Heterogeneous peer effects in education

Eleonora Patacchini, Edoardo Rainone and Yves Zenou

DOI/Link: https://doi.org/10.1016/j.jebo.2016.10.020

Reference: Patacchini, Eleonora, Edoardo Rainone and Yves Zenou (2017). "Heterogeneous Peer Effects in Education". *Journal of Economic Behavior & Organization*, 134(February), 190–227.

Journal of Economic Behavior & Organization 134 (2017) 190–227

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Heterogeneous peer effects in education☆

Eleonora Patacchini[a,b,c,d], Edoardo Rainone[e,1], Yves Zenou[f,g,*]

[a] Cornell University, United States
[b] EIEF, Italy
[c] IZA, Germany
[d] CEPR, UK
[e] Bank of Italy, Italy
[f] Monash University, Australia
[g] IFN, Sweden

ARTICLE INFO

Article history:
Received 13 November 2015
Received in revised form 2 August 2016
Accepted 23 October 2016
Available online 7 December 2016

JEL classification:
C31
D85
Z13

Keywords:
Spatial autoregressive model
Heterogeneous spillovers
2SLS estimation
Bayesian estimation
Education

ABSTRACT

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

1. Introduction

The influence of peers on educational outcomes has been widely studied in both economics and sociology (Sacerdote, 2011). However, many questions remain unanswered.[2] In particular, very little is known about the effect of school peers

☆ We thank the editor, two anonymous referees, Larry Blume, Chih-Sheng Hsieh, Alfonso Flores-Lagunes, Xiaodong Liu, Guido Kuersteiner, Francesca Molinari, Ingman Prucha, Theodoros Rapanos, John Rust and Frank Vella for very 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.

* Corresponding author at: Monash University, Department of Economics, Caulfield VIC 3145, Australia.
E-mail addresses: ep454@cornell.edu (E. Patacchini), edoardo.rainone@bancaditalia.it (E. Rainone), yves.zenou@monash.edu (Y. Zenou).

[1] The views expressed here do not necessarily reflect those of the Bank of Italy.

[2] 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.,

http://dx.doi.org/10.1016/j.jebo.2016.10.020

on the long-run outcomes of students. This is primarily due to the absence of data that provides both information on peers during teenage years and information on long-run outcomes. In addition, the mechanisms by which peers affect education are unclear.

In this paper, we analyze the long-run effects of peers' behavior on own educational outcomes by examining the role of different types of ties.

In the existing (enormous) literature on peer effects,[3] the term "heterogeneous peer effects" usually means that an individual response to peers' characteristics may vary by the level of the characteristic or that peer effects are different for different types of individuals (e.g. males versus females, whites versus blacks, etc.).[4] The possibility that an individual's response to peers' behavior may vary by peer type is usually overlooked.[5]

Our analysis is made possible by the unique information on friendship networks[6] among students in the United States provided by the AddHealth data. We exploit three unique features of the AddHealth data: (i) the nomination-based friendship information, which allows us to reconstruct the precise geometry of social contacts during high-school years, (ii) the variation in friendship network topology between Wave I and Wave II, which enables us to distinguish between *short-lived ties* and *long-lived ties* and (iii) the longitudinal dimension, which provides information about each individual and his/her friends' outcomes in adulthood.

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 outcomes (when adult) reported in the fourth wave in 2007–2008 (measured by the number of completed years of full time education). We define students as having a *long-lived tie* relationship if they have nominated each other in both waves (i.e. in Wave I in 1994–1995 and in Wave II in 1995–1996) and a *short-lived tie* relationship if they have nominated each other in one wave only. We also study the robustness of our results in terms of the definition and measurement of long and short-lived ties.

Our results show that there are strong and persistent peer effects in education. When looking at the role of *short-lived* and *long-lived ties* in education decisions, it appears that the education decisions of *short-lived* ties have no significant effect on individual long-run outcomes, regardless of whether peers interact 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 examining the effects of *weak* versus *strong* ties on labor-market outcomes. Strong ties are viewed as *stable* relationships and weak ties as *unstable* relationships.[7] Interestingly, compared to the literature on the labor market, we find the opposite result for educational outcomes.[8,9] 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 shape social norms, values and attitudes (see, e.g. Coleman, 1988; Wellman and Wortley, 1990). The collective value of "social networks", which is a relevant driver of long-run influences, need time and repeated interactions to be established (Putnam, 2000).

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

1992). 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; Ioannides, 2011, 2012).

[3] See Sacerdote (2014) for a recent review.

[4] See e.g. Griffith and Rask (2014), Tincani (2015), Yakusheva et al. (2014) and Arduini et al. (2014).

[5] A notable exception is Goldsmith-Pinkham and Imbens (2013), who use different network structures as a statistical exercise to investigate measurement errors in peer status. They estimate the model using a Bayesian approach.

[6] The economics of networks is a growing field. For overviews, see Jackson (2008, 2014), Blume et al. (2011), Ioannides (2012), Boucher and Fortin (2015), Graham (2015), Jackson and Zenou (2015), and Jackson et al. (2017).

[7] 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. For a formal analysis of the Granovetter's idea of weak and strong ties, see Zenou (2015).

[8] 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.

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

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

There are relatively 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. By exploiting a unique feature of the Israeli school placement system, which assigns peers randomly conditional on school choice, Lavy and Sand (2016) look at the impact of the number of pre-existing friends and their socioeconomic background on students' academic progress from elementary to middle school. They find that the number of friends and their characteristics have a positive effect, though the length of the relationship does not play any role.

Differently from this paper, our paper focuses on the analysis of endogenous rather than exogenous peer effects. This requires additional methodological efforts given that the OLS estimators are not valid due to the simultaneity between peers and own behavior. For that, we extend the Liu and Lee (2010) 2SLS approach to a network model with different interaction matrices. The asymptotic consistency and efficiency of the proposed estimators are proved. We also employ a Bayesian inferential method to integrate a network formation model with the study of behavior over the formed networks. Finally, we consider possible measurement errors in peer groups using a simulation experiment. Our results are robust to various types of network topology misspecifications.

One of the biggest issues in the peer effect literature mentioned above is the difficulty of identifying credible mechanisms through which the effects are obtained. Even the random assignment of peers (as, for example, in Sacerdote, 2001, and Lavy and Sand, 2016) does not address this problem as it only gives us internally valid estimates of peer effects on the outcome considered. Although our paper does not provide a definitive answer to the question of mechanism, it moves the literature forward by providing evidence on the effect of long- versus short-lived ties.

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

2. Data description

Our analysis is made possible by the use of a unique database on friendship networks from the National Longitudinal Survey of Adolescent Health (AddHealth). The AddHealth survey was 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 was 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, was then asked to compile a longer questionnaire containing more sensitive individual and household information (*in-home and parental data*). Those subjects were 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).[10] This information was 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).

[10] 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.

Moreover, one can distinguish between *long-lived* and *short-lived* ties in the data (a unique characteristic of our analysis). We define two students as having a *long-lived* friendship if they nominated each other in both waves (i.e. in Wave I in 1994–1995 *and* in Wave II in 1995–1996) and a *short-lived* friendship if they have nominated each other in one wave only (Wave I *or* Wave II). In Section 6.2.2 we check the robustness of our definitions when links may be erroneously observed as long or short-lived.[11]

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* by the number of completed years of full-time education in Wave IV.[12] The Wave IV study was designed as a follow-up of the nationally representative sample of 20,745 adolescents first interviewed in 1994 (Wave I). About 80% of the original sample were re-interviewed. Attrition can be considered as random (see Harris, 2013 for further details). Social contacts (i.e. friendship nominations) are, instead, collected in Waves I and II.

Our final sample 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% of the students do not nominate any friends and another 20% cannot be correctly linked. In addition, we exclude individuals in networks of 2–3 students or over 400 students, and exclude individuals who are not followed in Wave IV.[13]

In Table 1, we detail our sample selection procedure. We report the characteristics of five different samples, which correspond to the five different steps of our selection procedure. In column 1, we consider the full Wave I sample with 20,475 students. In column 2, we use the sample of students in Wave I who were also followed in Wave IV (15,701 students). In column 3, we display the sample of students obtained after the network construction procedure, i.e. when students with no nominations are eliminated (7,077 students). In column 4, we report the sample of students after having eliminated observations with missing values in variables (6,687 students). Finally, in column 5, we give the sample of students after having eliminated very small or very large networks. This is our sample with 1,819 students. Table 1 shows that the differences between these samples are never statistically significant. In Wave I, the mean and the standard deviation of network size are roughly 9.5 and 15, respectively. Roughly 61% of the nominations are not renewed in Wave II, and about 44% new ones are made. On average, these adolescents have roughly 30% long-lived ties and 70% short-lived ties. Further details on nomination data can be found in Table A1 in Appendix A. Appendix A also gives a precise definition of the variables used in our study as well as their descriptive statistics (see Table A1).[14]

3. Empirical model and identification strategy

3.1. Empirical model

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

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

where $y_{i,r,t+1}$ is the highest education level attained 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). Moreover, $x'_{i,r,t} = (x_{i,r,t}^1, \ldots, x_{i,r,t}^M)'$ indicates the M variables accounting for observable differences in individual characteristics of individual i at time t (parental education, neighborhood quality,

[11] In principle, a short-lived tie observed only in Wave II can be a long-lived tie if it is not severed later, while a tie observed in both Wave I and Wave II may be severed later, becoming a short-lived.

[12] 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 in which the highest qualification was achieved. Such information allows us to construct a reliable measure of each individual's completed years of education.

[13] 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 in these edge networks (see Calvó-Armengol et al., 2009). The representativeness of the sample is preserved. Summary statistics are available upon request.

[14] 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.

Table 1
Sample representativeness.

Sample	Wave I students	Wave IV students		Connected students		Students without missing values in variables		Students in networks of size 4–400	
	Mean (std)	Mean (std)	Difference [*p*-value]	Mean (std)	Difference [*p*-value]	Mean (std)	Difference [*p*-value]	Mean (std)	Difference [*p*-value]
Years of education		14.443		14.558	[0.509]	14.596	[0.503]	14.344	[0.479]
		(3.585)		(3.519)		(3.497)		(3.198)	
Female	0.505	0.532	[0.515]	0.533	[0.500]	0.535	[0.501]	0.526	[0.495]
	(0.500)	(0.499)		(0.499)		(0.499)		(0.499)	
Black or African American	0.232	0.230	[0.498]	0.201	[0.481]	0.202	[0.501]	0.135	[0.449]
	(0.422)	(0.421)		(0.401)		(0.402)		(0.341)	
Other races	0.203	0.186	[0.488]	0.196	[0.507]	0.191	[0.496]	0.084	[0.412]
	(0.402)	(0.389)		(0.397)		(0.393)		(0.277)	
Religion practice	3.929	3.963	[0.505]	3.925	[0.494]	3.916	[0.499]	3.875	[0.494]
	(1.796)	(1.771)		(1.774)		(1.773)		(1.809)	
Household size	3.614	3.607	[0.499]	3.623	[0.503]	3.622	[0.500]	3.388	[0.454]
	(1.656)	(1.625)		(1.597)		(1.552)		(1.273)	
Parent education	3.090	3.106	[0.505]	3.141	[0.510]	3.144	[0.501]	3.251	[0.532]
	(0.975)	(0.967)		(0.967)		(0.964)		(0.927)	
Mathematics score A	0.229	0.235	[0.505]	0.243	[0.505]	0.249	[0.504]	0.279	[0.519]
	(0.420)	(0.424)		(0.429)		(0.432)		(0.448)	
Mathematics score B	0.283	0.288	[0.504]	0.298	[0.506]	0.304	[0.503]	0.354	[0.530]
	(0.450)	(0.453)		(0.458)		(0.460)		(0.478)	
Mathematics score C	0.236	0.234	[0.498]	0.230	[0.498]	0.233	[0.501]	0.209	[0.484]
	(0.425)	(0.423)		(0.421)		(0.422)		(0.407)	
Mathematics score missing	0.092	0.085	[0.493]	0.080	[0.495]	0.065	[0.483]	0.036	[0.463]
	(0.289)	(0.278)		(0.272)		(0.246)		(0.187)	
Residential building quality	1.659	1.641	[0.494]	1.602	[0.487]	1.590	[0.496]	1.543	[0.483]
	(0.850)	(0.839)		(0.820)		(0.812)		(0.803)	
Student grade	9.669	9.633	[0.494]	9.688	[0.509]	9.666	[0.496]	8.973	[0.372]
	(1.635)	(1.635)		(1.636)		(1.634)		(1.368)	
Children		0.466		0.461	[0.497]	0.459	[0.499]	0.420	[0.478]
		(0.499)		(0.499)		(0.498)		(0.494)	
Religion practice (Wave IV)		1.639		1.649	[0.502]	1.655	[0.501]	1.537	[0.479]
		(1.607)		(1.600)		(1.600)		(1.583)	
Married		0.398		0.412	[0.508]	0.412	[0.500]	0.427	[0.508]
		(0.489)		(0.492)		(0.492)		(0.495)	
Observations	20,745	15,701		7,077		6,687		1,819	

Notes: *t*-tests for differences in means are performed. *p*-values are reported. Differences are computed w.r.t. the larger sample in the previous column.

etc.). Also $g^L_{i,r,t} = \sum_{j=1}^{n} g^L_{ij,r,t}$ and $g^S_{i,r,t} = \sum_{j=1}^{n} g^S_{ij,r,t}$ are the total number of long-lived and short-lived friends each individual i has in network r at time t. $\eta_{r,t}$ is the network fixed effect. Finally, $\epsilon_{i,r}$'s are i.i.d. innovations with zero mean and variance σ^2 for all i and r.

Let $\boldsymbol{Y}_r = (y_{1,r,t+1}, \ldots, y_{n_r,r,t+1})\prime$, $\boldsymbol{X}_r = (x_{1,r,t}, \ldots, x_{n_r,r,t})\prime$, and $\boldsymbol{\epsilon}_r = (\epsilon_{1,r}, \ldots, \epsilon_{n_r,r})\prime$. Denote the $n_r \times n_r$ adjacency matrix by $\boldsymbol{G}_r = [g_{ij,r}]$, the row-normalized of $\boldsymbol{G}_r$ by $\boldsymbol{G}^*_r$, and the n_r-dimensional vector of ones by $\boldsymbol{l}_{n_r}$. As above, let us split the adjacency matrix into two submatrices $\boldsymbol{G}^L_r$ and $\boldsymbol{G}^S_r$, which keep track of long-lived and short-lived friends, respectively. Then, model (1) can be written in matrix form as:

$$\boldsymbol{Y}_r = \phi^L \boldsymbol{G}^L_r \boldsymbol{Y}_r + \phi^S \boldsymbol{G}^S_r \boldsymbol{Y}_r + \boldsymbol{X}^*_r \beta + \eta_r \boldsymbol{l}_{n_r} + \boldsymbol{\epsilon}_r, \tag{2}$$

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

For a sample with $\bar{r}$ networks, stack the data by defining $Y = (Y'_1, \ldots, Y'_{\bar{r}})\prime$, $X^* = (X^{*\prime}_1, \ldots, X^{*\prime}_{\bar{r}})\prime$, $\epsilon = (\epsilon'_1, \ldots, \epsilon'_{\bar{r}})\prime$, $G = D(G_1, \ldots, G_{\bar{r}})$, $G^* = D(G^*_1, \ldots, G^*_{\bar{r}})$, $\iota = D(l_{n_1}, \ldots, l_{n_{\bar{r}}})$ and $\eta = (\eta_1, \ldots, \eta_{\bar{r}})\prime$, where $D(A_1, \ldots, A_K)$ is a block diagonal matrix in which the diagonal blocks are $n_k \times n_k$ matrices A_k. For the entire sample, the model is thus:

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

In this model, ϕ^L and ϕ^S represent *the endogenous effects* (the effect of friends' outcomes on own outcomes), while γ^L and γ^S represent *the contextual effect* (the effect of friends' exogenous characteristics on own outcomes). The vector of network fixed effects $\boldsymbol{\eta}$ captures *the correlated effect* [the propensity for agents in the same network to behave similarly because they have similar unobserved characteristics or face a similar (e.g. institutional) environment].[15]

3.2. Identification and estimation

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

Reflection problem. In linear-in-means models, simultaneity in the behavior of interacting agents introduces a perfect collinearity between the expected mean outcome of the group and its mean characteristics. Therefore, it is difficult to differentiate between the effect of 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). In the standard approach, the reflection problem arises because individuals are affected by all individuals belonging to their group and by nobody outside the group. In the case of social networks, however, this is almost 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 these two values are the same unless the network is complete (i.e. everybody is linked with everybody).

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

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

First, we follow the standard approach (e.g. Bramoullé et al., 2009) of using *network fixed effects*. Network fixed effects are a remedy for the selection bias that originates from the possible sorting of individuals with similar unobserved characteristics into a network. The underlying assumption is that such unobserved characteristics are common to all the individuals within each network. This is reasonable in our case study where the networks are quite small (see Section 2).[16] In our case, this assumption further implies that such unobserved characteristics are common to both short-lived and long-lived ties, which means that there should not be much difference between the friends who are long-lived and short-lived. We collect some evidence that supports this idea. Indeed, in Section 5.2 below (Tables 7 and 8), we provide evidence showing that there are no differences between peers in Waves I and II in terms of observable characteristics, so that the link formation between

[15] As an analogy with time series models, the model in (3) can be referred to as a SARARMA(p, q) with $p = 0$ and $q = 2$, where p and q are the maximum number of spatial lags for the error and the outcome, respectively.

[16] 93% of our networks have a size below 35.

these two waves is not significantly different. As a consequence, it seems reasonable to also assume that the influence of unobservable factors is the same for short and long-lived ties.

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

4. Estimation results

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

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

4.1. Long-run peer effects

Table 2 collects the estimation results of model (1), without distinguishing between long-lived and short-lived ties so that students i and j are friends, i.e. $g_{ij}=1$, if they have nominated each other in Wave I. In other words, we look at the impact of friends from Wave I on own educational attainment in Wave IV. In the first column, we report the OLS results. The other columns show the IV results, which are obtained using the IV estimators detailed in Appendix B, using an increasing set of controls. In the first panel (2SLS (1)), we only use lagged covariates (Wave I) in the specification whereas, in the second panel (2SLS (2)), we enrich the control sets with variables from Wave IV. The first-stage partial F-statistics (Stock et al., 2012; Stock and Yogo, 2005) reveal that our instruments are quite informative and the OIR test provides evidence in line with their validity. The first stage results are shown in Table 3. Our identification comes from the assumption that the characteristics of friends-of-friends affect own behavior only through their effects on friends' behavior. Table 3 shows that the more relevant friends-of-friends' characteristics in our application are parental education, grade level and math performance. To test whether the characteristics of friends of friends are balanced between Wave I (population) and Wave IV (sample) students, we perform a battery of balance tests. They are reported in Table 4. It appears that none of the differences is statistically significant. These results could be taken as evidence of non-systematic attrition (and hence random treatment). In Section 6.2, we check the robustness of our analysis with respect to network topology misspecification.

The results in Table 2 reveal that the effect of friends' education on own education is always significant and positive, suggesting that 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 own future educational attainment, 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,[17] in a group of two friends, a standard deviation increase in the years of education of the friend translates into a roughly 5.4% 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% increase of a standard deviation in the individual's educational attainment (roughly seven more months of education). This is a non-negligible effect, especially given our long list of controls and the fact that friendship networks might have changed over time. The influence of peers at school seems to be carried over time.

4.2. The role of long-lived ties

We would now like to determine how long-lived and short-lived ties affect educational choices by estimating the magnitude of ϕ^L and ϕ^S in Eq. (1). Table 5 displays the estimation results of Eq. (1).[18] We find that the educational choices of short-lived friends have no significant impact on individual educational outcomes (years of schooling) while the educational choices of long-lived friends do have a positive and significant effect on educational outcomes. In terms of magnitude, a standard deviation increase in aggregate years of education of peers nominated both in Waves I and II (long-lived friends) translates into roughly a 21% increase of a standard deviation in the individual's educational attainment (roughly 8.3 more months of education). In an average group of four best friends (linked to each other in a network), a standard deviation increase of each of the peers translates into two more years of education. This is quite an important effect. It suggests that *long-lived friends* rather than *short-lived friends* matter for educational outcomes in the long run.[19] Table 5 also shows that

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

[18] 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 results when using the alternative estimators in Appendix B remain qualitatively unchanged. The latter are available upon request.

[19] When estimating Eq. (1) including only long-lived ties (i.e. $\boldsymbol{G}^S=0$), we obtain comparable results.

Table 2
Long-run peer effects.

Dep. var. years of education	OLS	2SLS (1)			2SLS (2)		
		Finite IV	Many IV	Bias corrected	Finite IV	Many IV	Bias corrected
Peer effects (ϕ)	0.0406***	0.0053***	0.0049***	0.0049***	0.0064***	0.0058***	0.0059***
	(0.0138)	(0.0020)	(0.0019)	(0.0019)	(0.0027)	(0.0020)	(0.0020)
Female	0.9571***	0.9830***	0.9847***	0.9847***	0.7383	1.0434***	1.0435***
	(0.2038)	(0.1981)	(0.1981)	(0.1981)	(0.6124)	(0.2409)	(0.2409)
Black or African American	0.6035**	−0.1295	−0.1286	−0.1287	−0.2933	−0.1833	−0.1831
	(0.2823)	(0.4088)	(0.4087)	(0.4087)	(0.7723)	(0.4464)	(0.4464)
Other races	0.0150	−0.3730	−0.3728	−0.3728	−0.3874	−0.2930	−0.2927
	(0.2706)	(0.2917)	(0.2917)	(0.2917)	(0.4561)	(0.3111)	(0.3111)
Religion practice	−0.1344**	0.2770***	0.2775***	0.2775***	0.1289	0.2521***	0.2521***
	(0.0548)	(0.0539)	(0.0539)	(0.0539)	(0.1373)	(0.0599)	(0.0600)
Household size	0.2343***	0.0325	0.0327	0.0327	0.0247	0.0355	0.0356
	(0.0566)	(0.0576)	(0.0576)	(0.0576)	(0.0900)	(0.0606)	(0.0606)
Parent education	0.9074***	0.3213***	0.3217***	0.3217***	0.2275*	0.2263***	0.2262***
	(0.0888)	(0.0971)	(0.0971)	(0.0971)	(0.1528)	(0.1052)	(0.1052)
Mathematics score A	2.4166***	1.4439***	1.4461***	1.4460***	0.8917*	1.2389***	1.2385***
	(0.2549)	(0.2549)	(0.2549)	(0.2549)	(0.5065)	(0.2817)	(0.2817)
Mathematics score B	1.8623***	1.0180***	1.0172***	1.0172***	0.6889*	0.8942***	0.8941***
	(0.2423)	(0.2407)	(0.2407)	(0.2407)	(0.3966)	(0.2532)	(0.2533)
Mathematics score C	1.5210***	0.5405**	0.5419**	0.5419**	0.3423	0.5110**	0.5108**
	(0.2600)	(0.2593)	(0.2593)	(0.2593)	(0.4020)	(0.2709)	(0.2709)
Mathematics score missing	0.7269	0.5854	0.5849	0.5849	1.1488	0.6679	0.6677
	(0.4486)	(0.4305)	(0.4305)	(0.4305)	(0.8356)	(0.4471)	(0.4472)
Resid. building qual.	0.0085	0.2059**	0.2082**	0.2081**	0.1151*	0.1765*	0.1863*
	(0.0953)	(0.096)	(0.0967)	(0.0967)	(0.0669)	(0.1053)	(0.1053)
Student grade	1.1538***	0.4720***	0.4707***	0.4707***	0.4051***	0.5042***	0.5044***
	(0.0445)	(0.0853)	(0.0852)	(0.0852)	(0.1654)	(0.0914)	(0.0914)
Children					−1.5108	−1.2429**	−1.2438**
					(1.7632)	(0.6263)	(0.6263)
Religion practice (Wave IV)					1.0470	0.3429**	0.3428**
					(1.1265)	(0.1874)	(0.1874)
Married					−0.8750	−0.6021	−0.6024
					(2.1134)	(0.6382)	(0.6382)
Parental occupation dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contextual effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Network fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First stage *F* statistic		10.8882	5.1647		9.5496	5.1280	
OIR test *p*-value		0.1927	0.2603		0.4765	0.3882	
Observations	1,819	1,819	1,819	1,819	1,819	1,819	1,819
Networks	116	116	116	116	116	116	116

Notes: robust standard errors in parentheses.
Only lagged variables are used as instruments. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

the characteristics of short-lived ties are never statistically significant, whereas long-lived peers' parental education, race, religion practice, neighborhood quality and fertility decisions show significant correlations with own educational outcome.

5. Understanding the mechanisms

Our empirical results displayed in Table 5 suggest that the distinction between long-lived and short-lived friends is important for understanding long-run peer effects in education. In Section 5.1, we propose a simple theoretical model that may explain this evidence. The idea underlying the theoretical mechanism is that convergence of preferences and formation of social norms require long-term relationships between peers. In Section 5.2, we provide additional empirical evidence ruling out alternative explanations.

5.1. Theoretical framework

In order to understand how long-lived and short-lived ties influence long-run educational outcomes, we extend the DeGroot (1974) model as follows.[20] Consider a society consisting of a finite set of individuals $N = \{1, 2, \ldots, n\}$ who are linked

[20] Appendix C.1 contains the technical details of the DeGroot model.

Table 3
Long-run peer effects – 2SLS first stage results.

Dependent variable: *GY*			
Variables: *X*	*X* Own	*GX* Peers	G^2X Peers of peers (exclusion restrictions)
Female	0.4516	0.7482**	−0.2222
	(0.4456)	(0.3483)	(0.1738)
Black or African American	−1.1913*	−0.3352	0.1265
	(0.6487)	(0.4546)	(0.1264)
Other races	0.1140	0.7957*	0.0356
	(0.6163)	(0.4818)	(0.2197)
Religion practice	−0.1296	−0.0112	−0.0007
	(0.1202)	(0.0891)	(0.0431)
Household size	−0.2091	0.1924**	0.0392
	(0.1355)	(0.0948)	(0.0458)
Parent education	−0.3902*	0.6201***	0.1962***
	(0.1966)	(0.1520)	(0.0625)
Mathematics score A	−1.5474*	2.7857***	1.2068***
	(0.5663)	(0.4119)	(0.1749)
Mathematics score B	−1.0044*	2.4891***	0.9199***
	(0.5545)	(0.3843)	(0.1756)
Mathematics score C	−1.0947*	1.8637***	0.9009***
	(0.5913)	(0.4270)	(0.2005)
Mathematics score missing	−0.0857	1.8655***	0.4239
	(0.9780)	(0.7666)	(0.3482)
Resid. building qual.	−0.0623	−0.2335	−0.0974
	(0.1958)	(0.1545)	(0.0659)
Student grade	−0.5640***	1.0504***	−0.1511***
	(0.1559)	(0.0838)	(0.0263)
Network fixed effects		Yes	
Number of observations		1,819	
Number of networks		116	

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

Table 4
Friend-of-friends characteristics – balance tests.

Sample	Wave I Mean (std)	Wave IV Mean (std)	Difference [*p*-value]
Female	0.502	0.522	[0.511]
	(0.500)	(0.500)	
Black or African American	0.165	0.160	[0.496]
	(0.372)	(0.367)	
Other races	0.102	0.092	[0.490]
	(0.303)	(0.289)	
Religion practice	3.891	3.922	[0.505]
	(1.816)	(1.803)	
Household size	3.401	3.408	[0.501]
	(1.382)	(1.376)	
Parent education	3.231	3.229	[0.500]
	(0.952)	(0.942)	
Mathematics score A	0.257	0.265	[0.505]
	(0.437)	(0.441)	
Mathematics score B	0.319	0.320	[0.501]
	(0.466)	(0.467)	
Mathematics score C	0.223	0.217	[0.496]
	(0.417)	(0.412)	
Mathematics score missing	0.077	0.073	[0.496]
	(0.267)	(0.261)	
Residential building quality	1.557	1.554	[0.499]
	(0.808)	(0.808)	
Student grade	9.453	9.425	[0.495]
	(1.647)	(1.645)	
Observations	2,341	1,819	

Notes: *t*-tests for differences in means are performed. *p*-values are reported. Differences are computed w.r.t. the larger sample in the previous column.

Table 5
Long and short-lived ties: endogenous and exogenous long-run peer effects.

Dep. var. years of education		2SLS (1)		2SLS (2)	
		Long-lived ties	Short-lived ties	Long-lived ties	Short-lived ties
Endogenous effects		0.0410***	0.0060	0.0345***	0.0080
		(0.0139)	(0.0056)	(0.0155)	(0.0063)
Exogenous effects	Female	−0.2371	0.1667	−0.1011	−0.0838
		(0.3443)	(0.3149)	(0.4015)	(0.4168)
	Black or African American	−0.8318*	−0.5801	−1.1855**	−0.5474
		(0.4456)	(0.3965)	(0.6010)	(0.5313)
	Other races	−0.5138	−0.6105	−0.4395	−0.4940
		(0.4511)	(0.4415)	(0.4844)	(0.4593)
	Religion practice	0.1448*	−0.0013	0.1849*	0.0046
		(0.0898)	(0.0866)	(0.1012)	(0.1058)
	Household size	0.1159	−0.0076	0.1194	−0.0098
		(0.0923)	(0.0836)	(0.0986)	(0.0868)
	Parent education	0.2368**	−0.0258	0.4310***	0.0088
		(0.1237)	(0.1409)	(0.1799)	(0.1617)
	Mathematics score A	−0.3064	0.4166	−0.5945	0.2195
		(0.4174)	(0.3910)	(0.4577)	(0.4260)
	Mathematics score B	0.0639	0.2386	0.0336	0.1960
		(0.3860)	(0.3639)	(0.3980)	(0.3794)
	Mathematics score C	0.0939	0.3332	−0.0195	0.1886
		(0.4187)	(0.4120)	(0.4415)	(0.4379)
	Mathematics score missing	−1.1887	−0.7823	−1.0367	−0.6235
		(0.7444)	(0.7210)	(0.7972)	(0.7482)
	Resid. building qual.	0.4558***	−0.0949	0.3357**	−0.1316
		(0.1480)	(0.1501)	(0.1598)	(0.1646)
	Student grade	0.0269	−0.0178	0.1311	−0.0321
		(0.0764)	(0.0701)	(0.1046)	(0.0886)
	Children			−2.1908**	0.5665
				(1.0434)	(0.9400)
	Religion practice (Wave IV)			0.3068	0.0156
				(0.3109)	(0.2959)
	Married			−0.1707	0.3332
				(1.0073)	(0.8907)
Individual socio-demographic		Yes	Yes	Yes	Yes
Family background		Yes	Yes	Yes	Yes
Network fixed effects		Yes	Yes	Yes	Yes
Observations		1,819	1,819	1,819	1,819
Networks		116	116	116	116

Notes: we report bias-corrected 2SLS estimates. Robust standard errors in parentheses.
See Table 2. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

in a directed network and who would like to gather information about an unknown parameter θ. In our context, assume that there are two states of the world so that θ can be equal to either: {It is worth continuing studying} or {It is not worth continuing studying}. What is key in the DeGroot model is that agents update their beliefs by repeatedly taking *weighted averages* of their neighbors' beliefs (where neighbors are the people directly linked to each individual) with p_{ij} being the weight that agent i places on the current belief of agent j in forming his or her belief for the next period. If the network is strongly connected and at least some individuals listen to themselves, then, in the limit, everybody belonging to the same network will converge to a consensus and the influence of each person will depend on her position in the network (see Proposition 4 in Appendix C). This means that, if there are *repeated interactions* between students from the same network, then they will all adopt the same social norm, which could be either {It is worth continuing studying} or {It is not worth continuing studying} depending on what the "influential" students believe.[21]

We use this theoretical framework to understand why, in our empirical results, short-lived friends have no impact on own education decision while long-lived friends have an impact. Indeed, there are two types of friend relationships between students in a network $\boldsymbol{G}$: *short-lived* friendships ($l = S$) and *long-lived* friendships ($l = L$). As previously stated, we define *short-lived* friends as students who interact with each other only *once* while *long-lived* friends are students who interact *for a longer time*. Quite naturally, we assume that each agent has a long-lived relationship with him/herself, i.e. $g_{ii}^L = 1$ and $g_{ii}^S = 0$.

[21] Because students have parents with different incomes or students have different costs of studying (some like to study while others do not), we can explain why, within a network with the same social norm, students make different decisions concerning the number of years they will spend in college. In particular, the students whose parents have low income or students who have a high disutility of studying may not go to college even if the social norm in their network of friends says that it is worth continuing studying.

As in Section 3, this implies two types of adjacency matrices: $\mathbf{G}^L = \{g_{ij}^L\}$, where $g_{ij}^L = 1$ if i and j are long-lived friends and $\mathbf{G}^S = \{g_{ij}^S\}$, where $g_{ij}^S = 1$ if i and j are short-lived friends, with $\mathbf{G}^L + \mathbf{G}^S = \mathbf{G}$. Denote by $\tilde{\mathbf{G}}^L$ and $\tilde{\mathbf{G}}^S$ the row-normalized matrices of $\mathbf{G}^L$ and $\mathbf{G}^S$, respectively. Because *short-lived* friendships only interact once, they will influence the beliefs of each other only in the initial period. The updating stops there since *short-lived* friends do not meet anymore and thus students only update their beliefs once. On the contrary, *long-lived* friends interact repeatedly and thus update their beliefs continuously as in the standard DeGroot model. Therefore, all students in the network will reach a consensus after many interactions, even though only long-lived friends will matter in the long run.

How do we solve this model? In the first period, both short-lived and long-lived friends influence each other so that $\mathbf{b}^{(1)} = \tilde{\mathbf{G}}\mathbf{b}^{(0)}$, where $\mathbf{b}^{(t)}$ is the vector of beliefs of *all students* at time t (i.e. both short-lived and long-lived students) and $\tilde{\mathbf{G}}$ is the row-normalized matrix of $\mathbf{G}$. Now relabel $\mathbf{b}^{(1)}$ as $\tilde{\mathbf{b}}^{(0)}$, i.e. $\tilde{\mathbf{b}}^{(0)} := \mathbf{b}^{(1)}$. We are now in the framework of the DeGroot model where the initial beliefs are given by $\tilde{\mathbf{b}}^{(0)}$. As a result, we can apply Proposition 4 in Appendix C, which implies that, if the network $\tilde{\mathbf{G}}^L$ is strongly connected and if at least some agents pay attention to themselves, i.e. $g_{ii} > 0$ for some i, then all students will reach a consensus in the long run, which is determined by:

$$\mathbf{b}^{\infty} = \lim_{t\to\infty}\left(\tilde{\mathbf{G}}^L\right)^t \tilde{\mathbf{b}}^{(0)} = \lim_{t\to\infty}\left(\tilde{\mathbf{G}}^L\right)^t \tilde{\mathbf{G}}\mathbf{b}^{(0)}. \tag{4}$$

In Eq. (4), we see clearly the distinct influence of short-lived and long-lived friends. The updating matrix $\tilde{\mathbf{G}}^L$ only depends on long-lived friends because they interact over time to reach a consensus. However, the initial beliefs are a function of the beliefs of both short-lived and long-lived friends since $\tilde{\mathbf{G}}$ includes both $\tilde{\mathbf{G}}^L$ and $\tilde{\mathbf{G}}^S$. In Eq. (4), we assume that all students have both long-lived and short-lived friends[22] so that they are all included in the convergence process $(\mathbf{b})^{\infty}$. In Appendix C.2, we provide an example where we calculate the consensus with short-lived and long-lived ties and show that this consensus is different compared to the case where all friends are long-lived ties.

As a result, a possible interpretation of our evidence is that the *strength of interactions* between two students may affect how much they learn, their human capital accumulation and how much they value achievement. It also shapes social norms that accumulate over time, which affect years of schooling both directly and indirectly. This idea is related to Akerlof's and Kranton's (2002) concept of identity in economics, in which learning in 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.[23]

Observe that the aim of the model is to give some economic intuition of why different types of links (and thus friends) have different impacts on long-run outcomes (steady-state). We are not directly testing this model in the data, i.e. Eq. (1) is not a reduced form of the model presented in this section. For example, in our data, short-lived links live only one period while long-lived links live two periods. In our model, long-lived links last an infinite number of periods while short-lived links last one period.[24]

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

5.2. Additional evidence

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

In our analysis, we identified long-lived ties as peers nominated in both Wave I and Wave II. This definition implies that long-lived friends are more likely to be peers at the time of college decisions. One could thus put forward another explanation for why only long-lived ties influence education decisions: it could be *decision proximity*, so that friends in later grades (grades 10–12) are more likely to impact college decisions and also are more likely to be long-lived ties. In other words, is it really the strength of social interactions or is it the timing of friendship formation that is crucial for future educational outcomes?

---

[22] Observe that if some students have only short-lived friends, then there will be no consensus. Indeed, even if just one student i has only short-lived friends, then $g_{ii}^L = 1$ and $g_{ij}^L = 0$ for all other j. Hence, the network $\tilde{\mathbf{G}}^L$ will not be strongly connected, and the baseline deGroot model will not work properly (i.e. depending on the network structure there will either be more than one component with a different "consensus" each, or no consensus at all).

[23] 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.

[24] One way to make the model closer to the data is to assume that long-lived friends (i.e. those nominated in Waves I and II) are still friends in Wave IV while short-lived ones are not. This seems quite reasonable and would then imply that short-lived links live only one period while long-lived links live an infinite number of periods.

Table 6
Heterogeneous endogenous peer effects in grades 10–12.

Dep. var. years of education		
	2SLS	2SLS
Long-lived ties (ϕ^L)	0.0452***	0.0485**
	(0.0191)	(0.0212)
Earlier short-lived ties (ϕ^{S_1})	0.0037	0.0027
	(0.0241)	(0.0331)
Later short-lived ties (ϕ^{S_2})	0.0003	0.0049
	(0.0224)	(0.0257)
Individual socio-demographic	Yes	Yes
Family background	Yes	Yes
Residential neighborhood	Yes	Yes
Contextual effects	Yes	Yes
Network fixed effects	Yes	Yes
Wave IV controls	No	Yes
Observations	628	628
Networks	41	41

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

We would therefore like to disentangle the *decision proximity* effect from the *strength of interaction* effect. To do this, we select Wave II students in the later grades (grades 10–12) and distinguish between different types of short-lived ties. We estimate a modified version of model (1)

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

This model disentangles the effects of short-lived ties who have been nominated in the past (Wave I), i.e. $\phi^S=\phi^{S_1}$, from the effect of short-lived ties who have been nominated at the time when college decisions are made (Wave II), i.e. $\phi^S=\phi^{S_2}$. If the decision proximity matters, then coefficient ϕ^{S_2} should be significant while ϕ^{S_1} should not.

Table 6 contains the estimation results. The empirical results reveal that the educational decisions of short-lived ties continue to show a non-significant effect on individual outcomes, regardless of whether peers interact in lower or higher grades, highlighting the crucial role of long-lived ties in college decision.

Another concern is that peers nominated in different time periods may have a different long-run effects 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 7 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 there are no significant differences between these peers in terms of observable characteristics.

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

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

In this model, $g_{ij,r,t}=1$ if there is a link between i and j belonging to network r at time t (where t = Wave I, Wave II), $x^m_{i,r,t}$ 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 γ^m captures the differences between the importance of these characteristics in link formation between Wave I and Wave II. Estimating Eq. (5), Table 8 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 also perform an F-test that tests the joint significance of the γ parameters.[25] Table 8 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,

[25] The idea is similar to the *Chow test* in time series analysis to investigate the existence of a structural break (see e.g. Chow, 1960; Hansen, 2000, 2001).

Table 7
Peer characteristics – comparison between different types of peers.

Ties	Short-lived		Long-lived	Differences		
	Wave I Mean (std)	Wave II Mean (std)	Wave I and II Mean (std)	$\Delta_{WI,WII}$ [*p*-value]	$\Delta_{WII,WI\&II}$ [*p*-value]	$\Delta_{WI,WI\&II}$ [*p*-value]
Years of education	14.963 (3.483)	14.616 (3.352)	14.907 (3.390)	[0.471]	[0.476]	[0.495]
Female	0.483 (0.500)	0.428 (0.496)	0.473 (0.500)	[0.469]	[0.474]	[0.495]
Black or African American	0.064 (0.245)	0.013 (0.115)	0.043 (0.204)	[0.426]	[0.449]	[0.474]
Other races	0.093 (0.291)	0.057 (0.233)	0.053 (0.225)	[0.462]	[0.495]	[0.457]
Religion practice	3.983 (1.823)	3.882 (1.893)	3.837 (1.861)	[0.485]	[0.493]	[0.478]
Household size	3.336 (1.373)	3.323 (1.344)	3.473 (1.379)	[0.497]	[0.469]	[0.472]
Parent education	3.333 (0.901)	3.219 (0.828)	3.193 (0.832)	[0.463]	[0.491]	[0.455]
Mathematics score A	0.235 (0.425)	0.212 (0.410)	0.250 (0.434)	[0.484]	[0.475]	[0.490]
Mathematics score B	0.277 (0.448)	0.316 (0.466)	0.307 (0.462)	[0.476]	[0.494]	[0.482]
Mathematics score C	0.248 (0.432)	0.242 (0.429)	0.237 (0.426)	[0.497]	[0.496]	[0.493]
Mathematics score missing	0.078 (0.269)	0.084 (0.278)	0.070 (0.256)	[0.494]	[0.485]	[0.491]
Residential building quality	1.556 (0.775)	1.589 (0.788)	1.523 (0.786)	[0.479]	[0.455]	[0.478]
Student grade	10.439 (0.497)	10.434 (0.497)	10.403 (0.491)	[0.500]	[0.500]	[0.500]
Children	0.478 (0.500)	0.478 (0.500)	0.453 (0.499)	[0.500]	[0.500]	[0.500]
Religion practice (Wave IV)	1.436 (1.625)	1.320 (1.556)	1.457 (1.661)	[0.500]	[0.500]	[0.500]
Married	0.498 (0.501)	0.498 (0.501)	0.493 (0.501)	[0.500]	[0.497]	[0.498]

Notes: *t*-tests for differences in means are performed. *p*-values are reported.

Table 8
Network formation in Wave I and Wave II OLS estimation results.

Variable	γ coefficient	Std. error
Female	0.0025	(0.019)
Black or African American	−0.0107	(0.044)
Other races	−0.0265	(0.036)
Religion practice	0.0032	(0.008)
Household size	0.0125	(0.008)
Parent education	−0.0073	(0.013)
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)
Residential building quality	0.0085	(0.012)
Student grade	−0.0226	(0.016)
Network fixed effects	Yes	
F-test *p* value	0.6083	
Observations	6,932	

Notes: we report bias-corrected 2SLS estimates. Robust standard errors in parentheses.

Tables 7 and 8 provide evidence showing that there are no significant 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.

Finally, we investigate whether there are 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 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 9. It appears that the two networks are topologically very similar.

Table 9
Network structure – comparison between Wave I and Wave II.

Network structure indicators	Wave I	Wave II
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.

6. Robustness checks

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

6.1. Endogenous network formation

Our identification strategy hinges on the use of *network fixed effects* to control for unobserved factors driving both network formation and behavior in networks. If there are student-level unobservables that drive both peer choice and outcome choice, this strategy fails. A possible way to tackle this issue is to *simultaneously* estimate network formation and outcomes. This strategy can be pursued by using parametric modeling assumptions and Bayesian inferential methods that allow integrating a network formation with the study of behavior over the formed networks. Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2015) propose two slightly different ways to implement this approach. In Goldsmith-Pinkham and Imbens (2013), unobservables are dichotomous and only one network is considered. As we have multiple networks in our data, we follow Hsieh and Lee (2015).[26] They present a model with one peer type. We implement an extension of their method for heterogeneous peer effects (short-lived and long-lived ties). If there is an unobservable characteristic that drives the choice of long-lived ties[27] and is correlated with $\epsilon_{i,r}$, then $g^L_{ij,r}$ is endogenous and estimates of model (1) are biased. By failing to account for similarities in (unobserved) characteristics, similar behaviors might mistakenly be attributed to peer influence when they simply result from similar characteristics. Let $z_{i,r}$ denote such an *unobserved characteristic*, which influences the link formation process. Let us also assume that $z_{i,r}$ is correlated with $\epsilon_{i,r}$ in model (1) according to a bivariate normal distribution

$$(z_{i,r}, \epsilon_{i,r}) \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_z^2 & \sigma_{\varepsilon z}\\ \sigma_{\varepsilon z} & \sigma_\varepsilon^2\end{pmatrix}\right).$$

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

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

and

$$P(g_{ij,r,t} = 1 | x_{ij,r}, z_{i,r}, z_{j,r}, g_{ij,r,t-1}, \gamma_t, \theta_t, \lambda) = \Lambda\left(\gamma_{0,t} + \lambda g_{ij,r,t-1} + \sum_k |x_{i,r} - x_{j,r}|\gamma_{k,t} + |z_{i,r} - z_{j,r}|\theta_t\right), \quad (7)$$

where $\Lambda(\cdot)$ is a logistic function. Homophily behavior in the unobserved characteristics implies that $\theta_\tau < 0$, where $\tau = t-1, t$, which means that the closer two individuals are in terms of unobservable characteristics, the higher is the probability that they are friends. The same argument holds for observables. If $\sigma_{\varepsilon z}$ and θ_τ are different from zero, then networks $g^L_{ij,r}$ and $g^S_{ij,r}$ in model (1) are endogenous. From models (6) and (7), the probability of observing a short-lived tie is then given by:

$$\begin{aligned} P(g^S_{ij,r} &= 1 | x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, \gamma_{t-1}, \theta_{t-1}) = P(g_{ij,r,t} = 1 | x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, g_{ij,r,t-1} = 0) \times P(g_{ij,r,t-1} \\ &= 0 | x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}) + P(g_{ij,r,t} = 0 | x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, g_{ij,r,t-1} = 1) \times P(g_{ij,r,t-1} \\ &= 1 | x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}) \end{aligned}$$

[26] Another difference between those two procedures is that Goldsmith-Pinkham and Imbens (2013) set the same unobservables in both link formation and outcome equation while Hsieh and Lee (2015) use different unobservables for those equations and let them be correlated.
[27] The reasoning is the same for short-lived ties.

whereas the probability of observing a long-lived tie is equal to:

$$P(g^L_{ij,r} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, \gamma_{t-1}, \theta_{t-1}) = P(g_{ij,r,t} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_t, \theta_t, \lambda, g_{ij,r,t-1} = 1)$$
$$\times P(g_{ij,r,t-1} = 1|x_{ij,r}, z_{i,r}, z_{j,r}, \gamma_{t-1}, \theta_{t-1}).$$

In this way, we have modeled the probability of being a long-lived or short-lived tie including unobservables that are allowed to be correlated with the error term in the outcome equation.[28] Joint normality implies $E(\epsilon_{i,r}|z_{i,r}) = (\sigma_{\varepsilon z}/\sigma^2_z)z_{i,r}$, when conditioning on $z_{i,r}$. Hence, the outcome equation is

$$y_{i,r} = \phi^L \sum_{j=1}^{n_r} g^L_{ij,r} y_{j,r} + \phi^S \sum_{j=1}^{n_r} g^S_{ij,r} y_{j,r} + x'_{i,r}\beta + \frac{1}{g^L_{i,r}} \sum_{j=1}^{n_r} g^L_{ij,r} x'_{j,r}\delta^L + \frac{1}{g^S_{i,r}} \sum_{j=1}^{n_r} g^S_{ij,r,t} x'_{j,r}\delta^S + \eta_r + \frac{\sigma_{\varepsilon z}}{\sigma^2_z} z_{i,r} + u_{i,r}, \tag{8}$$

where $u_{i,r} \sim N(0, \sigma^2_z - (\sigma^2_{\varepsilon z}/\sigma^2_z))$. Note that if no correlation exists ($\sigma_{\varepsilon z} = 0$), then estimating Eq. (8) or (1) is equivalent. Given the complexity of this framework, it is convenient to simultaneously estimate the parameters of Eqs. (6)–(8) with a Bayesian approach. Bayesian inference requires the computation of marginal distributions for all parameters. However, since this requires integration of complicated distributions over several variables, a closed form solution is not readily available and Markov Chain Monte Carlo (MCMC) techniques are usually employed to obtain random draws from posterior distributions. The unobservable variable ($z_{i,r}$) is thus generated according to the joint likelihood of link formation and outcome; it is drawn in each MCMC step together with the parameters of the model. The Gibbs sampling algorithm allows us to draw random values for each parameter from their posterior marginal distribution, given previous values of other parameters. Once stationarity of the Markov Chain has been achieved, the random draws can be used to study the empirical distributions of the posterior.[29]

The extended model (6)–(8) is more demanding than model (1) in terms of identification conditions. Identification in the baseline model (1) rests on the exogeneity of the $\boldsymbol{X}$ variables and on the presence of intransitivities in the exogenous network topology as captured by the matrix $\boldsymbol{G}$ (Bramoullé et al., 2009). As a matter of fact, $\boldsymbol{G}^2\boldsymbol{X}$ is the exclusion restriction. Following Hsieh and Lee (2015) and Goldsmith-Pinkham and Imbens (2013), identification in the extended model (6)–(8) requires an additional source of exogenous variation through absolute values of differences $|x_i - x_j|$. Indeed, in the extended model (6)–(8), the dyad-specific regressors used in the network formation model are naturally excluded from the outcome equation (Hsieh and Lee, 2015). As a consequence, covariates affect links through absolute values of differences $|x_i - x_j|$, while they affect outcomes directly. This constitutes a form of exclusion restriction that relies on nonlinearities (Hsieh and Lee, 2015).

Table 10 panel (b) collects the results that are obtained when estimating Eqs. (8), (6) and (7) simultaneously. Panel (a) shows the estimation results of the model without distinguishing between short-lived and long lived ties (homogeneous peer effects). The first column in both panels reports the 2SLS results for comparison. Table 10 reveals that $\sigma_{\varepsilon z}$ is not significantly different from zero for both the models with homogeneous and heterogeneous peer effects (columns (3) and (6) of Table 10). The Bayesian estimates are close to the 2SLS estimates.[30] This evidence is thus in support of our identification strategy. Indeed, our list of controls and network fixed effects, together with the temporal lag between when friends are chosen and when education levels are attained, seem to account for unobserved factors driving both network formation and behavior over networks.

6.2. *Measurement errors in network topology*

6.2.1. *Directed networks*

Our empirical investigation has assumed that friendship relationships are symmetric, i.e. $g_{ij} = g_{ji}$. We now check the sensitivity of our results 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% of the relationships in our dataset are not reciprocal. 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.[31] Table 11 shows the estimation results of model (1) for directed networks. The results remain qualitatively unchanged and are only slightly higher in magnitude.

---

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

[29] See Appendix E for more details on the estimation procedure. An introduction to Monte Carlo methods in Bayesian econometrics can be found in Chib (1996) and Casella and Robert (2004).

[30] Those estimates are slightly different from those collected in Table 5 because the computational burden of Bayesian estimation forced us to drop small networks, since they create computational problems when some covariates are constant across the same network, and big networks, because they slow the computation time excessively.

[31] As highlighted by Wasserman and Faust (1994), centrality indices for directional relationships generally focus on choices made (outdegrees). The estimation results, however, remain qualitatively unchanged if we define the link using the nominations received (indegrees).

Table 10
Robustness check: endogenous network formation.

Dependent variable: years of education						
	(a) Homogeneous peer effects			(b) Heterogeneous peer effects		
	2SLS	Bayesian estimation		2SLS	Bayesian estimation	
	Outcome eq. without link form. (1)	Link form. (2)	Outcome eq. with link form. (3)	Outcome eq. without link form. (4)	Link form. (5)	Outcome eq. with link form. (6)
Peer effects (ϕ)	0.0045***		0.0040**			
	(0.0017)		(0.0016)			
Long-lived ties (ϕ^L)				0.0097**		0.0087**
				(0.0042)		(0.0044)
Short-lived ties (ϕ^S)				0.0020		0.0017
				(0.0021)		(0.0021)
Unobservables						
$\sigma_{\epsilon z}$			0.2303			−0.1524
			(0.1887)			(0.0933)
θ		−3.6833***				
		(1.4227)				
$\theta_{t-1=1}$					0.7942***	
					(0.1262)	
$\theta_{t=2}$					0.5214***	
					(0.1168)	
Individual socio-demographic	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes
Residential neighborhood	Yes	Yes	Yes	Yes	Yes	Yes
Contextual effects	Yes	Yes	Yes	Yes	Yes	Yes
Network fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	932	433,846	932	932	433,846	932
Networks	33	33	33	33	33	33

Notes: columns (2), (3), (5) and (6) report the means and the standard deviations of the posterior distributions of the parameters. We let our chain run for 200,000 iterations, discarding the first 20,000 iterations. Ergodicity of the Markov chain is achieved. ** $p < 0.05$; *** $p < 0.01$.

Table 11
Heterogeneous endogenous peer effects – directed networks.

Dep. var. years of education		
	2SLS (1)	2SLS (2)
Long-lived ties (ϕ^L)	0.0409***	0.0474***
	(0.0138)	(0.0183)
Short-lived ties (ϕ^S)	0.0043	0.0052
	(0.0071)	(0.0070)
Individual socio-demographic	Yes	Yes
Family background	Yes	Yes
Residential neighborhood	Yes	Yes
Contextual effects	Yes	Yes
Network fixed effects	Yes	Yes
Observations	1,819	1,819
Networks	116	116

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

6.2.2. Link misspecification

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 ties. Indeed, our empirical analysis finds 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 to perfectly identify long-lived and short-lived ties. To be more precise, we use simulated data to answer the following questions: Do our results change if some links are not assigned to the correct 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 $p > q$, and reports another individual (neither a short-lived nor a long-lived tie, $l = N$), i.e. an unconnected individual, with probability r, with $r < q < p$.

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

True	Observed			
	L	S	N	
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_{KM}, which are the probabilities of moving from state K to state M, with $K, M = \{S, L, N\}$. In our numerical exercise, we gradually change those elements from 0 to 1 at a pace of 0.005, i.e. $P_{KM} = [0, 0.005, 0.010, \ldots]$.

Our misspecification experiment can be summarized by the following table:

True	Observed		
	L	S	N
L	·	P_{LS}	P_{LN}
S	P_{SL}	·	P_{SN}
N	P_{NL}	P_{NS}	·

For ease of computation, we proceed in two steps. First, we change ties from long-lived to short-lived and vice versa, i.e. we change

$$P_{LS} = \frac{2p(1-p)}{p^2 + 2p(1-p)} \quad \text{and} \quad P_{SL} = \frac{q^2}{q^2 + 2q(1-q)}.$$

Second, for each combination of P_{LS} and P_{SL}, we change ties to non-ties and vice versa, i.e. we change

$$P_{LN} = \frac{(1-p)^2}{p^2 + 2p(1-p) + (1-p)^2}, \quad P_{NL} = \frac{n^2}{n^2 + 2n(1-n) + (1-n)^2}, P_{SN} = \frac{(1-q)^2}{q^2 + 2q(1-q) + (1-q)^2}$$

$$\text{and} \quad P_{NS} = \frac{(1-n)^2}{n^2 + 2n(1-n) + (1-n)^2}.$$

In this framework, higher probabilities are associated with greater deviation 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_{NS}, we draw one hundred network structures (samples) of a size equal to the real one ($n=1,819$). Then, we estimate model (2) replacing the real $\boldsymbol{G}_r^L$ and $\boldsymbol{G}_r^S$ matrices with the simulated ones in turn so that, in total, we estimate model (2) eighty thousand times for each type of estimator described in Appendix B.[32]

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 by 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.

Table 12 displays the results of our simulation experiment for the 2SLS bias-corrected lagged estimator.[33] Fig. 1 depicts the evidence. Both Table 12 and Fig. 1 show the estimates of long-lived and short-lived tie effects with 90% confidence bands, in the upper and lower panel, respectively.[34]

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 Fig. 1 (and upper panel of Table 12) 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 results show that the long-lived tie effects remain statistically significant for levels of P_{LS} and P_{LN} in the range of 0.005 and 0.35. Fig. 2 depicts the conditional results (i.e. P_{LS} conditional on $P_{LN}=0$ in the upper panel and P_{LN} conditional on $P_{LS}=0$ in the lower panel). The upper panel shows that long-lived-tie effects remain statistically significant up to a percentage of randomly replaced links with short-lived ties of about 35%. The lower panel shows a similar result when increasing the percentage of links randomly replaced by zeros. This evidence implies that even if we do not observe or we imprecisely observe a portion of each individual's long-lived ties, our results on the existence of this effect still hold.

Our second question concerns what percentages of replacement is needed to have a significant effect of short-lived ties. The lower panel of Fig. 1 (and lower panel of Table 12) shows the estimates of the 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 results show that we need to replace almost 70% of the short-lived ties with long-lived ties before finding an effect that is statistically different from zero. 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 Fig. 2 shows this evidence more clearly by depicting the conditional results (i.e. P_{SL} conditional on $P_{SN}=0$ in the upper panel and P_{SN} conditional on $P_{SL}=0$ in the lower panel). These results show that the effects of short-lived ties are found to be important only when the large majority of long-lived ties are labeled as short-lived ties.

To summarize, in this section we have shown that the importance of long-lived ties ϕ^L is reduced and becomes insignificant when we convert more than 35% of the long-lived ties into short-lived ties. At the same time, the strength of short-lived ties ϕ^S is increasing and becomes significant when we replace more than 60% of the short-lived ties with long-lived ties. To illustrate this result, consider a student i who has 20 friends, 10 long-lived ties $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and 10 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

[32] The simulation exercise takes the network component as fixed. Bridges between two unconnected social networks are not allowed.

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

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

Table 12
Simulated evidence: ties misspecification.

Long-lived ties								
P_{LN}								
		7%	21%	35%	49%	63%	77%	91%
P_{LS}	7%	[0.008;0.037;0.066]	[0.007;0.039;0.071]	[0.004;0.041;0.078]	[−0.001;0.040;0.082]	[−0.009;0.037;0.084]	[−0.020;0.033;0.086]	[−0.041;0.016;0.073]
	21%	[0.006;0.036;0.066]	[0.003;0.037;0.070]	[0.002;0.039;0.077]	[−0.004;0.038;0.081]	[−0.014;0.034;0.083]	[−0.028;0.025;0.079]	[−0.038;0.019;0.076]
	35%	[0.001;0.032;0.062]	[−0.001;0.033;0.068]	[−0.004;0.034;0.072]	[−0.012;0.031;0.074]	[−0.018;0.030;0.077]	[−0.030;0.024;0.077]	[−0.041;0.016;0.072]
	49%	[−0.006;0.025;0.056]	[−0.008;0.026;0.060]	[−0.012;0.027;0.065]	[−0.019;0.024;0.067]	[−0.024;0.024;0.072]	[−0.035;0.018;0.072]	[−0.044;0.013;0.070]
	63%	[−0.014;0.016;0.047]	[−0.017;0.017;0.051]	[−0.021;0.017;0.055]	[−0.025;0.017;0.060]	[−0.033;0.015;0.062]	[−0.039;0.013;0.065]	[−0.048;0.010;0.067]
	77%	[−0.021;0.008;0.038]	[−0.023;0.010;0.043]	[−0.027;0.009;0.046]	[−0.031;0.010;0.051]	[−0.037;0.009;0.056]	[−0.045;0.007;0.059]	[−0.047;0.011;0.068]
	91%	[−0.026;0.002;0.030]	[−0.031;0.000;0.032]	[−0.034;0.002;0.037]	[−0.037;0.003;0.043]	[−0.044;0.001;0.047]	[−0.046;0.005;0.056]	[−0.047;0.011;0.068]
Short-lived ties								
P_{SN}								
		7%	21%	35%	49%	63%	77%	91%
P_{SL}	7%	[−0.004;0.008;0.020]	[−0.006;0.008;0.021]	[−0.009;0.007;0.023]	[−0.013;0.006;0.026]	[−0.019;0.005;0.029]	[−0.026;0.004;0.035]	[−0.031;0.005;0.042]
	21%	[−0.004;0.009;0.022]	[−0.007;0.008;0.024]	[−0.011;0.007;0.024]	[−0.016;0.006;0.027]	[−0.021;0.005;0.031]	[−0.028;0.004;0.036]	[−0.032;0.005;0.042]
	35%	[−0.004;0.010;0.024]	[−0.007;0.009;0.025]	[−0.012;0.007;0.026]	[−0.016;0.006;0.029]	[−0.020;0.007;0.033]	[−0.028;0.004;0.036]	[−0.031;0.006;0.043]
	49%	[−0.002;0.013;0.027]	[−0.005;0.012;0.029]	[−0.009;0.010;0.030]	[−0.013;0.010;0.033]	[−0.020;0.008;0.036]	[−0.024;0.008;0.041]	[−0.028;0.009;0.046]
	63%	[0.001;0.016;0.031]	[−0.002;0.015;0.033]	[−0.005;0.015;0.035]	[−0.010;0.013;0.037]	[−0.014;0.014;0.042]	[−0.023;0.010;0.043]	[−0.027;0.010;0.046]
	77%	[0.004;0.020;0.035]	[0.001;0.019;0.036]	[−0.002;0.018;0.038]	[−0.007;0.017;0.041]	[−0.012;0.016;0.045]	[−0.017;0.016;0.049]	[−0.025;0.012;0.049]
	91%	[0.008;0.023;0.038]	[0.006;0.023;0.041]	[0.002;0.022;0.042]	[−0.002;0.021;0.045]	[−0.009;0.020;0.048]	[−0.015;0.017;0.050]	[−0.024;0.013;0.049]

Notes: [10% confidence bound; mean; 90% confidence bound]. Observations = 1,819. Networks = 116. Network fixed effects, contextual effects, parental occupation dummies and individual socio-demographic characteristics included.

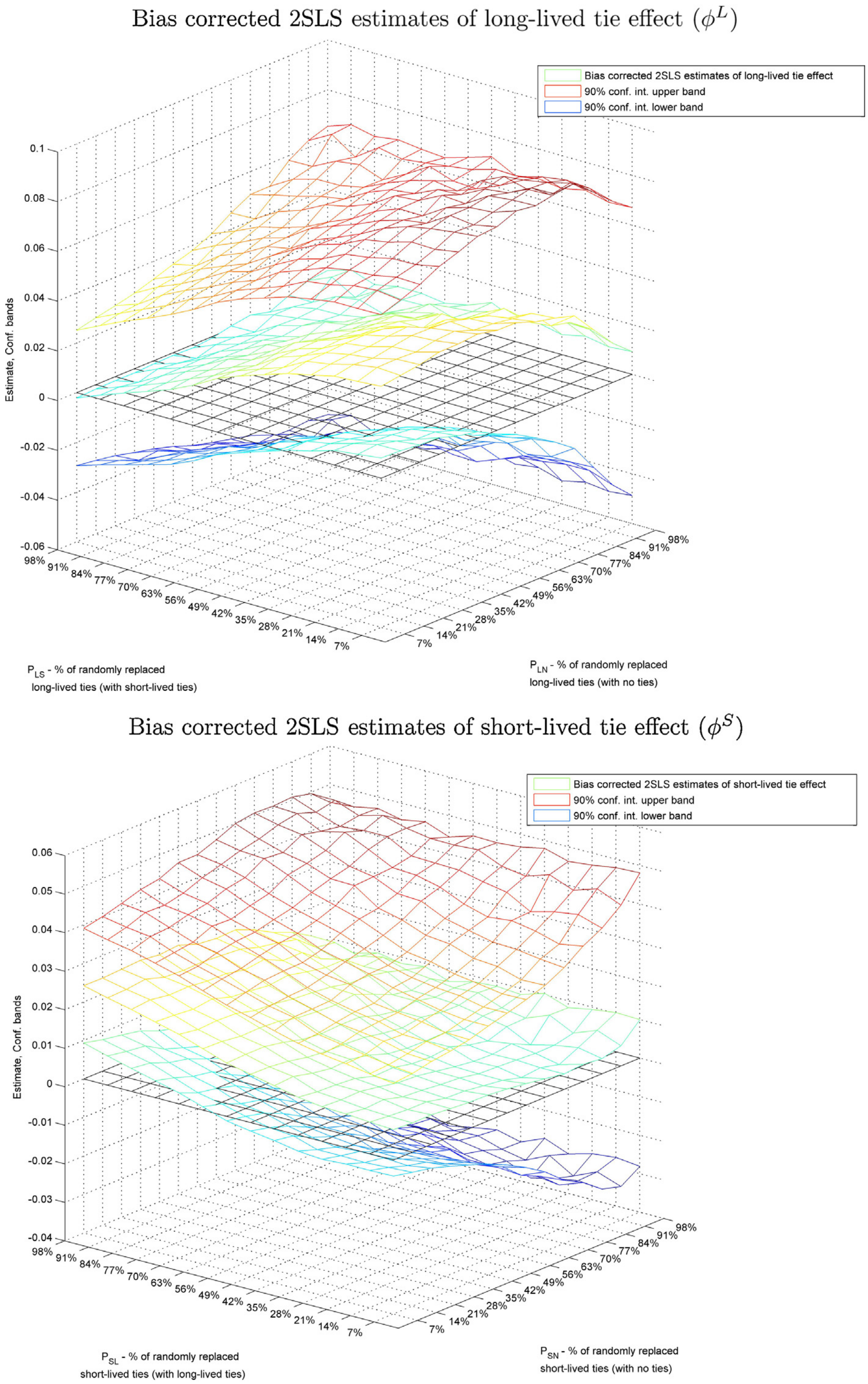


Fig. 1. Misspecification of long-lived ties and short-lived ties: numerical simulation. Notes: For each combination of replacement rates, we plot the average estimate of peer effects. Standard errors are derived assuming drawing independence and accounting for both within and between sample variation.

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.

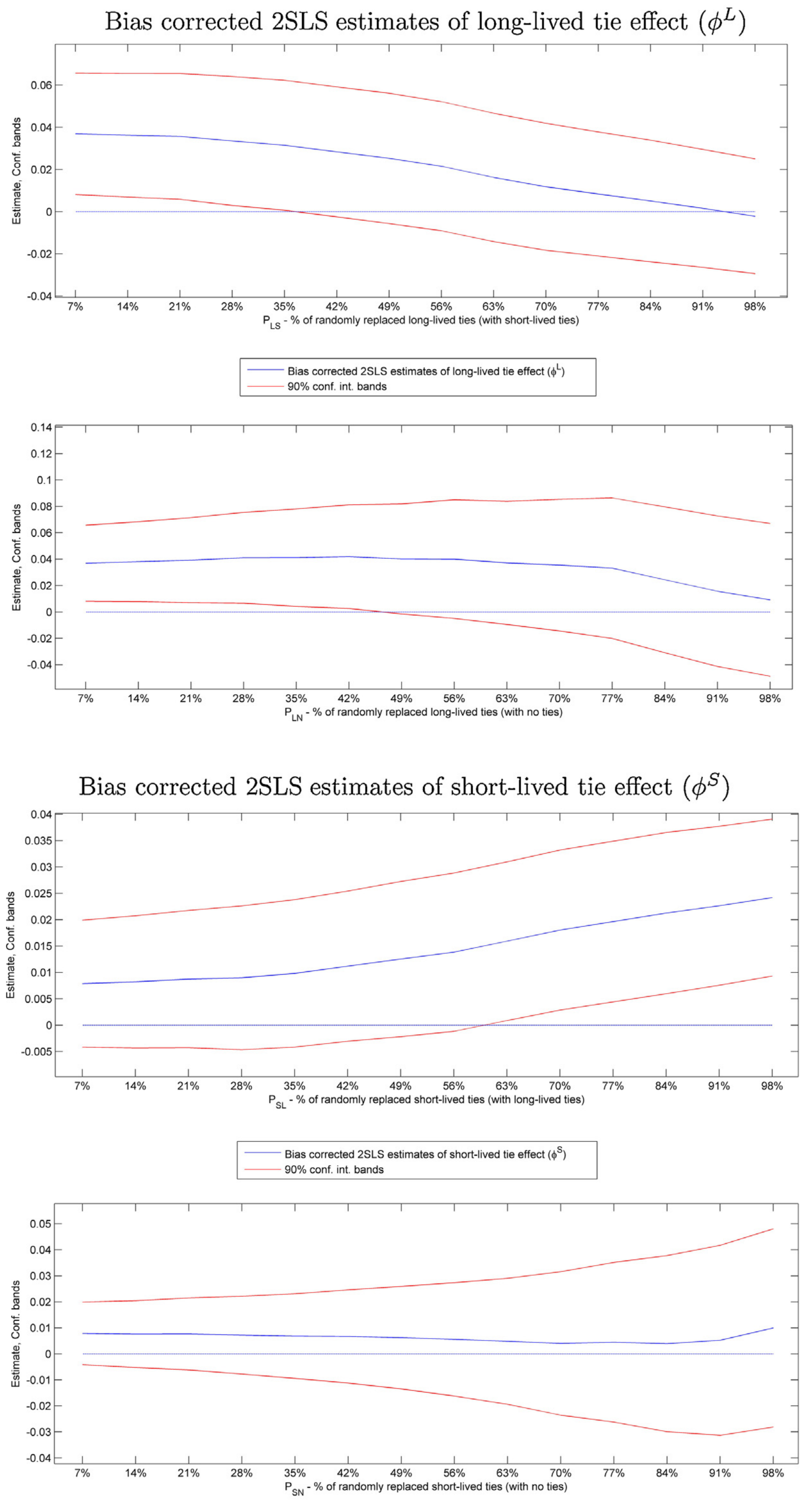


Fig. 2. Simulation experiment. Notes: For each replacement rate, we plot the average estimate of peer effects. Standard errors are derived assuming drawing independence and accounting for both within and between sample variation.

6.2.3. Missing links

In our analysis, the identification assumption (exclusion restriction) is that the characteristics of friends of friends only affect an individual through their effects on his/her direct friends. Nevertheless, it is a well-known fact that real-world networks have high levels of clustering and closure relative to an Erdos–Renyi graph. Therefore, it could be much more likely that links are missing between an individual and a friend of friend (FoF, hereafter). Let W be the true matrix of connections among individuals, and G the observed one. Let π be the probability that a link exists between a FoF:

$$\pi = P\left(w_{ij} = 1 | g_{ij} = 0, \sum_k g_{ik} g_{kj} \neq 0\right).$$

Let θ be the probability that a link exists between students who have no friend in common:

$$\theta = P\left(w_{ij} = 1 | g_{ij} = 0, \sum_k g_{ik} g_{kj} = 0\right).$$

A non-random measurement error of the sort described above would imply that $\pi > \theta$. Let $P_{FoF} = \pi / \left(\pi + \theta\right)$. Higher values of P_{FoF} imply a larger systematic error. As a result, if our estimates are sensitive to this issue, we should observe a marked departure from our point estimates, even for relatively small values of P_{FoF}.

In this section, we check the robustness of our results with respect to departures from the benchmark $P_{FoF} = 0$ by performing a variation of the simulation exercise described in the previous section.

We split non-observed links (N) in two types: (i) links between students with common friends, N_{FoF}, and (ii) links between students without common friends ($N - N_{FoF}$). For each P_{NS} and P_{NL}, we then add a constant number of links with an increasing probability of drawing from N_{FoF}, thus gradually increase P_{FoF} from 0 to 1 at a pace of 0.005. For the sake of simplicity, we hold P_{LS} and P_{SL} equal to zero. We vary the number of added links up to three times the number of existing links. Then, we estimate model (2) replacing the real $\boldsymbol{G}_r^L$ and $\boldsymbol{G}_r^S$ matrices with the simulated ones in turn so that, in total, we estimate model (2) 80,000 times for each type of estimator described in Appendix B.[35]

Table 13 show the results of our simulation experiment for the *2SLS bias-corrected lagged estimator* with 90% confidence bands.[36] Fig. 3 depicts the evidence. The upper panels of Fig. 3 and Table 13 show the estimates of long-lived tie effects, the lower panels show the estimates of short-lived tie effects, for each combination of percentage of additional links and P_{FoF}.

The relevant question here is whether we observe a significant departure from our point estimate (i.e. when $P_{FoF} = P_{NS} = P_{NL} = 0$) when the percentage of friends of friends turned into friends increases. Meaningful comparisons across estimates can be made within each column of Table 13, that is across networks of the same density. The results show that the estimates of the long-lived ties effects remain statistically significant and the point estimates roughly constant as we depart from the observed network topology. The evidence on the short-lived ties effect remains unchanged too: the estimates are never significantly different from zero and are almost unchanged within each column.

7. Short-run versus long-run effects

Thus 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 their educational choices. Using Addhealth data for Wave I only, Calvó-Armengol et al. (2009) study the *current* effect of peers on education, finding that peers do affect the current education activity (i.e. grades) of students. They do not differentiate between different types of peers.

To further investigate this issue, we now contrast the long-run effects and 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 Eq. (1):

$$y_{i,r,t} = \phi^L \sum_{j=1}^{n_r} g_{ij,r,t}^L y_{j,r,t} + \phi^S \sum_{j=1}^{n_r} g_{ij,r,t}^S y_{j,r,t} + x_{i,r,t}\delta + \frac{1}{g_{i,r,t}^L} \sum_{j=1}^{n_r} g_{ij,r,t}^L x'_{j,r,t} \gamma^L + \frac{1}{g_{i,r,t}^S} \sum_{j=1}^{n_r} g_{ij,r,t}^S x'_{j,r,t} \gamma^S + \eta_{r,t} + \epsilon_{i,r,t}, \quad (9)$$

where $y_{i,r,t}$ is now the grades 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 of long-run effects, we exploit variations in link formation in Waves I and II to differentiate between long-lived ties and short-lived ties. We then look at how these different types of peers affect each student's grades obtained in Wave II. The identification

[35] The simulation exercise takes the network component as fixed. Bridges between two unconnected social networks are not allowed.

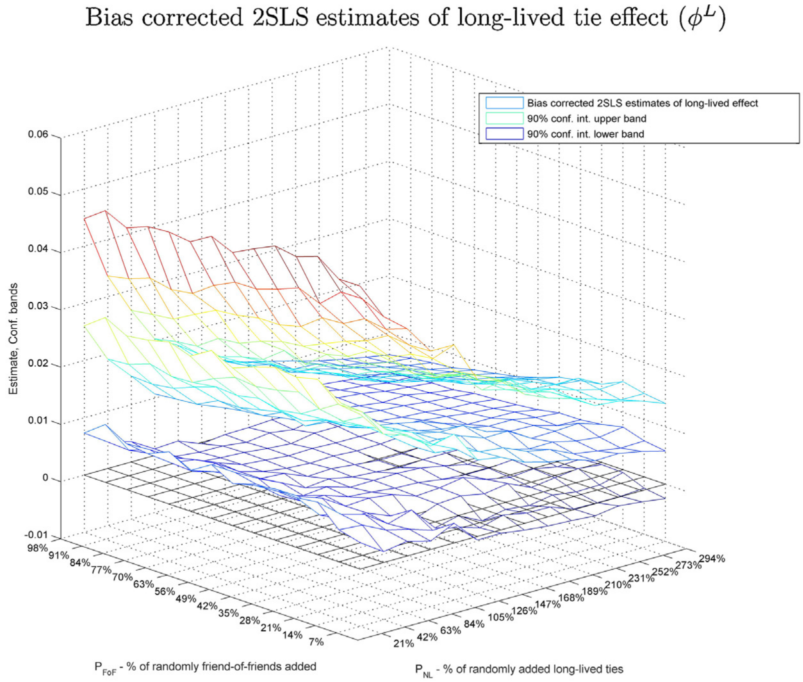
[36] Standard errors have been calculated as in the previous exercise. The simulation results for the other estimators are similar. They are available upon request.

Table 13
Simulated evidence: exclusion restriction violation.

Long-lived ties P_{NL}		7%	21%	35%	49%	63%	77%	91%
of which								
P_{FoF}	7%	[0.005;0.027;0.049]	[0.003;0.021;0.040]	[0.003;0.019;0.035]	[−0.001;0.013;0.026]	[−0.002;0.010;0.021]	[−0.002;0.008;0.017]	[−0.002;0.007;0.016]
	21%	[0.009;0.031;0.052]	[0.005;0.022;0.039]	[0.003;0.017;0.030]	[0.000;0.012;0.025]	[−0.001;0.009;0.020]	[−0.001;0.008;0.018]	[−0.002;0.007;0.015]
	35%	[0.009;0.030;0.051]	[0.005;0.021;0.037]	[0.003;0.016;0.029]	[0.001;0.012;0.023]	[0.002;0.011;0.020]	[0.000;0.008;0.016]	[−0.001;0.006;0.013]
	49%	[0.007;0.028;0.048]	[0.003;0.018;0.033]	[0.002;0.014;0.025]	[0.001;0.011;0.020]	[0.001;0.009;0.017]	[−0.001;0.006;0.013]	[0.000;0.006;0.012]
	63%	[0.007;0.027;0.047]	[0.004;0.018;0.032]	[0.002;0.013;0.023]	[0.001;0.009;0.017]	[0.000;0.007;0.014]	[0.000;0.006;0.012]	[0.000;0.005;0.010]
	77%	[0.008;0.027;0.047]	[0.004;0.017;0.030]	[0.001;0.010;0.020]	[0.000;0.008;0.015]	[0.000;0.006;0.012]	[0.000;0.005;0.010]	[0.000;0.004;0.009]
	91%	[0.010;0.029;0.047]	[0.003;0.015;0.027]	[0.001;0.009;0.018]	[0.000;0.007;0.014]	[0.000;0.005;0.011]	[0.000;0.005;0.009]	[0.000;0.004;0.007]

Short-lived ties P_{NS}		7%	21%	35%	49%	63%	77%	91%
of which								
P_{FoF}	7%	[−0.014;0.015;0.045]	[−0.008;0.015;0.038]	[−0.010;0.015;0.041]	[−0.015;0.016;0.047]	[−0.023;0.017;0.058]	[−0.013;0.017;0.046]	[−0.040;0.018;0.077]
	21%	[−0.011;0.015;0.041]	[−0.009;0.015;0.039]	[−0.017;0.017;0.050]	[−0.015;0.017;0.049]	[−0.017;0.017;0.051]	[−0.016;0.017;0.050]	[−0.018;0.018;0.054]
	35%	[−0.014;0.016;0.045]	[−0.012;0.016;0.044]	[−0.011;0.016;0.044]	[−0.020;0.018;0.055]	[−0.021;0.018;0.057]	[−0.021;0.018;0.058]	[−0.023;0.018;0.059]
	49%	[−0.009;0.015;0.038]	[−0.010;0.016;0.042]	[−0.021;0.018;0.056]	[−0.014;0.018;0.049]	[−0.031;0.019;0.069]	[−0.021;0.019;0.058]	[−0.027;0.019;0.065]
	63%	[−0.012;0.016;0.043]	[−0.015;0.017;0.049]	[−0.021;0.018;0.057]	[−0.021;0.018;0.058]	[−0.028;0.019;0.066]	[−0.058;0.020;0.097]	[−0.050;0.020;0.090]
	77%	[−0.009;0.015;0.039]	[−0.018;0.017;0.052]	[−0.022;0.018;0.058]	[−0.021;0.018;0.058]	[−0.024;0.019;0.062]	[−0.052;0.020;0.092]	[−0.139;0.020;0.179]
	91%	[−0.013;0.016;0.044]	[−0.019;0.018;0.054]	[−0.034;0.019;0.072]	[−0.047;0.019;0.086]	[−0.039;0.019;0.078]	[−0.110;0.020;0.150]	[−0.254;0.020;0.295]

Notes: [10% confidence bound; mean; 90% confidence bound]. Observations = 1,819. Networks = 116. Network fixed effects, contextual effects, parental occupation dummies and individual socio-demographic characteristics included.



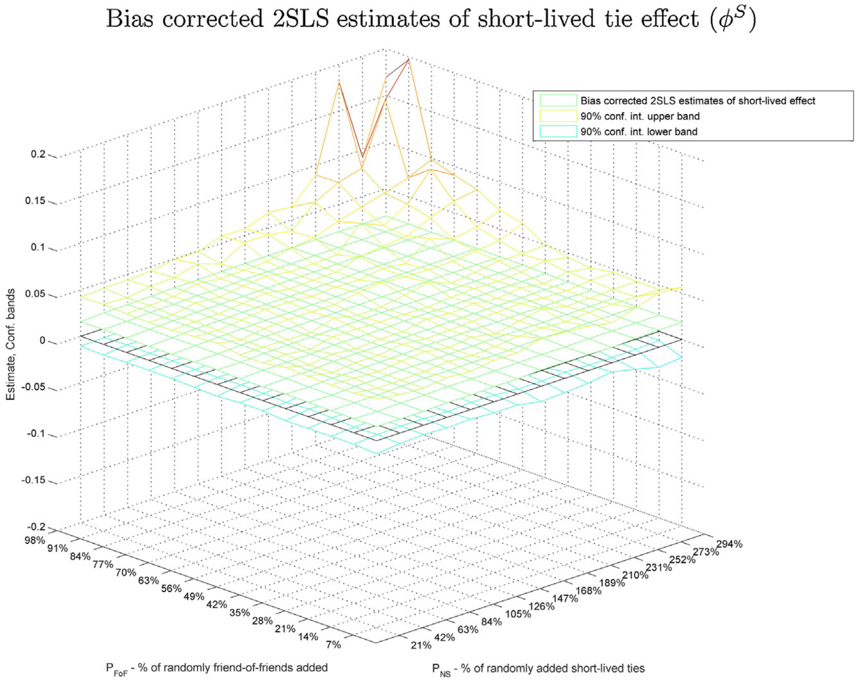


Fig. 3. Exclusion restriction violation: numerical simulation. Notes: For each combination of replacement rates, we plot the average estimate of peer effects. Standard errors are derived assuming drawing independence and accounting for both within and between sample variation.

and estimation strategy remains unchanged with the difference that now, we cannot use IV variables lagged in time (see Appendix B).

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

Table 14
Long and short-lived ties: short-run effects.

Dep. var. GPA		
	2SLS	Bayesian estimation
Long-lived ties (ϕ^L)	0.0238***	0.0253***
	(0.0097)	(0.0093)
Short-lived ties (ϕ^S)	0.0079*	0.0121**
	(0.0046)	(0.0051)
Unobservables ($\sigma_{\epsilon z}$)		0.0053
		(0.1036)
Individual socio-demographic	Yes	Yes
Family background	Yes	Yes
Residential neighborhood	Yes	Yes
Contextual effects	Yes	Yes
Network fixed effects	Yes	Yes
Observations	1,819	932
Networks	116	33

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

Table 15
Heterogeneous endogenous peer effects – comparison between different educational outcomes.

Dep. var.	GPA 2SLS (1)	College 2SLS (2)	Years of education 2SLS (3)
Long-lived ties (ϕ^L)	0.0238***	0.0261**	0.0410***
	(0.0097)	(0.0133)	(0.0139)
Short-lived ties (ϕ^S)	0.0079*	0.0118	0.0060
	(0.0046)	(0.0073)	(0.0056)
GPA		0.1123***	0.2584***
		(0.0141)	(0.0425)
College			2.8462***
			(0.0956)
Individual socio-demographic	Yes	Yes	Yes
Family background	Yes	Yes	Yes
Residential neighborhood	Yes	Yes	Yes
Contextual effects	Yes	Yes	Yes
Network fixed effects	Yes	Yes	Yes
Observations	1,819	1,819	1,819
Networks	116	116	116

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

composite index (labeled as GPA index or grade point average index) is the first principal component.[37] It ranges between 0 and 6.09, with a mean equal to 2.29 and a standard deviation equal to 1.49.

The estimation results of model (9) are contained in Table 14. 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.

Observe that these results are not directly comparable to those obtained in Lavy and Sand (2016). They test the effect of length of acquaintance, comparing the effect of friends from kindergarten to friends who meet in later years, and find that the length of acquaintance does not influence the outcome. While Lavy and Sand (2016) examine the impact of the number of pre-existing friends and their socioeconomic background on GPA, we estimate the impact of (different types of) friends' educational outcomes on own educational outcomes. This is why we view our results as complementary to those obtained in Lavy and Sand (2016).

To better understand the potential mechanisms driving our results, we can also look at the *college enrollment choice*. In Table 15, we show the impact of long-lived and short-lived ties on GPA (column (1), which corresponds to the first column in Table 14), college enrollment (column (2)) and years of education (column (3), which corresponds to the two columns under 2SLS (1) in Table 5). The evidence seems to point towards the fact that long-lived ties rather than short-lived ties matter when college decisions are made. They then remain the peers who matter when explaining education achievement in the long run (conditional on college decisions).

[37] The index explains roughly 56% 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.

As a whole, our results suggest that all peers matter for education (i.e. grades) in the short run, but only long-lived ties matter for future educational choices (i.e. college enrollment and years of schooling). This is consistent with the model developed in Section 5.1 where, in the very short-run, both long-lived and short-lived friends have an impact on current beliefs while, in the long run, only long-lived friends have an impact on educational decisions (since the latter are influenced by the social norm emerging from the iterations of beliefs). This is also consistent with the fact that, in the short-run analysis, the outcome is the students' grades while, in the long-term, the outcome is the number years of education, where social norms matter more.

8. Concluding remarks

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

Using the Moving to Opportunity (MTO) programs, Chetty et al. (2016) compare the long-run outcomes of children who moved to better neighborhoods as toddlers with those who moved in their late teens. They show that the exposure to low-poverty neighborhoods has a positive and significant impact on college education and adult earnings only for children who moved when they were less than 13 years old. That is, what matters for long-run outcomes are the years of exposure to a good neighborhood. If we think of friendship networks as "neighborhoods", then this result is similar to the one obtained in our paper. More generally, our evidence is in line with several studies in sociology and economics (e.g. Coleman, 1988; Wellman and Wortley, 1990; Akerlof and Kranton, 2002), where long-lasting social interactions affect how much students value achievement, their human capital accumulation, and how social norms are formed.

Our evidence implies that, if there are social norms that do not favor education among a group of close friends, then it may be difficult for these students to perform well at school and to attend college.[38] Our finding on the importance of long-lived ties is relevant for educational policy makers. In the array of policies that can be proposed to foster education, our analysis suggests that those attempting to change the social norm of students may be effective. A prominent example is the charter-school policy.[39] In particular, the "No Excuses policy" (Angrist et al., 2010, 2012) is a highly standardized and widely replicated charter model that features a long school day, an extended school year, selective teacher hiring, strict behavior norms, and emphasizes traditional reading and math skills. The main objective is to change the social norms of disadvantaged kids by emphasizing discipline.[40] This is a typical policy that is in accordance with our results since its aim is to change the social norm of students in terms of education. Such policies have very strong positive impact on the educational outcomes of students (see e.g. Angrist et al., 2010, 2012; Fryer, 2014; Epple et al., 2016).

More generally, by highlighting the importance of long-lived peers on educational outcomes, we believe that our work provides new insights into the role of social dynamics in educational attainment.

Appendix A. Data appendix

Table A1 provides a detailed description of the variables used in our study as well as the summary statistics for our sample. Among the individuals selected in our sample, 53% are female and 19% are blacks. The average parental education is high-school graduate. Roughly 10% have parents working in a managerial occupation, another 10% in the office or sales sector, 20% in a professional/technical occupation, and roughly 30% have parents in manual occupations. On average our individuals come from households of about four people. In Wave IV, 42% of our adolescents are now married and nearly half of them (43%) have at least a son or a daughter. The mean intensity in religion practice slightly decreases during the transition from adolescence to adulthood. Roughly, 30% 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.

[38] This is the case in inner cities in the US where, mainly because of urban segregation, African Americans tend to interact mostly with other African Americans and these relationships tend to be persistent over time (see e.g. Sigelman et al., 1996; Tuch et al., 1999; Topa, 2001; Zenou, 2013; Patacchini and Zenou, 2016). As a result, the social norms are quite strong and may not always favor education. Indeed, empirical studies in the US (and also in the UK) have found that African American students in poor areas may be ambivalent about learning standard English and performing well at school because this may be regarded as "acting white" (Fordham and Ogbu, 1986; Wilson, 1987; Fryer and Torelli, 2010; Battu and Zenou, 2010).

[39] A charter school is a public school chartered under the auspices of a state government. While charter laws vary across states, two defining characteristics are: (*i*) charter schools cannot charge tuition fees, and (*ii*) charter schools are not permitted to impose admission requirements and, if oversubscribed, must select from their applicants by lottery. For a recent overview on the charter school policies, see Epple et al. (2016).

[40] Angrist et al. (2012) focus on special needs students that may be underserved. Their results show average achievement gains of 0.36 standard deviations in math and 0.12 standard deviations in reading for each year spent at a charter school.

Table A1
Data description 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
Family background			
Household size	Number of people living in the household	3.40 (1.34)	1–11
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 1–5. 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
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
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 religious services in the past 12 months?”, coded as 0 = never, 1 = a few times, 2 = several times, 3 = once a month, 4 = 2 or 3 times a month, 5 = once a week, 6 = more than once a week.	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 B. IV estimation – technical details

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

$$\boldsymbol{Y} = \boldsymbol{Z}\theta + \boldsymbol{\iota}\cdot\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{10}$$

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

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 $\boldsymbol{J} = D(\boldsymbol{J}_1, \ldots, \boldsymbol{J}_{\bar{r}})$, where $\boldsymbol{J}_r = \boldsymbol{I}_{n_r} - (1/n_r)\boldsymbol{l}'_{n_r}\boldsymbol{l}_{n_r}$. The network fixed effect can be eliminated by a transformation with $\boldsymbol{J}$ such that:

$$\boldsymbol{JY} = \boldsymbol{JZ}\theta + \boldsymbol{J\epsilon}. \tag{11}$$

Let $\boldsymbol{M} = (\boldsymbol{I} - \phi^L\boldsymbol{G}^L - \phi^S\boldsymbol{G}^S)^{-1}$. The equilibrium outcome vector $\boldsymbol{Y}$ in (10) is then given by the reduced form equation:

$$\boldsymbol{Y} = \boldsymbol{M}(\boldsymbol{X}^*\beta + \boldsymbol{\iota}\cdot\boldsymbol{\eta}) + \boldsymbol{M\epsilon}. \tag{12}$$

It follows that $\boldsymbol{G}^L\boldsymbol{Y} = \boldsymbol{G}^L\boldsymbol{MX}^*\beta + \boldsymbol{G}^L\boldsymbol{M}\iota\eta + \boldsymbol{G}^L\boldsymbol{M}\epsilon$ and $\boldsymbol{G}^S\boldsymbol{Y} = \boldsymbol{G}^S\boldsymbol{MX}^*\beta + \boldsymbol{G}^S\boldsymbol{M}\iota\eta + \boldsymbol{G}^S\boldsymbol{M}\epsilon$. $\boldsymbol{G}^L\boldsymbol{Y}$ and $\boldsymbol{G}^S\boldsymbol{Y}$ are correlated with $\boldsymbol{\epsilon}$ because $\mathrm{E}[(\boldsymbol{G}^L\boldsymbol{M\epsilon})'\boldsymbol{\epsilon}] = \sigma^2 tr(\boldsymbol{G}^L\boldsymbol{M}) \neq 0$ and $\mathrm{E}[(\boldsymbol{G}^S\boldsymbol{M\epsilon})'\boldsymbol{\epsilon}] = \sigma^2 tr(\boldsymbol{G}^S\boldsymbol{M}) \neq 0$. Hence, in general, (11) cannot be consistently estimated by OLS.[41] If $\boldsymbol{G}$ is row-normalized such that $\boldsymbol{G}\cdot\boldsymbol{l}_n = \boldsymbol{l}_n$, where $\boldsymbol{l}_n$ 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 $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$. They first consider the 2SLS estimator based on the conventional instrumental matrix for the estimation of (11): $\boldsymbol{Q}_1 = \boldsymbol{J}(\boldsymbol{GX}^*, \boldsymbol{X}^*)$ (*finite-IVs 2SLS*). Then, they propose to use additional instruments (IVs) $J\,G\iota$ and enlarge the instrumental matrix: $Q_2 = (Q_1, J\,G\iota)$ (many-IVs 2SLS). The additional IVs of $J\,G\iota$ 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 $\boldsymbol{Q}_1$. 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 number of IVs. The 2SLS could be asymptotic biased when the number of IVs increases too fast relative to the sample size (see, e.g. Bekker, 1994; Bekker and van der Ploeg, 2005; Hansen et al., 2008). As detailed in Section 2, in this empirical study, we have a number of small networks. Liu and Lee (2010) also propose a bias-correction procedure based on the estimated leading-order many-IV bias (*bias-corrected 2SLS*). The bias-corrected many-IV 2SLS estimator is properly centered, asymptotically normally distributed, and efficient when the average group size is sufficiently large. Thus, it is the more appropriate estimator in our case study.[42]

Let us now derive the best 2SLS estimator for Eq. (11). From the reduced form Eq. (10), we have $E(\boldsymbol{Z}) = [\boldsymbol{G}^L\boldsymbol{M}(\boldsymbol{X}^*\beta + \boldsymbol{\iota}\cdot\boldsymbol{\eta}), \boldsymbol{G}^S\boldsymbol{M}(\boldsymbol{X}^*\beta + \boldsymbol{\iota}\cdot\boldsymbol{\eta}), \boldsymbol{X}^*]$. The best IV matrix for $\boldsymbol{JZ}$ is given by

$$\boldsymbol{J}f = \boldsymbol{JE}(\boldsymbol{Z}) = J[\boldsymbol{G}^L\boldsymbol{M}(\boldsymbol{X}^*\beta + \boldsymbol{\iota}\cdot\boldsymbol{\eta}), \boldsymbol{G}^S\boldsymbol{M}(\boldsymbol{X}^*\beta + \boldsymbol{\iota}\cdot\boldsymbol{\eta}), \boldsymbol{X}^*], \tag{13}$$

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

$$(\boldsymbol{G}^L\boldsymbol{X}^*, \boldsymbol{G}^S\boldsymbol{X}^*, \boldsymbol{G}^S\boldsymbol{G}^L\boldsymbol{X}^*, (\boldsymbol{G}^L)^2\boldsymbol{X}^*, (\boldsymbol{G}^S)^2\boldsymbol{X}^*, (\boldsymbol{G}^S)^2\boldsymbol{G}^L\boldsymbol{X}^*, (\boldsymbol{G}^S)^2(\boldsymbol{G}^L)^2\boldsymbol{X}^*, \ldots),$$

and

$$(\boldsymbol{G}^L\boldsymbol{\iota}, \boldsymbol{G}^S\boldsymbol{\iota}, \boldsymbol{G}^S\boldsymbol{G}^L\boldsymbol{\iota}, (\boldsymbol{G}^L)^2\boldsymbol{\iota}, (\boldsymbol{G}^S)^2\boldsymbol{\iota}, (\boldsymbol{G}^S)^2\boldsymbol{G}^L\boldsymbol{\iota}, (\boldsymbol{G}^S)^2(\boldsymbol{G}^L)^2\boldsymbol{\iota}, \ldots),$$

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

$$\boldsymbol{Q}_\infty = \boldsymbol{J}(\boldsymbol{G}^L(\boldsymbol{G}^L\boldsymbol{X}^*, \boldsymbol{G}^S\boldsymbol{X}^*, \boldsymbol{G}^S\boldsymbol{G}^L\boldsymbol{X}^*, \ldots, \boldsymbol{G}^L\boldsymbol{\iota}, \boldsymbol{G}^S\boldsymbol{\iota}, \boldsymbol{G}^S\boldsymbol{G}^L\boldsymbol{\iota}, \ldots), \boldsymbol{G}^S(\boldsymbol{G}^L\boldsymbol{X}^*, \boldsymbol{G}^S\boldsymbol{X}^*, \boldsymbol{G}^S\boldsymbol{G}^L\boldsymbol{X}^*, \ldots, \boldsymbol{G}^L\boldsymbol{\iota}, \boldsymbol{G}^S\boldsymbol{\iota}, \boldsymbol{G}^S\boldsymbol{G}^L\boldsymbol{\iota}, \ldots), \boldsymbol{X}^*).$$

Let $\boldsymbol{Q_K}$ be an $n\times K$ submatrix of $\boldsymbol{Q_\infty}$ (with $K \geq 3m+2$) including X^*. Let Q_L be an $n\times K_L$ submatrix of $Q_{L\infty} = G^L(G^LX^*, G^SX^*, G^SG^LX^*, \ldots, G^L\iota, G^S\iota, G^SG^L\iota, \ldots)$ and Q_S an $n\times K_S$ submatrix of $Q_{S\infty} = G^S(G^LX^*, G^SX^*, G^SG^LX^*, \ldots, G^L\iota, G^S\iota, G^SG^L\iota, \ldots)$. We assume that $(K_L/K_S) = 1$. Let $\boldsymbol{P_K} = \boldsymbol{Q_K}(\boldsymbol{Q}'_{\boldsymbol{K}}\boldsymbol{Q_K})^{-1}\boldsymbol{Q}'_{\boldsymbol{K}}$ be the projector of $\boldsymbol{Q_K}$. The resulting 2SLS estimator is given by:

$$\hat{\theta}_{2sls} = \left(\boldsymbol{Z}'\boldsymbol{P_K}\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{P_K}\boldsymbol{y}. \tag{14}$$

---

[41] 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.

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

B.1. 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 ϵ are i.i.d with zero mean, variance σ^2 and a moment of order higher than four exists.
Assumption C2: The elements of $\boldsymbol{X}^*$ are uniformly bounded constants, $\boldsymbol{X}^*$ has the full rank k and $\lim_{n\to\infty}\boldsymbol{X}^{*\prime}\boldsymbol{X}^*$ exists and is nonsingular.
Assumption C3: The sequences of matrices $\{\boldsymbol{G}^L\}$, $\{\boldsymbol{G}^S\}$, $\{\boldsymbol{M}\}$ are uniformly bounded.
Assumption C4: $\bar{\boldsymbol{H}} = \lim_{n\to\infty}(1/n)f'f$ is a finite non singular matrix.
Assumption C5: There exists a $\boldsymbol{K}\times(3m+2)$ matrix π_K such that $\|f-\boldsymbol{Q_K}\pi_K\|_\infty \to 0$ 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 $\boldsymbol{\Psi}_{K,S}=\boldsymbol{P_K}\boldsymbol{G}^L\boldsymbol{M}$ and $\boldsymbol{\Psi}_{K,W}=\boldsymbol{P_K}\boldsymbol{G}^S\boldsymbol{M}$. The following proposition shows consistency and asymptotic normality of the 2SLS estimator (14).

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

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

$$\frac{1}{n}\boldsymbol{Z}'\boldsymbol{P_K}\boldsymbol{Z}=\boldsymbol{H}-e_f+\frac{1}{n}v'\boldsymbol{P_K}f+\frac{1}{n}f'\boldsymbol{P_K}v+\frac{1}{n}v'\boldsymbol{P_K}v=\boldsymbol{H}+\boldsymbol{O}(\boldsymbol{tr}(e_f))+\boldsymbol{O}_p(\sqrt{K/n})+\boldsymbol{O}_p(K/n)=\bar{\boldsymbol{H}}+o_p(1),$$

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

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

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

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

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

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 $\sigma^2\bar{\boldsymbol{H}}^{-1}$ 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 $Q_{\boldsymbol{K}}$, 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

$$\hat{\theta}_{c2sls}=\left(\boldsymbol{Z}'\boldsymbol{P}_{\boldsymbol{Q_K}}\boldsymbol{Z}\right)^{-1}\left[\boldsymbol{Z}'\boldsymbol{P}_{\boldsymbol{Q_K}}\boldsymbol{y}-\tilde{\sigma}^2\left[tr\left(\tilde{\boldsymbol{\Psi}}_{K,L}\right), tr\left(\tilde{\boldsymbol{\Psi}}_{K,S}\right), \boldsymbol{0}_{3m\times 1}\right]'\right], \tag{15}$$

where $\tilde{\boldsymbol{\Psi}}_{K,L}=\boldsymbol{P_K}\boldsymbol{G}^L\boldsymbol{M}(\tilde{\phi}^L,\tilde{\phi}^S)$ and $\tilde{\boldsymbol{\Psi}}_{K,S}=\boldsymbol{P_K}\boldsymbol{G}^S\boldsymbol{M}(\tilde{\phi}^L,\tilde{\phi}^S)$ are estimated with initial $\sqrt{n}$-consistent estimators of $\tilde{\sigma}$, $\tilde{\phi}^L$ and $\tilde{\phi}^S$. This estimator adjusts the 2SLS estimator by the estimated leading order bias $b_{2s\,ls}$, which is presented in Proposition 1.

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

Proof. We need to show that

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

where $\tilde{\boldsymbol{M}} = \boldsymbol{M}(\tilde{\phi}^L, \tilde{\phi}^S)$. Given Proposition 1, this is quite straightforward since

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

because $\tilde{\boldsymbol{M}}-\boldsymbol{M} = \tilde{\boldsymbol{M}}[(\tilde{\phi}^L-\phi^L)\boldsymbol{G}^L\boldsymbol{M} + (\tilde{\phi}^S-\phi^S)\boldsymbol{G}^S\boldsymbol{M}]$, 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*: $\hat{\boldsymbol{\theta}}_{2sls1} = (\boldsymbol{Z}'\boldsymbol{P}_1\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{P}_1\boldsymbol{y}$, where $\boldsymbol{P}_1 = \boldsymbol{Q}_1(\boldsymbol{Q}_1'\boldsymbol{Q}_1)_1^{-1}\boldsymbol{Q}_1'$ and $\boldsymbol{Q}_1$ contains the linearly independent columns of $\boldsymbol{J}[\boldsymbol{X}^*, \boldsymbol{GX}^*, \boldsymbol{GGX}^*]$.

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

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

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

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

(vi) *Bias-corrected lagged*: $\hat{\boldsymbol{\theta}}_{c2slsL} = (\boldsymbol{Z}'\boldsymbol{P}_4\boldsymbol{Z})^{-1}\{\boldsymbol{Z}'\boldsymbol{P}_4\boldsymbol{y} - \tilde{\sigma}^2_{2sls1}[tr\left(\boldsymbol{P_4 G}^L\tilde{\boldsymbol{M}}\right), tr\left(\boldsymbol{P_4 G}^S\tilde{\boldsymbol{M}}\right), \boldsymbol{0}_{3m\times 1}]'\}$ where $\tilde{\boldsymbol{M}} = (\boldsymbol{I} - \tilde{\phi}^L_{2sls1L}\boldsymbol{G}^L - \tilde{\phi}^S_{2sls1L}\boldsymbol{G}^S)^{-1}$, and $\tilde{\sigma}^2_{2sls1L}$, $\tilde{\phi}^L_{2sls1L}$ and $\tilde{\phi}^S_{2sls-1L}$ are $\sqrt{n}$-consistent initial estimators of σ^2, ϕ^L and ϕ^S obtained by *Finite-IV lagged*.

Appendix C. The DeGroot model with heterogeneous peers

C.1. The DeGroot model

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

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

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

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

DeGroot (1974) assumes that agents update their beliefs by repeatedly taking *weighted averages* of their neighbors' beliefs with p_{ij} being the weight that agent i places on the current belief of agent j in forming his or her belief for the next period. To be more precise, in the beginning, each agent $i \in N$ assigns a (relative) *weight* p_{ij} on each of his/her neighbors j such that $\sum_{j=1}^{j=n} p_{ij} = 1$. Observe that p_{ij} is the direct influence of agent j on i so that $p_{ij} = 0$ for $j \notin D(i)$.

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

$$p_{ij} = \frac{g_{ij}\pi_j}{\sum_{k=1}^{k=n} g_{ik}\pi_k}. \tag{16}$$

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

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

$$b_i^{(t)} = \sum_{j=1}^{j=n} p_{ij} b_j^{(t-1)}, \tag{17}$$

or, in matrix form

$$\boldsymbol{b}^{(t)} = \boldsymbol{P}\boldsymbol{b}^{(t-1)}. \tag{18}$$

The *limiting beliefs* can be calculated as a function of the initial beliefs and weights. They are given by:

$$\boldsymbol{b}^{\infty} = \lim_{t\to\infty} \boldsymbol{P}^t \boldsymbol{b}^{(0)}, \tag{19}$$

where $\boldsymbol{P}^t$ is the matrix of cumulative influences in period t.

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

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

Definition 3. Agents in network $\boldsymbol{G}$ reach a consensus if for any $\boldsymbol{b}(0) \in \mathbb{R}^n$, we have:

$$\lim_{t\to+\infty} \left[b_i(t) - b_j(t)\right] = 0 \quad \text{for all} \quad (i, j) \in N^2.$$

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

Proposition 4. *[DeGroot (1974)] Assume that network $\boldsymbol{G}$ is strongly connected and agents follow the average-based updating process described by expression (18). Then all agents in $\boldsymbol{G}$ reach a consensus, which, for each agent $i \in N$, is given in the limit by:*

$$\left(\lim_{t\to+\infty} \boldsymbol{P}^t \boldsymbol{b}\right)_i = \boldsymbol{e}\boldsymbol{b} \quad \text{for every} \quad \boldsymbol{b} \in [0, 1]^n,$$

where $\boldsymbol{e}$ is the left-hand unit eigenvector of matrix $\boldsymbol{P}$.

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

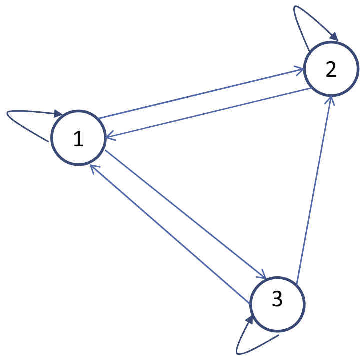


Figure C1. Directed network.

Corollary 2. *Assume that the network* $\boldsymbol{G}$ *is undirected so that* $g_{ij} = 1$ *means that i and j pay attention to each other and define* $p_{ij} = g_{ij}/d_i(\boldsymbol{G})$*, where* $d_i(\boldsymbol{G}) = \sum_{j \in N} g_{ij}$ *is the degree of i. Then, under the same assumptions as in Proposition 4, the same convergence result holds where* e_i *is now given by:*

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

This is a particular case of Proposition 4 where we assume that each agent equally split attention among his or her neighbors in an undirected network. In that case, Corollary 2 shows that social influence is proportional to the agent's degree. An interesting feature of Corollary 2 is that social influence is only determined by the *degree distribution* of the network and not by any structural properties of the network such as the average path length or the centrality of the network. However, Golub and Jackson (2010) shows that the *speed of convergence* is affected by a structural property of the network, namely *homophily*. They show that homophily does not alter agents' social influence and therefore has no effect on the long-term learning but significantly reduces the speed of convergence.

An example of convergence with the DeGroot model. To illustrate the notations and the results in Proposition 4, let us consider the following network (Fig. C1): where the adjacency matrix $\boldsymbol{G}$ and the row normalized adjacency matrix $\tilde{\boldsymbol{G}}$ are given by (outdegrees):

$$\boldsymbol{G} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \tilde{\boldsymbol{G}} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}. \tag{20}$$

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

$$\boldsymbol{P} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

Assuming that the initial beliefs are given by: $\boldsymbol{b}^{(0)} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^{\mathrm{T}}$, it is easily verified that:

$$\boldsymbol{b}^{(1)} = \boldsymbol{P}\boldsymbol{b}^{(0)} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/2 \\ 1/2 \end{pmatrix},$$

$$\boldsymbol{b}^{(2)} = \boldsymbol{P}\boldsymbol{b}^{(1)} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1/3 \\ 1/2 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 0.444 \\ 0.417 \\ 0.417 \end{pmatrix}.$$

By continuing iterating, there is convergence to the following consensus:[43]

$$\boldsymbol{b}^{\infty} = \lim_{t \to \infty} \boldsymbol{P}^t \boldsymbol{b}^{(0)} = \begin{pmatrix} 3/7 \\ 2/7 \\ 2/7 \end{pmatrix}. \tag{21}$$

[43] As shown in Proposition 4, to calculate the limiting beliefs, one needs to solve the following equation:

This means that

$$\boldsymbol{P}^t = \begin{pmatrix} 3/7 & 2/7 & 2/7 \\ 3/7 & 2/7 & 2/7 \\ 3/7 & 2/7 & 2/7 \end{pmatrix},$$

which implies that a consensus is reached. In other words, no matter what are the initial beliefs $\boldsymbol{b}^{(0)}$, all agents end up with limiting beliefs corresponding to the entries of $\boldsymbol{b}^{\infty} = \lim_{t\to\infty} \boldsymbol{P}^t \boldsymbol{b}^{(0)}$ where

$$b_1^{\infty} = b_2^{\infty} = b_3^{\infty} = \frac{3}{7} b_1^{(0)} + \frac{2}{7} b_2^{(0)} + \frac{2}{7} b_3^{(0)}. \tag{22}$$

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

$$b_1^{\infty} = b_2^{\infty} = b_3^{\infty} = \frac{3}{7} = 0.429. \tag{23}$$

If there are two states of the world and if the consensus is on the state {It is worth continuing to study}, then this means that all students in this network agree that with 42.9% that it is worth continuing to study.

C.2. The DeGroot model with short-lived and long-lived ties

Let us consider the model of Section 5.1 and consider the network described in Fig. C1. The adjacency matrix $\boldsymbol{G}$ and the row-normalized one $\tilde{\boldsymbol{G}}$ are given by (20). Assume that both agents 1 and 2 and agents 1 and 3 have a *long-lived* friendship while agents 2 and 3 have a *short-lived* friendship. As stated above, we assume that each agent has a long-lived relationship with him/herself, i.e. $g_{ii}^L = 1$ and $g_{ii}^S = 0$. We have:

$$\boldsymbol{G}^L = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{G}^S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{so that} \quad \boldsymbol{G}^L + \boldsymbol{G}^S = \boldsymbol{G}.$$

Let us row-normalize these matrices so that

$$\tilde{\boldsymbol{G}}^L = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \quad \text{and} \quad \tilde{\boldsymbol{G}}^S = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

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

Let us first determine the initial beliefs. We have:

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

Now we can determine the consensus among all the students where the updates is only on the matrix $\tilde{\boldsymbol{G}}^L$ for the long-lived students. It is easily shown that (since $\boldsymbol{P} = \tilde{\boldsymbol{G}}^L$):

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

so that there is convergence to the following consensus:

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

$$\boldsymbol{b}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{b}^{\mathrm{T}},$$

since

$$\left(\lim_{t\to+\infty} \boldsymbol{P}^t \boldsymbol{b}\right)_i = \boldsymbol{eb} \quad \text{for every} \quad \boldsymbol{b} \in [0, 1]^n,$$

where $\boldsymbol{e}$ is the left-hand unit eigenvector of $\boldsymbol{P}$.

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

$$(b_1)^{\infty} = (b_2)^{\infty} = \left(b_3^L\right)^{\infty} = \frac{3}{7}\tilde{b}_1^{(0)} + \frac{2}{7}\tilde{b}_2^{(0)} + \frac{2}{7}\tilde{b}_3^{(0)}.$$

In the text, to calculate the initial beliefs, we assume that

$$\tilde{\boldsymbol{b}}^{(0)} = \tilde{\boldsymbol{G}}\boldsymbol{b}^{(0)}.$$

Since $\boldsymbol{b}^{(0)} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^{\mathrm{T}}$, we have:

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

Therefore, we can then calculate the *consensus* reached by all agents in the network. We have:

$$\frac{3}{7}\tilde{b}_1^{(0)} + \frac{2}{7}\tilde{b}_2^{(0)} + \frac{2}{7}\tilde{b}_3^{(0)} = \frac{3}{7}\frac{1}{3} + \frac{2}{7}\frac{1}{2} + \frac{2}{7}\frac{1}{3} = \frac{8}{21} = 0.38.$$

If the consensus is on the state {It is worth continuing studying}, this means that the three students reach a consensus for which there will agree that it is worth continuing studying with probability 0.38. This example shows that the beliefs converge over time for all the students and that they reach a consensus but it also shows that agent 1 has more influence than agents 2 and 3 over the limiting beliefs. Observe that, compared to the example above (see (23)) where we did not differentiate between short-lived and long-lived friends, the consensus of continuing studying here leads to a lower probability since $0.38 < 0.429$, even though we update on the same matrix $\tilde{\boldsymbol{G}}^{L} = \boldsymbol{P}$. This is because of the influence of the short-lived friends who change the initial beliefs from $\boldsymbol{b}^{(0)}$ to $\tilde{\boldsymbol{b}}^{(0)} = \tilde{\boldsymbol{G}}\boldsymbol{b}^{(0)}$.

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

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

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

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

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

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

where B^{*} 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:

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

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 $m = \sum_i (d_i/n)$ the average number of links among nodes. Assortativity is defined as:

$$A(\boldsymbol{G}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(d_i - m)(d_j - m)g_{ij}}{\sum_{i=1}^{n}(d_i - m)^2}.$$

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

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

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

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

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

Appendix E. Bayesian estimation

Prior and posteriors distributions. In order to draw random values from the marginal posterior distributions of parameters we need to set prior distributions of those parameters. Once priors and likelihoods are specified, we can derive marginal posterior distributions of parameters and draw values from them. Given the link formation, the probability of observing the short- and long-lasting networks, $\boldsymbol{G}_r^L$ and $\boldsymbol{G}_r^S$ is

$$\begin{aligned} P(\boldsymbol{G}_r^L | x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}) &= \prod_{i \neq j} P(g_{ij,r,t-1}^L | x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}), \\ P(\boldsymbol{G}_r^S | x_{i,r}, x_{j,r}, z_{i,r}, z_{j,r}) &= \prod_{i \neq j} P(g_{ij,r,t}^S | x_{ij,r}, z_{i,r}, z_{j,r}). \end{aligned}$$

Following Hsieh and Lee (2015) our prior distributions are

$$\begin{aligned} z_{i,r} &\sim N(0, 1), \\ \omega &\sim N_{2K+3}(\omega_0, \Omega_0), \\ \phi^L &\sim U[-\kappa_L, \kappa_L], \\ \phi^S &\sim U[-\kappa_S, \kappa_S], \\ \beta^* &\sim N_{3K+1}(\beta_0, B_0), \\ (\sigma_\varepsilon^2, \sigma_{\varepsilon z}) &\sim TN_2(\sigma_0, \Sigma_0), \\ \eta_r | \sigma_\eta &\sim N(0, \sigma_\eta), \\ \sigma_\eta &\sim IG\left(\frac{\varsigma_0}{2}, \frac{\zeta_0}{2}\right), \end{aligned}$$

where $\omega = (\delta^L, \theta^L, \delta^S, \theta^S)$, $\kappa_L = (1/\kappa) - |\phi^L|$, $\kappa_S = (1/\kappa) - |\phi^S|$ and $\kappa = 1/\max(\min(\max_i(\sum_j g_{ij}^S), \max_j(\sum_i g_{ij}^S)), \min(\max_i(\sum_j g_{ij}^L), \max_j(\sum_i g_{ij}^L)))$ from Gershgorin Theorem, $U[\cdot]$, $TN_2(\cdot)$ and $IG(\cdot)$ are respectively the uniform, bivariate truncated normal, and inverse gamma distributions. Those distributions depend on hyper-parameters (like β_0) that are set by the econometrician. It follows that the marginal posteriors are

$$P(\boldsymbol{Z}_r|\boldsymbol{Y}_r, \boldsymbol{G}_r^S, \boldsymbol{G}_r^L, \rho) \propto \prod_{r=1}^{\bar{r}}\prod_{i}^{n_r}\phi(z_{i,r})P(\boldsymbol{Y}_r, \boldsymbol{G}_r^S, \boldsymbol{G}_r^L|\boldsymbol{Z}_r, \rho),$$

$$P(\omega|\boldsymbol{Y}_r, \boldsymbol{G}_r^S, \boldsymbol{G}_r^L) \propto \boldsymbol{\phi}^{2K+3}(\omega, \omega_0, \Omega_0)\prod_{r=1}^{\bar{r}}P(\boldsymbol{G}_r^S, \boldsymbol{G}_r^L|\boldsymbol{Z}_r, \omega),$$

$$P(\phi^S, \phi^S|\boldsymbol{Y}_r, \boldsymbol{G}_r^S, \boldsymbol{G}_r^L, \boldsymbol{Z}_r, \beta, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) \propto \prod_{r=1}^{\bar{r}}P(\boldsymbol{Y}_r|\boldsymbol{G}_r^S, \boldsymbol{G}_r^L, \boldsymbol{Z}_r, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}),$$

$$P(\beta^*|\boldsymbol{Y}_r, \boldsymbol{G}_r^S, \boldsymbol{G}_r^L, \boldsymbol{Z}_r, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \phi^S, \phi^L) \propto \boldsymbol{\phi}^{3K+2}(\tilde{\beta}, \tilde{\boldsymbol{B}}),$$

$$P(\sigma_\varepsilon^2, \sigma_{\varepsilon z}|\boldsymbol{Y}_r, \boldsymbol{G}_r^S, \boldsymbol{G}_r^L, \boldsymbol{Z}_r, \phi^S, \phi^L) \propto \boldsymbol{\phi}_T^2((\sigma_\varepsilon^2, \sigma_{\varepsilon z}), \sigma_0, \Sigma_0)\prod_{r=1}^{\bar{r}}P(\boldsymbol{Y}_r|\boldsymbol{G}_r^L, \boldsymbol{G}_r^S, \boldsymbol{Z}_r, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta),$$

$$P(\eta_r|\boldsymbol{Y}_r, \boldsymbol{G}_r^L, \boldsymbol{G}_r^S, \boldsymbol{Z}_r, \phi^S, \phi^L, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta) \propto \phi(\eta_r, \tilde{\eta}_r, \tilde{M}_r),$$

$$P(\sigma_\eta|\boldsymbol{Y}_r, \boldsymbol{G}_r^L, \boldsymbol{G}_r^S, \boldsymbol{Z}_r, \phi^S, \phi^L, \sigma_\varepsilon^2, \sigma_{\varepsilon z}) \propto \iota\gamma\left(\frac{\varsigma_0+\bar{r}}{2}, \frac{\zeta_0+\sum_{r=1}^{\bar{r}}\eta_r^2}{2}\right),$$

where $\rho = (\omega, \phi^S, \phi^L, \beta^*, \sigma_\varepsilon^2, \sigma_{\varepsilon z}, \sigma_\eta, \eta)$, $\boldsymbol{\phi}^l(\cdot)$ is the multivariate l-dimensional normal density function, $\boldsymbol{\phi}_T^l(\cdot)$ is the truncated counterpart, $\iota\gamma(\cdot)$ is the inverse gamma density function. $\tilde{\beta} = \tilde{B}(B_0^{-1}\beta_0 + \sum_{r=1}^{\bar{r}}\boldsymbol{X}_r'\boldsymbol{V}_r(\boldsymbol{S}_r\boldsymbol{Y}_r - \sigma_{\varepsilon z}\boldsymbol{Z}_r))$, $\tilde{B} = (B_0^{-1} + \sum_{r=1}^{\bar{r}}\boldsymbol{X}_r'\boldsymbol{V}_r\boldsymbol{X}_r)^{-1}$, $\tilde{\eta}_r = (\sigma_\varepsilon^2 - \sigma_{\varepsilon z}^2)^{-1}\tilde{M}_r\boldsymbol{l}_{n_r}'(\boldsymbol{S}_r\boldsymbol{Y}_r - \sigma_{\varepsilon z}\boldsymbol{Z}_r - \boldsymbol{X}_r^*\beta^*)$, and $\tilde{M}_r = (\sigma_\eta^{-2} + (\sigma_\varepsilon^2 - \sigma_{\varepsilon z}^2)^{-1}\boldsymbol{l}_{n_r}'\boldsymbol{l}_{n_r})^{-1}$, where $\boldsymbol{V}_r = (\sigma_\varepsilon^2 - \sigma_{\varepsilon z}^2)I_{n_r} + \sigma_\eta^2\boldsymbol{l}_{n_r}\boldsymbol{l}_{n_r}'$, where $\boldsymbol{X}_r^* = (\boldsymbol{X}_r, \boldsymbol{G}_r^{*S}\boldsymbol{X}_r, \boldsymbol{G}_r^{*L}\boldsymbol{X}_r)$. The posteriors of β^*, $\{\eta_r\}$ and σ_η are available in closed forms and a usual Gibbs Sampler is used to draw from them. The other parameters are drawn using the Metropolis–Hastings (M–H) algorithm (Metropolis-within-Gibbs).

Sampling algorithm. We start our algorithm by picking $(\omega^{(1)}, \phi^{L(1)}, \phi^{S(1)}, \beta^{*(1)}, \sigma_\varepsilon^{2(1)}, \sigma_{\varepsilon z}^{(1)}, \sigma_\eta^{(1)}, \eta^{(1)})$ as starting values. For $\beta^{*(1)}, \eta^{(1)}, \phi^{L(1)}, \phi^{S(1)}$ we use OLS estimates, while we set the variances-covariances $\sigma_\varepsilon^{2(1)}, \sigma_{\varepsilon z}^{(1)}, \sigma_\eta^{(1)}$ at 0.[44] We ought to draw samples of $z_{i,r}^t$ from $P(z_{i,r}|Y_r, G_r^L, G_r^S, \rho)$, $\quad i = 1, \ldots, n$. To do this, we first draw a candidate $\tilde{z}_{i,r}^t$ from a normal distribution with mean $z_{i,r}^{(t-1)}$, then we rely on a M-H decision rule: if $\tilde{z}_{i,r}^t$ is accepted we set $z_{i,r}^t = \tilde{z}_{i,r}^t$, otherwise $z_{i,r}^t = z_{i,r}^{t-1}$. Once all $z_{i,r}$ are sampled, we move to the sampling of β^*. By specifying a normal prior and a normal likelihood we can now easily sample β^t from a multivariate normal distribution. A diffuse prior for σ_ϵ^2 allows us to sample it from an inverse Chi-squared distribution. We follow the Bayesian spatial econometric literature by sampling ϕ^L, ϕ^S from uniform distributions with support $[-\kappa_L, \kappa_L]$ and $[-\kappa_S, \kappa_S]$, as defined above. A M–H step is then performed over a normal likelihood: if accepted, then $\phi^{S^t} = \tilde{\phi}^{S^t}$ and $\phi^{L^t} = \tilde{\phi}^{L^t}$. For network fixed effects we deal again with normal prior and normal likelihood, so η is easily sampled from a multivariate normal. We sample $\sigma_\varepsilon^2, \sigma_{\varepsilon z}$ from a truncated bivariate normal over an admissible region Ξ such that the variance-covariance matrix is positive definite. Acceptance or rejection is determined by the usual M–H decision rule. At each of the M–H steps, the algorithm accepts the new values if the likelihood is higher than the current one.

References

Akerlof, G.A., Kranton, R.E., 2002. Identity and schooling: some lessons for the economics of education. J. Econ. Lit. 40, 1167–1201.

Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A., Walters, C.R., 2010. Inputs and impacts in charter schools: KIPP Lynn. Am. Econ. Rev. Papers Proc. 100, 239–243.

Angrist, J.D., Dynarski, S.M., Kane, T.J., Pathak, P.A., Walters, C.R., 2012. Who benefits from KIPP? J. Policy Anal. Manag. 31, 837–860.

Arduini, T., Patacchini, E., Rainone, E., 2014. Identification and Estimation of Outcome Response with Heterogeneous Treatment Externalities. Bank of Italy Temi di Discussione (Working Paper) No. 974.

Battu, H., Zenou, Y., 2010. Oppositional identities and employment for ethnic minorities. Evidence for England. Econ. J. 120, F52–F71.

Bekker, P., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62, 657–681.

Bekker, P., van der Ploeg, J., 2005. Instrumental variable estimation based on grouped data. Stat. Neerlandica 59, 239–267.

Bifulco, R., Fletcher, J.M., Ross, S.L., 2011. The effect of classmate characteristics on post-secondary outcomes: evidence from the Add Health. Am. Econ. J. Econ. Policy 3, 25–53.

Blume, L.E., Brock, W.A., Durlauf, S.N., Ioannides, Y.M., 2011. Identification of social interactions. In: Benhabib, J., Bisin, A., Jackson, M.O. (Eds.), In: Handbook of Social Economics, vol. 1B. Elsevier Science, Amsterdam, pp. 853–964.

Boucher, V., Fortin, B., 2015. Some challenges in the empirics of the effects of networks. In: Bramoulle', Y., Rogers, B.W., Galeotti, A. (Eds.), Oxford Handbook on the Economics of Networks. Oxford University Press, Oxford, pp. 277–302.

Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. J. Econ. 150, 41–55.

[44] The algorithm is robust to different starting values. However, speed of convergence may increase significantly.

Brooks-Gunn, J., Duncan, G.J., Kato Klebanov, P., Sealand, N., 1993. Do neighborhoods influence child and adolescent development. Am. J. Sociol. 99, 353–395.
Calvó-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. Rev. Econ. Stud. 76, 1239–1267.
Casella, G., Robert, C., 2004. Monte Carlo Statistical Methods. Springer Verlag, Berlin.
Chetty, R., Hendren, N., Katz, L., 2016. The effects of exposure to better neighborhoods on children: new evidence from the Moving to Opportunity experiment. Am. Econ. Rev. 106, 855–902.
Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. J. Econ. 75, 79–97.
Chow, G.C., 1960. Test of equality between sets of coefficients in two linear regressions. Econometrica 28, 591–603.
Coleman, J., 1988. Social capital in the creation of human capital. Am. J. Sociol. 94, 95–120.
Corcoran, M., Gordon, R., Laren, D., Solon, G., 1992. The association between men's economic status and their family and community origins. J. Human Resour. 27, 575–601.
Currarini, S., Jackson, M.O., Pin, P., 2009. An economic model of friendship: homophily, minorities, and segregation. Econometrica 77, 1003–1045.
Currarini, S., Jackson, M.O., Pin, P., 2010. Identifying the roles of race-based choice and chance in high school friendship network formation. Proc. Natl. Acad. Sci. U.S.A. 107, 4857–4861.
Datcher, L., 1982. Effects of communuty and family background on achievement. Rev. Econ. Stat. 64, 32–41.
DeGroot, M.H., 1974. Reaching a consensus. J. Am. Stat. Assoc. 69, 118–121.
De Giorgi, G., Pellizzari, M., Redaelli, S., 2010. Identification of social interactions through partially overlapping peer groups. Am. Econ. J. Appl. Econ. 2, 241–275.
DeMarzo, P., Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. Q. J. Econ. 118, 909–968.
Donald, S.G., Newey, W.K., 2001. Choosing the number of instruments. Econometrica, 1161–1191.
Durlauf, S.E., 2004. Neighborhood effects. In: Henderson, J.V., Thisse, J.-F. (Eds.), In: Handbook of Regional and Urban Economics, vol. 4. Elsevier Science, Amsterdam, pp. 2173–2242.
Epple, D., Romano, R., Zimmer, R., 2016. Charter schools: a survey of research on their characteristics and effectiveness. In: Hanushek, E.A., Machin, S.J., Woessmann, L. (Eds.), In: Handbook of the Economics of Education, vol. 5. Elsevier Science, Amsterdam, pp. 139–237.
Evans, W.N., Oates, W.E., Schwab, R.M., 1992. Measuring peer group effects: a study of teenage behavior. J. Pol. Econ. 100, 966–991.
Fordham, S., Ogbu, J.U., 1986. Black student' school success: coping with the burden of acting White. Urban Rev. 18, 176–206.
Fryer Jr, R.G., 2014. Injecting charter school best practices into traditional public schools: Evidence from field experiments. Q. J. Econ. 129, 1355–1407.
Fryer Jr., R.G., Torelli, P., 2010. An empirical analysis of 'acting white'. J. Public Econ. 94, 380–396.
Goldsmith-Pinkham, P., Imbens, G.W., 2013. Social networks and the identification of peer effects. J. Business Econ. Stat. 31, 253–264.
Golub, B., Jackson, M.O., 2010. Naive learning in social networks and the wisdom of crowds. Am. Econ. J. Microecon. 2, 112–149.
Golub, B., Jackson, M.O., 2012. How homophily affects the speed of learning and best-response dynamics. Q. J. Econ. 127, 1287–1338.
Graham, B.S., 2015. Methods of identification in social networks. Annu. Rev. Econ. 7, 465–485.
Granovetter, M.S., 1973. The strength of short-lived ties. Am. J. Sociol. 78, 1360–1380.
Granovetter, M.S., 1974. Getting A Job: A Study of Contacts and Careers. Harvard University Press, Cambridge, MA.
Granovetter, M.S., 1983. The strength of short-lived ties: a network theory revisited. Sociol. Theory 1, 201–233.
Griffith, A.L., Rask, K.N., 2014. Peer effects in higher education: a look at heterogeneous impacts. Econ. Educ. Rev. 39, 65–77.
Hansen, B.E., 2000. Testing for structural change in conditional models. J. Econ. 97, 93–115.
Hansen, B.E., 2001. The new econometrics of structural change: dating breaks in U.S. labor productivity. J. Econ. Perspect. 15, 117–128.
Hansen, C., Hausman, J., Newey, W., 2008. Estimation with many instrumental variables. J. Business Econ. Stat. 26, 398–422.
Harris, K.M., 2013. The Add Health Study: Design and Accomplishments. Carolina Population Center, University of North Carolina at Chapel Hill.
Hsieh, C.-S., Lee, L.-F., 2015. A social interaction model with endogenous friendship formation and selectivity. J. Appl. Econ. 31, 301–319.
Ioannides, Y.M., 2011. Neighborhood effects and housing. In: Benhabib, J., Bisin, A., Jackson, M.O. (Eds.), In: Handbook of Social Economics, vol. 1B. Elsevier Science, Amsterdam, pp. 1281–1340.
Ioannides, Y.M., 2012. From Neighborhoods to Nations: The Economics of Social Interactions. Princeton University Press, Princeton.
Ioannides, Y.M., Topa, G., 2010. Neighborhood effects: accomplishments and looking beyond them. J. Reg. Sci. 50, 343–362.
Jackson, M.O., 2008. Social and Economic Networks. Princeton University Press, Princeton.
Jackson, M.O., 2014. Networks in the understanding of economic behaviors. J. Econ. Perspect. 28, 3–22.
Jackson, M.O., Rogers, B.W., Zenou, Y., 2017. The economic consequences of social network structure. J. Econ. Lit. 55 (1) (forthcoming).
Jackson, M.O., Zenou, Y., 2015. Games on networks. In: Young, P., Zamir, S. (Eds.), In: Handbook of Game Theory, vol. 4. Elsevier, Amsterdam, pp. 91–157.
Lai, G., Lin, N., Leung, S.-Y., 1998. Network resources, contact resources, and status attainment. Social Netw. 20, 159–178.
Lavy, V., Sand, E., 2016. The effect of social networks on student's academic and non-cognitive behavioral outcomes: evidence from conditional random assignment of friends in school. Unpublished manuscript. University of Warwick.
Lee, L.-F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econ. Theory 18, 252–277.
Lee, L.-F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econ. Theory 26, 187–230.
Lee, L.-F., Liu, X., Lin, X., 2010. Specification and estimation of social interaction models with network structures. Econ. J. 13, 145–176.
Lin, X., 2010. Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables. J. Labor Econ. 28, 825–860.
Lin, N., Ensel, W.M., Vaughn, J.C., 1981. Social resources and strength of ties: structural factors in occupational status attainment. Am. Sociol. Rev. 46, 393–405.
Liu, X., Lee, L.-F., 2010. GMM estimation of social interaction models with centrality. J. Econ. 159, 99–115.
Liu, X., Patacchini, E., Zenou, Y., Lee, L.-F., 2012. Criminal Networks: Who Is the Key Player? CEPR Discussion Paper No. 8772.
Manski, C.F., 1993. Identification of endogenous effects: the reflection problem. Rev. Econ. Stud. 60, 531–542.
Marsden, P.V., Hurlbert, J.S., 1988. Social resources and mobility outcomes: a replication and extension. Social Forces 66, 1038–1059.
Patacchini, E., Zenou, Y., 2008. The strength of weak ties in crime. Eur. Econ. Rev. 52, 209–236.
Patacchini, E., Zenou, Y., 2016. Racial identity and education in social networks. Social Netw. 44, 85–94.
Putnam, R., 2000. Bowling Alone: The Collapse and Revival of American Community. Simon and Schuster, New York.
Sacerdote, B., 2001. Peer effects with random assignment: results from Dartmouth roomates. Q. J. Econ. 116, 681–704.
Sacerdote, B., 2011. Peer effects in education: how might they work, how big are they and how much do we know thus far? In: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), In: Handbook of Economics of Education, vol. 3. Elevier Science, Amsterdam, pp. 249–277.
Sacerdote, B., 2014. Experimental and quasi-experimental analysis of peer effects: two steps forward? Annu. Rev. Econ. 6 (1), 253–272.
Sigelman, L., Bledsoe, T., Welch, S., Combs, M.W., 1996. Making contact? Black–White social interaction in an urban setting. Am. J. Sociol. 101, 1306–1332.
Stock, J.H., Wright, J.H., Yogo, M., 2012. A survey of weak instruments and weak identification in generalized method of moments. J. Business Econ. Stat. 20, 518–529.
Stock, J.H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Andrews, D.W.K., Stock, J.H. (Eds.), Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, Cambridge, pp. 80–108.
Tincani, M.M., 2015. Heterogeneous Peer Effects and Rank Concerns: Theory and Evidence. Unpublished manuscript, University College London.
Topa, G., 2001. Social interactions, local spillovers and unemployment. Rev. Econ. Stud. 68, 261-295.1.

Tuch, S.A., Sigelman, L., Macdonald, J.A., 1999. Trends: race relations and America youth, 1976–1995. Public Opin. Q. 63, 109–148.
Wasserman, S., Faust, K., 1994. Social Network Analysis. Methods and Applications. Cambridge University Press, Cambridge.
Wellman, B., Wortley, S., 1990. Different strokes from different folks: community ties and social support. Am. J. Sociol. 96, 558–588.
Wilson, W.J., 1987. The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy. University of Chicago Press, Chicago.
Yakubovich, V., 2005. Short-lived ties, information, and influence: how workers find jobs in a local Russian labor market. Am. Sociol. Rev. 70 (3), 408–421.
Yakusheva, O., Kapinos, K.A., Eisenberg, D., 2014. Estimating heterogeneous and hierarchical peer effects on body weight using roommate assignments as a natural experiment. J. Human Resour. 49, 234–261.
Zax, J.S., Rees, D.I., 2002. IQ, academic performance, environment, and earnings. Rev. Econ. Stat. 84, 600–616.
Zenou, Y., 2013. Spatial versus social mismatch. J. Urban Econ. 74, 113–132.
Zenou, Y., 2015. A dynamic model of weak and strong ties in the labor market. J. Labor Econ. 33, 891–932.