L}\right), \operatorname{tr}\left(\widetilde{\mathbf{\Psi}}_{K,S}\right), \mathbf{0}_{3m\times 1}\right]'\right], \tag{19}$$

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

**Proposition 3** *Under assumptions C1-C5, if $K/n \to 0$ and $\widetilde{\sigma}, \widetilde{\phi}^L$ and $\widetilde{\phi}^S$ are $\sqrt{n}-$consistent initial estimators of $\sigma$, $\phi^L$ and $\phi^S$, then $\sqrt{n}\left(\widehat{\theta}_{c2sls} - \theta\right) \xrightarrow{d} N\left(0, \sigma^2\overline{\mathbf{H}}^{-1}\right)$.*

**Proof:** We need to show that

$$\{\widetilde{\sigma}^2\left[\operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^L\widetilde{\mathbf{M}}\right), \operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^S\widetilde{\mathbf{M}}\right)\right]' - \sigma^2\left[\operatorname{tr}\left(\mathbf{\Psi}_{K,L}\right), \operatorname{tr}\left(\mathbf{\Psi}_{K,S}\right)\right]'\}/\sqrt{n} = o_p(1)$$

where $\widetilde{\mathbf{M}} = \mathbf{M}(\widetilde{\phi}^L, \widetilde{\phi}^S)$. Given Proposition 2, this is quite straightforward since

$$\begin{aligned}
& \{\widetilde{\sigma}^2\left[\operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^L\widetilde{\mathbf{M}}\right), \operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^S\widetilde{\mathbf{M}}\right)\right]' - \sigma^2\left[\operatorname{tr}\left(\mathbf{\Psi}_{K,L}\right), \operatorname{tr}\left(\mathbf{\Psi}_{K,S}\right)\right]'\}/\sqrt{n} \\
= & \sqrt{n}(\widetilde{\sigma}^2 - \sigma^2)\left[\operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^L\widetilde{\mathbf{M}}\right), \operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^S\widetilde{\mathbf{M}}\right)\right]'/n \\
& + \sqrt{n}\sigma^2\left\{\operatorname{tr}\left[\mathbf{P_K}\mathbf{G}^L(\widetilde{\mathbf{M}} - \mathbf{M})\right], \operatorname{tr}\left[\mathbf{P_K}\mathbf{G}^S(\widetilde{\mathbf{M}} - \mathbf{M})\right]\right\}'/n \\
= & \sqrt{n}(\widetilde{\sigma}^2 - \sigma^2)\left[\operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^L\widetilde{\mathbf{M}}\right), \operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^S\widetilde{\mathbf{M}}\right)\right]'/n \\
& + \sqrt{n}\sigma^2\left[(\widetilde{\phi}^L - \phi^L)\operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^L\widetilde{\mathbf{M}}\mathbf{G}^L\mathbf{M}\right), 0\right]'/n \\
& + \sqrt{n}\sigma^2\left[(\widetilde{\phi}^S - \phi^S)\operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^L\widetilde{\mathbf{M}}\mathbf{G}^S\mathbf{M}\right), (\widetilde{\phi}^L - \phi^L)\operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^S\widetilde{\mathbf{M}}\mathbf{G}^L\mathbf{M}\right)\right]'/n \\
& + \sqrt{n}\sigma^2\left[0, (\widetilde{\phi}^S - \phi^S)\operatorname{tr}\left(\mathbf{P_K}\mathbf{G}^S\widetilde{\mathbf{M}}\mathbf{G}^S\mathbf{M}\right)\right]'/n \\
= & \mathbf{O}_p(\sqrt{K/n}) \\
= & o_p(1)
\end{aligned}$$

because $\widetilde{\mathbf{M}} - \mathbf{M} = \widetilde{\mathbf{M}}[(\widetilde{\phi}^L - \phi^L)\mathbf{G}^L\mathbf{M} + (\widetilde{\phi}^S - \phi^S)\mathbf{G}^S\mathbf{M}]$, as a special case of Lemma C.11 in Lee and Liu (2010). ▮

The 2SLS estimators of $\boldsymbol{\theta} = (\phi^L, \phi^S, \boldsymbol{\beta}')'$ considered in this paper are:

(i) *Finite-IV* : $\widehat{\boldsymbol{\theta}}_{2sls1} = (\mathbf{Z}'\mathbf{P}_1\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_1\mathbf{y}$, where $\mathbf{P}_1 = \mathbf{Q}_1(\mathbf{Q}_1'\mathbf{Q}_1)_1^{-1}\mathbf{Q}_1'$ and $\mathbf{Q}_1$ contains the linearly independent columns of $\mathbf{J}[\mathbf{X}^*, \mathbf{G}\mathbf{X}^*, \mathbf{G}\mathbf{G}\mathbf{X}^*]$.

(ii) *Many-IV* : $\widehat{\boldsymbol{\theta}}_{2sls2} = (\mathbf{Z}'\mathbf{P}_2\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_2\mathbf{y}$, where $\mathbf{P}_2 = \mathbf{Q}_2(\mathbf{Q}_2'\mathbf{Q}_2)_2^{-1}\mathbf{Q}_2'$ and $\mathbf{Q}_2$ contains the linearly independent columns of $[\mathbf{Q}_1, \mathbf{J}\mathbf{G}^L\boldsymbol{\iota}, \mathbf{J}\mathbf{G}^S\boldsymbol{\iota}]$.

(iii) *Bias-corrected*: $\widehat{\boldsymbol{\theta}}_{c2sls} = (\mathbf{Z}'\mathbf{P}_2\mathbf{Z})^{-1}\{\mathbf{Z}'\mathbf{P}_2\boldsymbol{y} - \widetilde{\sigma}^2_{2sls1}[\text{tr}\left(\mathbf{P_2}\mathbf{G}^L\widetilde{\mathbf{M}}\right), \text{tr}\left(\mathbf{P_2}\mathbf{G}^S\widetilde{\mathbf{M}}\right), \mathbf{0}_{3m\times 1}]'\}$, where $\tilde{\mathbf{M}} = (\mathbf{I} - \widetilde{\phi}^L_{2sls1}\mathbf{G}^L - \widetilde{\phi}^S_{2sls1}\mathbf{G}^S)^{-1}$, and $\widetilde{\sigma}^2_{2sls1}, \widetilde{\phi}^L_{2sls1}$ and $\widetilde{\phi}^S_{2sls1}$ are $\sqrt{n}$-consistent initial estimators of $\sigma^2, \phi^L$ and $\phi^S$ obtained by *Finite-IV.*

(iv) *Finite-IV lagged*: $\widehat{\boldsymbol{\theta}}_{2sls1L} = (\mathbf{Z}'\mathbf{P}_3\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_3\mathbf{y}$, where $\mathbf{P}_3 = \mathbf{Q}_{1L}(\mathbf{Q}_{1L}'\mathbf{Q}_{1L})_{1L}^{-1}\mathbf{Q}_{1L}'$ and $\mathbf{Q}_{1L}$ contains the linearly independent and lagged in time columns of $\mathbf{J}[\mathbf{X}^*, \mathbf{G}\mathbf{X}^*, \mathbf{G}\mathbf{G}\mathbf{X}^*]$.

(v) *Many-IV lagged*: $\widehat{\boldsymbol{\theta}}_{2sls2L} = (\mathbf{Z}'\mathbf{P}_4\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}_4\mathbf{y}$, where $\mathbf{P}_4 = \mathbf{Q}_{2L}(\mathbf{Q}_{2L}'\ \mathbf{Q}_{2L})^{-1}\mathbf{Q}_{2L}'$ and $\mathbf{Q}_{2L}$ contains the linearly independent columns of $[\mathbf{Q}_{1L}, \mathbf{J}\mathbf{G}^L\boldsymbol{\iota}, \mathbf{J}\mathbf{G}^S\boldsymbol{\iota}]$

(vi) *Bias-corrected lagged*:

$$\widehat{\boldsymbol{\theta}}_{c2slsL} = (\mathbf{Z}'\mathbf{P}_4\mathbf{Z})^{-1}\{\mathbf{Z}'\mathbf{P}_4\mathbf{y} - \widetilde{\sigma}^2_{2sls1}[\text{tr}\left(\mathbf{P_4}\mathbf{G}^L\widetilde{\mathbf{M}}\right), \text{tr}\left(\mathbf{P_4}\mathbf{G}^S\widetilde{\mathbf{M}}\right), \mathbf{0}_{3m\times 1}]'\}$$

where $\tilde{\mathbf{M}} = (\mathbf{I} - \widetilde{\phi}^L_{2sls1L}\mathbf{G}^L - \widetilde{\phi}^S_{2sls1L}\mathbf{G}^S)^{-1}$, and $\widetilde{\sigma}^2_{2sls1L}, \widetilde{\phi}^L_{2sls1L}$ and $\widetilde{\phi}^S_{2sls-1L}$ are $\sqrt{n}$-consistent initial estimators of $\sigma^2, \phi^L$ and $\phi^S$ obtained by *Finite-IV lagged.*

## Appendix D: 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**

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

**Betweenness centrality**

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

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

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 betweeness 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 $m = \sum_i \frac{d_i}{n}$ 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:

$$Cl_i = \frac{\sum_{l \in N_i(g)} \sum_{k \in N_i(g)} g_{lk}}{n_i(g)[n_i(g) - 1]}.$$

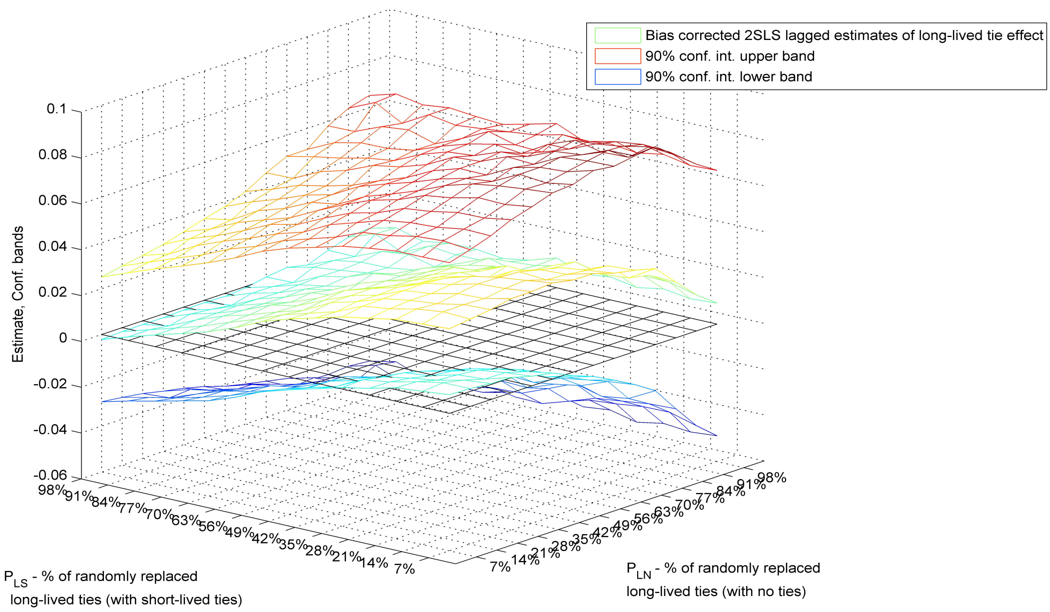
For the other nodes (singleton and with only one link), the value is imposed to be equal to zero. At the network level:

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

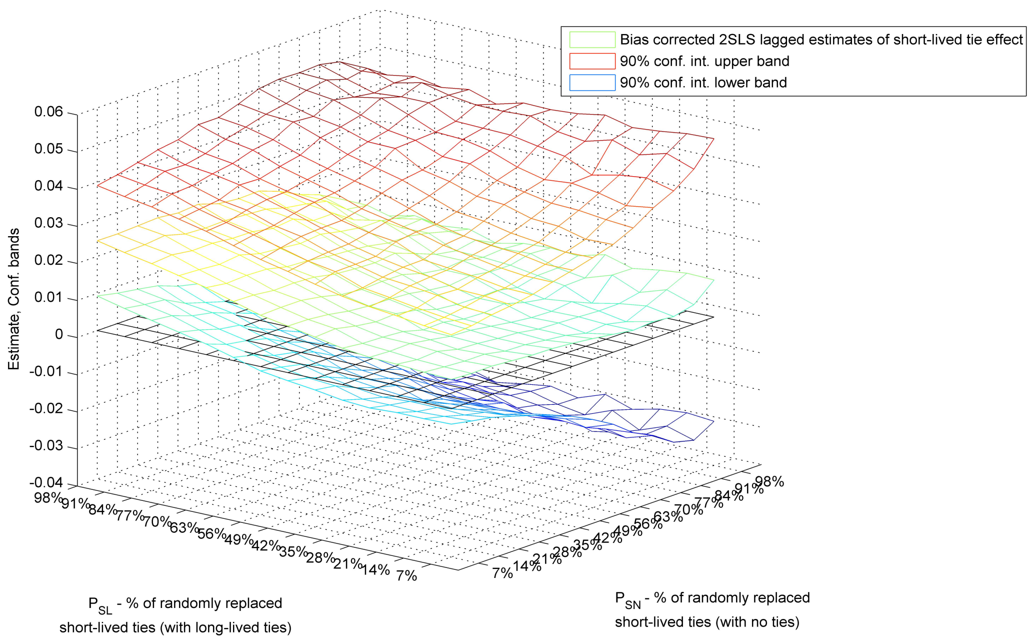
This is a simple weighted mean of the clustering coefficients in which each node has a weight proportional to the number of possible connections among its direct links.

Figure 3. Misspecification of long-lived ties and short-lived ties
Numerical simulation

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



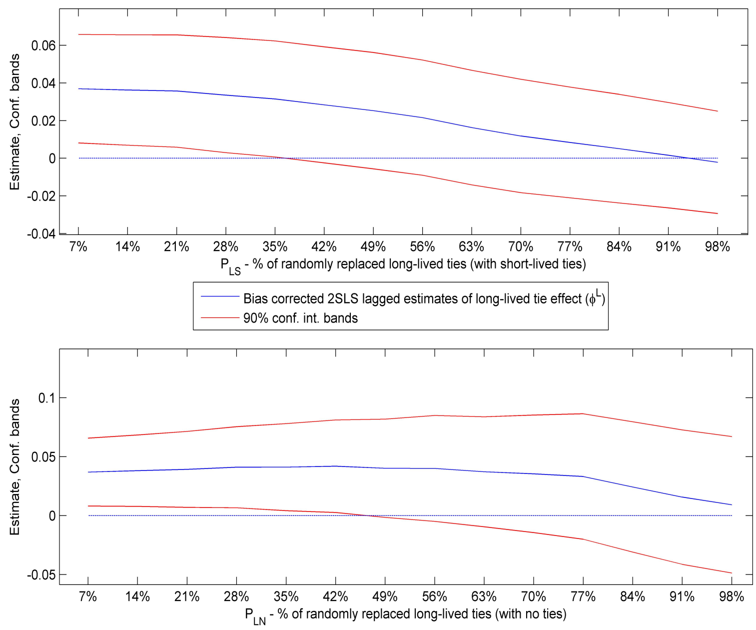

Bias corrected 2SLS lagged estimates of short-lived tie effect ($\phi^S$)



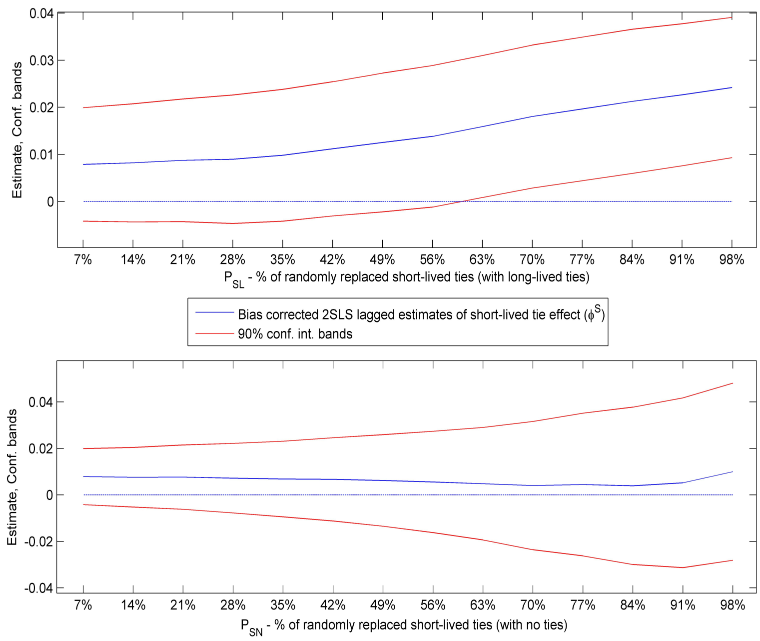

Notes: For each combination of replacement rates, we plot the average estimate of peer effects. Standard errors are derived assuming drawing independence and accounting for both within and between sample variation.

Figure 4. Simulation experiment
Single replacement effect

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

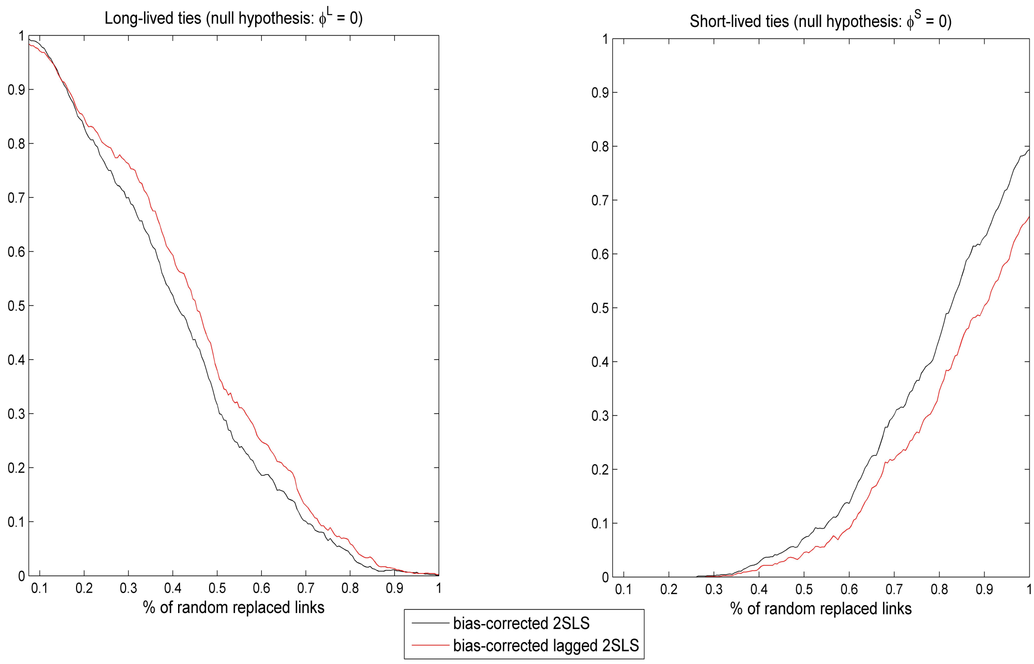



Bias corrected 2SLS lagged estimates of short-lived tie effect ($\phi^S$)




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 5. Rejection rates of null hypothesis
Comparing estimators




Notes: the rejection rate of null hypothesis for long-lived and short-lived tie effects.

Table 1: Link formation
logit estimation results

| Type of link | Long-lived | Short-lived |
|---|---|---|
| Female | -0.8325*** | 0.1482*** |
| | (0.057) | (0.035) |
| Black or African American | -1.7351*** | -0.8800*** |
| | (0.187) | (0.098) |
| Other races | 0.0149 | 0.1184** |
| | (0.074) | (0.046) |
| Student grade | -1.6270*** | -1.0470*** |
| | (0.045) | (0.022) |
| Religion Practice | 0.0110 | 0.0082 |
| | 0.0110 | 0.0082 |
| Mathematics score A | 0.1441*** | -0.0056 |
| | (0.047) | (0.030) |
| Mathematics score B | 0.0893** | 0.0220 |
| | (0.045) | (0.029) |
| Mathematics score C | 0.1796*** | 0.0202 |
| | (0.048) | (0.031) |
| Mathematics score missing | -0.1862* | -0.1080* |
| | (0.098) | (0.059) |
| Parent education | -0.0365 | -0.0275 |
| | (0.028) | (0.018) |
| Household Size | 0.0356* | 0.0407*** |
| | (0.019) | (0.012) |
| Residential building quality | -0.0661** | -0.1261*** |
| | (0.029) | (0.019) |
| Network fixed effects | yes | yes |
| Parent occupation dummies | yes | yes |
| Observations | 2,211 | 5,572 |

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

Table 2: Endogenous network formation
Testable implications

| | | | | | | |
|---|---|---|---|---|---|---|
| | **Dep.Var.** $|\hat{\epsilon}_i - \hat{\epsilon}_j|$ | | | | | |
| $\mathbf{g_{ij} = 1}$ | Full set of controls | | | Res. Build. Qual. unobserved | | |
| Threshold $\tau$ | $\tau = 25\%$ | $\tau = 35\%$ | $\tau = 45\%$ | $\tau = 25\%$ | $\tau = 35\%$ | $\tau = 45\%$ |
| ***Panel a) long-lived ties*** | | | | | | |
| $\mathbf{q^L_{ij} = \hat{g}^L_{ij}}$ | nc | -241.4128 | 395.0707 | 563.7353** | 491.9242*** | 300.8880*** |
| | nc | (641.4571) | (238.6882) | (165.686) | (30.267) | (62.905) |
| Constant | nc | 3.5840*** | 3.0330*** | 3.0264*** | 3.1502*** | 3.3745*** |
| | nc | (0.5734) | (0.2751) | (0.180) | (0.045) | (0.109) |
| Observations | nc | 88 | 134 | 128 | 180 | 212 |
| Network fixed effects | nc | yes | yes | yes | yes | yes |
| $P(q^L_{ij} < \tau \mid g^L_{ij} = 1)$ | 1,30% | 3.98% | 6.06% | 6% | 8% | 10% |
| $P(q^L_{ij} > \tau \mid g^L_{ij} = 1)$ | 98,70% | 96.02% | 93.94% | 94% | 92% | 90% |
| ***Panel b) short-lived ties*** | | | | | | |
| $\mathbf{q^S_{ij} = \hat{g}^S_{ij}}$ | 11.2581 | -10.1276 | 46.8319 | 94.2945*** | 87.3961** | 82.2564*** |
| | (13.245) | (12.412) | (33.631) | (18.761) | (29.597) | (24.108) |
| Constant | 4.2560*** | 4.3126*** | 4.1472*** | 3.2987*** | 3.2872*** | 3.3010*** |
| | (0.057) | (0.055) | (0.180) | (0.104) | (0.196) | (0.162) |
| Observations | 427 | 447 | 629 | 685 | 979 | 999 |
| Network fixed effects | yes | yes | yes | yes | yes | yes |
| $P(q^S_{ij} < \tau \mid g^S_{ij} = 1)$ | 7,66% | 8,02% | 11,29% | 12% | 18% | 18% |
| $P(q^S_{ij} > \tau \mid g^S_{ij} = 1)$ | 92,34% | 91,98% | 88,71% | 88% | 82% | 82% |

| | | | | | | |
|---|---|---|---|---|---|---|
| $\mathbf{g_{ij} = 0}$ | Full set of controls | | | Res. Build. Qual. unobserved | | |
| Threshold $\tau$ | $\tau = 95\%$ | $\tau = 85\%$ | $\tau = 75\%$ | $\tau = 95\%$ | $\tau = 85\%$ | $\tau = 75\%$ |
| ***Panel a) long-lived ties*** | | | | | | |
| $\mathbf{q^L_{ij} = \hat{g}^L_{ij}}$ | -4.0325 | -4.7153 | -4.9591 | -5.5639*** | -5.5276*** | -5.6703*** |
| | (2.863) | (3.583) | (3.104) | (0.780) | (0.787) | (0.749) |
| Constant | 3.0655*** | 3.1771*** | 3.2152*** | 3.2582*** | 3.5647*** | 3.6339*** |
| | (0.553) | (0.624) | (0.526) | (0.054) | (0.026) | (0.018) |
| Observations | 1,321 | 1,883 | 2,053 | 13,004 | 40,563 | 68,311 |
| Network fixed effects | yes | yes | yes | yes | yes | yes |
| $P(q^L_{ij} > \tau \mid g^L_{ij} = 0)$ | 0,00% | 0,00% | 0,00% | 0,00% | 0,01% | 0,02% |
| $P(q^L_{ij} < \tau \mid g^L_{ij} = 0)$ | 100,00% | 100,00% | 100,00% | 100,00% | 99,99% | 99,98% |
| ***Panel b) short-lived ties*** | | | | | | |
| $\mathbf{q^S_{ij} = \hat{g}^S_{ij}}$ | -4.4922 | -4.1874 | -3.9915* | -2.9780*** | -2.8409*** | -2.7277*** |
| | (3.014) | (2.860) | (2.283) | (0.745) | (0.430) | (0.465) |
| Constant | 3.5821*** | 3.4906*** | 3.4188*** | 3.2749*** | 3.4266*** | 3.5596*** |
| | (0.920) | (0.768) | (0.558) | (0.107) | (0.030) | (0.023) |
| Observations | 1,149 | 1,844 | 2,600 | 11,852 | 39,065 | 66,466 |
| Network fixed effects | yes | yes | yes | yes | yes | yes |
| $P(q^S_{ij} > \tau \mid g^S_{ij} = 0)$ | 0,00% | 0,00% | 0,00% | 0,00% | 0,01% | 0,02% |
| $P(q^S_{ij} < \tau \mid g^S_{ij} = 0)$ | 100,00% | 100,00% | 100,00% | 100,00% | 99,99% | 99,98% |

Notes: nc = not computed, number of observations < 30. $\hat{g}^M_{ij}$ with $M = L, S$ is estimated with logit model.
Threshold $\tau$ based on percentiles of the empirical distributions of $q^L_{ij}$ and $q^S_{ij}$.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Both link formation and outcome equation are estimated excluding these variables.

Table 3: Long-run peer effects

| Dep.Var. Years of Education | 2SLS | | | 2SLS Lagged | | |
|---|---|---|---|---|---|---|
| | Finite IV | Many IV | Bias Corrected | Finite IV | Many IV | Bias Corrected |
| Peer effects ($\phi$) | 0.0057 *** | 0.0052 *** | 0.0052 *** | 0.0064 *** | 0.0058 *** | 0.0059 *** |
| | (0.0020) | (0.0019) | (0.0019) | (0.0026) | (0.0020) | (0.0020) |
| Female | 0.9702 *** | 0.9718 *** | 0.9718 *** | 0.7383 | 1.0433 *** | 1.0434 *** |
| | (0.2004) | (0.2004) | (0.2004) | (0.6123) | (0.2408) | (0.2408) |
| Black or African American | -0.1346 | -0.1341 | -0.1341 | -0.2933 | -0.1833 | -0.1830 |
| | (0.4088) | (0.4088) | (0.4088) | (0.7722) | (0.4464) | (0.4464) |
| Other races | -0.3445 | -0.3447 | -0.3447 | -0.3873 | -0.2930 | -0.2927 |
| | (0.2912) | (0.2911) | (0.2911) | (0.4561) | (0.3111) | (0.3111) |
| Religion Practice | 0.2515 *** | 0.2521 *** | 0.2521 *** | 0.2515 *** | 0.2521 *** | 0.2521 *** |
| | (0.0540) | (0.0540) | (0.0540) | (0.1373) | (0.0599) | (0.0599) |
| Household Size | 0.0325 | 0.0327 | 0.0327 | 0.0247 | 0.0355 | 0.0355 |
| | (0.0574) | (0.0574) | (0.0574) | (0.0900) | (0.0605) | (0.0605) |
| Parent education | 0.2890 *** | 0.2895 *** | 0.2895 *** | 0.2275 * | 0.2262 *** | 0.2262 *** |
| | (0.0969) | (0.0969) | (0.0969) | (0.1528) | (0.1051) | (0.1051) |
| Mathematics score A | 1.3550 *** | 1.3578 *** | 1.3578 *** | 0.8917 * | 1.2388 *** | 1.2385 *** |
| | (0.2551) | (0.2551) | (0.2551) | (0.5065) | (0.2816) | (0.2816) |
| Mathematics score B | 0.9691 *** | 0.9683 *** | 0.9683 *** | 0.6889 * | 0.8941 *** | 0.8941 *** |
| | (0.2399) | (0.2399) | (0.2399) | (0.3966) | (0.2532) | (0.2532) |
| Mathematics score C | 0.5254 *** | 0.5272 *** | 0.5272 *** | 0.3422 | 0.5110 ** | 0.5107 ** |
| | (0.2583) | (0.2583) | (0.2583) | (0.4019) | (0.2709) | (0.2709) |
| Mathematics score missing | 0.6214 | 0.6210 | 0.6210 | 1.1487 | 0.6678 | 0.6677 * |
| | (0.4286) | (0.4286) | (0.4286) | (0.8355) | (0.4471) | (0.4471) |
| Resid. building qual. | 0.1771 ** | 0.1798 ** | 0.1797 ** | 0.0950 ** | 0.1264 ** | 0.1262 ** |
| | (0.0966) | (0.0966) | (0.0966) | (0.1669) | (0.1053) | (0.1053) |
| Student grade | 0.4781 *** | 0.4766 *** | 0.4766 *** | 0.4051 *** | 0.5042 *** | 0.5043 *** |
| | (0.0851) | (0.0850) | (0.0850) | (0.1654) | (0.0914) | (0.0914) |
| Children | -0.4171 *** | -0.4157 *** | -0.4157 *** | -1.5107 | -1.2429 ** | -1.2438** |
| | (0.1668) | (0.1668) | (0.1668) | (1.7631) | (0.6263) | (0.6263) |
| Religion Practice (Wave 4) | 0.1511 *** | 0.1514 *** | 0.1514 *** | 1.0470 | 0.3429 ** | 0.3427 ** |
| | (0.0514) | (0.0514) | (0.0514) | (1.1264) | (0.1873) | (0.1873) |
| Married | -0.1168 | -0.1166 | -0.1166 | -0.8750 | -0.6020 | -0.6024 |
| | (0.1655) | (0.1655) | (0.1655) | (2.113) | (0.6381) | (0.6381) |
| Parental occupation dummies | Yes | Yes | Yes | Yes | Yes | Yes |
| Contextual effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Network fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| First stage F statistic | 1210.64 | 700.35 | | 1240.51 | 708.83 | |
| OIR test p-value | 0.503 | 0.461 | | 0.554 | 0.503 | |
| Observations | 1819 | 1819 | 1819 | 1819 | 1819 | 1819 |
| Networks | 116 | 116 | 116 | 116 | 116 | 116 |

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

Table 4: Long and short-lived ties: Long-run effects

| Dep.Var. Years of Education | Total IV | Lagged IV |
|---|---|---|
| Long-lived ties ($\phi^L$) | 0.0317*** | 0.0345*** |
| | (0.0137) | (0.0155) |
| Short-lived ties ($\phi^S$) | 0.0062 | 0.0080 |
| | (0.0055) | (0.0063) |
| 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 | 1819 |
| Networks | 116 | 116 |

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

Table 5: Short-lived ties in different grades (10 - 12)

| Dep.Var. Years of Education | Total IV | Lagged IV |
|---|---|---|
| Long-lived ties ($\phi^L$) | 0.0419*** | 0.0485** |
| | (0.0187) | (0.0212) |
| Short-lived in lower grades ($\phi^{S_1}$) | 0.0015 | 0.0027 |
| | (0.0237) | (0.0331) |
| Short-lived in higher grades ($\phi^{S_2}$) | 0.0011 | 0.0049 |
| | (0.0207) | (0.0257) |
| 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 | 628 | 628 |
| Networks | 41 | 41 |

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

Table 6: Peer characteristics

| | Wave I | | | | Wave II | | | | Wave I and Wave II | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | mean | std | min | max | mean | std | min | max | mean | std | min | max |
| Years of education | 14.96 | 3.48 | 9.00 | 23.00 | 14.62 | 3.35 | 9.00 | 23.00 | 14.91 | 3.39 | 9.00 | 23 .00 |
| GPA | 2.89 | 0.87 | 0.00 | 4.85 | 2.91 | 0.94 | 0.00 | 4.81 | 2.88 | 0.97 | 0.00 | 4.82 |
| Female | 0.48 | 0.50 | 0.00 | 1.00 | 0.43 | 0.50 | 0.00 | 1.00 | 0.47 | 0.50 | 0.00 | 1.00 |
| Black or African American | 0.06 | 0.24 | 0.00 | 1.00 | 0.01 | 0.12 | 0.00 | 1.00 | 0.04 | 0.20 | 0.00 | 1.00 |
| Other races | 0.09 | 0.29 | 0.00 | 1.00 | 0.06 | 0.23 | 0.00 | 1.00 | 0.05 | 0.23 | 0.00 | 1.00 |
| Religion Practice | 3.98 | 1.82 | 1.00 | 7.00 | 3.88 | 1.89 | 1.00 | 7.00 | 3.84 | 1.86 | 1.00 | 7.00 |
| Household Size | 3.34 | 1.37 | 1.00 | 10.00 | 3.32 | 1.34 | 1.00 | 10.00 | 3.47 | 1.38 | 1.00 | 10.00 |
| Two married parent family | 0.75 | 0.44 | 0.00 | 1.00 | 0.77 | 0.42 | 0.00 | 1.00 | 0.80 | 0.40 | 0.00 | 1.00 |
| Parent education | 3.33 | 0.90 | 1.00 | 5.00 | 3.22 | 0.83 | 1.00 | 5.00 | 3.19 | 0.83 | 1.00 | 5.00 |
| Mathematics score A | 0.24 | 0.42 | 0.00 | 1.00 | 0.21 | 0.41 | 0.00 | 1.00 | 0.25 | 0.43 | 0.00 | 1.00 |
| Mathematics score B | 0.28 | 0.45 | 0.00 | 1.00 | 0.32 | 0.47 | 0.00 | 1.00 | 0.31 | 0.46 | 0.00 | 1.00 |
| Mathematics score C | 0.25 | 0.43 | 0.00 | 1.00 | 0.24 | 0.43 | 0.00 | 1.00 | 0.24 | 0.43 | 0.00 | 1.00 |
| Mathematics score missing | 0.08 | 0.27 | 0.00 | 1.00 | 0.08 | 0.28 | 0.00 | 1.00 | 0.07 | 0.26 | 0.00 | 1.00 |
| GPA | 2.33 | 1.53 | 0.00 | 6.09 | 2.27 | 1.48 | 0.00 | 6.09 | 2.25 | 1.55 | 0.00 | 6.09 |
| Residential uilding quality | 1.56 | 0.78 | 1.00 | 4.00 | 1.59 | 0.79 | 1.00 | 4.00 | 1.52 | 0.79 | 1.00 | 4.00 |
| Student grade | 10.44 | 0.50 | 10.00 | 11.00 | 10.43 | 0.50 | 10.00 | 11.00 | 10.40 | 0.49 | 10.00 | 11.00 |
| Children | 0.48 | 0.50 | 0.00 | 1.00 | 0.48 | 0.50 | 0.00 | 1.00 | 0.45 | 0.50 | 0.00 | 1.00 |
| Religion Practice (Wave 4) | 1.44 | 1.62 | 0.00 | 5.00 | 1.32 | 1.56 | 0.00 | 5.00 | 1.46 | 1.66 | 0.00 | 5.00 |
| Married | 0.50 | 0.50 | 0.00 | 1.00 | 0.50 | 0.50 | 0.00 | 1.00 | 0.49 | 0.50 | 0.00 | 1.00 |

Notes: Differences between means are never statistically significant at conventional levels of significance

Table 7: Network formation in Wave I and Wave II

Model (8) OLS estimation results

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

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

Table 8: Wave I and Wave II network structure

| | Wave I | Wave II |
|---|---|---|
| Network structure indicators | | |
| Density | 0.0010 | 0.0007 |
| Betweeness | 0.0040 | 0.0055 |
| Closeness | 0.0131 | 0.0090 |
| Assortativity | 2.4105 | 2.9041 |
| Clustering coefficient | 0.0784 | 0.1516 |

Notes: Network structure indicators are described in Appendix D.

Table 9: Directed networks

| Dep.Var. Years of Education | Total IV | Lagged IV |
|---|---|---|
| Long-lived ties ($\phi^L$) | 0.0393*** | 0.0474*** |
| | (0.0160) | (0.0183) |
| Short-lived ties ($\phi^S$) | 0.0049 | 0.0052 |
| | (0.0063) | (0.0070) |
| 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 | 1819 |
| Networks | 116 | 116 |

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

| Dep.Var. GPA | Total IV |
|---|---|
| Long-lived ties ($\phi^L$) | 0.0238*** |
| | (0.0097) |
| Short-lived ties ($\phi^S$) | 0.0079* |
| | (0.0046) |
| Individual socio-demographic | yes |
| Family Background | yes |
| Protective Factors | yes |
| Residential neighborhood | yes |
| Contextual Effects | yes |
| Network Fixed Effects | yes |
| Observations | 1819 |
| Networks | 116 |

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