An Aspiring Friend Is a Friend Indeed: On the Mechanisms Behind Peer Influences on Human Capital Accumulation

Jessica Gagete-Miranda *

Abstract

This paper studies friends' influence on the process of human capital accumulation and the mechanisms behind such an influence. I combine novel data on Brazilian students' networks with administrative data and investigate whether friends' high school completion impacts students' high school completion. The employed methodology acknowledges that social cliques are formed endogenously and models friendship formation based on students' unobserved social ability and their random interaction opportunities. I then use the attributes of predicted friends of friends as instrumental variables for friends' school completion. The results show substantial peer effects: An extra friend graduating from high school increases the likelihood of students' graduation by 6.62 percent. Friends' influence is greater for black students and students whose mothers did not complete high school. Focusing on the mechanisms behind such an impact, I show that aspirations and effort in the school spread through students' networks, but perceptions about schooling returns and fear of stigma do not. A mediation analysis underscores the role of aspirations as the primary driver of friends' influence.

Keywords: Peer effects, Aspirations, Human Capital Accumulation

JEL Codes: I24, I25, I29

---

*Department of Economics, Management and Statistics, University of Milano-Bicocca – Piazza dell'Ateneo Nuovo, 1 - 20126, Milan, Italy; jessica.gagetemiranda@unimib.it

I thank Fernando Botelho, Ricardo Madeira, and Marcos A. Rangel for sharing the survey used in this work. Botelho, Madeira and Rangel had IRB approval for the survey from Fundacao Instituto de Pesquisas Economicas BR-T1158/FIPE, granted on the July 26, 2010.

I am grateful to Marcos A. Rangel, Paolo Pinotti, Eliana La Ferrara, and Massimo Anelli for their guidance over the course of this project. I also appreciate the valuable comments of Sule Alan, Samuel Berlinski, Matias Busso, Sebastian Calonico, Jane Cooley Fruehwirth, Armin Falk, Seth Gershenson, Catherine Michaud-Leclerc, Michele Pellizzari, Stephen L. Ross, Ana Trindade Ribeiro, and participants at the Summer School on Socioeconomic Inequality, Alp-Pop Conference, RIDGE Forum, the Empirics and Methods in Economics Conference, APPAM Fall Research Conference, PacDev Conference, MWIEDC Conference, IZA Workshop in the Economics of Education, the E-con of Education Student Conference, the AMSE Summer School in The Economics of Networks, the EALE Conference, the SEA Annual Meeting, the Psychology & Economics of Poverty Convening, the NOVAFRICA Conference, the Naples Workshop in Networks and Development, and seminars at Bocconi University, Brunel University London, the Center for Evaluation and Development, the Duke Network Analysis Center, Economic Research Southern Africa, Insper Institute of Education and Research, the Sanford School of Public Policy at Duke, Universidad del Chile, the University of Milano-Bicocca, and Universidad del Rosario. The usual caveats apply.

This paper circulated before with the title "An aspiring friend is a friend indeed: school peers and college aspirations in Brazil".

1 Introduction

Under-investment in human capital among disadvantaged youth is widespread across developed and developing countries and has detrimental consequences for growth and inequality (Galor, 2011; Galor and Moav, 2004). While the bottleneck in developed countries is usually the acquisition of tertiary education (e.g. Bailey and Dynarski, 2011), developing countries struggle with an even more complex challenge: the low completion rates of secondary education. In Brazil, for instance, about 33% of the adult population has not completed upper secondary education – or high school –, and about 19% of the youth in the age range for upper secondary education was out of school in 2017 (OECD, 2019).

Although financial constraints are an essential driver of low investment in human capital, psychological and informational barriers frequently keep students from achieving higher levels of education even in the absence of such constraints (Dynarski et al., 2021; Hoxby and Avery, 2012; Jensen, 2010; Kearney and Levine, 2014; La Ferrara, 2019). In this context, individuals' social networks might play a crucial role in either breaking such barriers or making them even more salient through their influence on students' aspirations, effort, performance, and perceptions about schooling returns. Indeed, recent literature has shed light on the importance of peers and family for educational choices (e.g. Altmejd et al., 2021; Barrios-Fernández, 2022). However, data and methodological challenges make the evidence of the mechanisms behind such an influence scarce.

The present paper combines administrative data with a unique social network dataset collected among middle-school students in Brazil to answer two questions. First, do school friends impact students' likelihood of graduating from high school? Second, what are the mechanisms behind such an impact?

To answer the first question, I investigate how students' decision to graduate from high school depends on their friends' graduation decisions. I address the reflection problem (Manski, 1993) and the endogeneity of friendship formation by leveraging the network structure of my data. Specifically, I first model friendship formation based on students' unobserved social ability and their random opportunities to interact in the school. I document that, after controlling for senders' and receivers' fixed effects, a student i (sender) is more likely to befriend another student j (receiver) if both i and j share their first name initial. This characteristic is an important predictor of friendship formation because, as I show, students tend to be allocated into classes in alphabetical order when they first enroll in middle schools, which randomly increases their chances of meeting and interacting. Next, based on the model's predicted connections, the identification strategy addresses the reflection problem by using predicted friends of friends' characteristics as instrumental variables for friends' outcomes (Bramoullé et al., 2009; De Giorgi et al., 2010). It also uses network fixed effects and a broad set of controls to eliminate other possible correlated effects.

I find that an extra friend graduating from high school increases, on average, 6.62 percent the likelihood that a student will also graduate from high school. This impact is even higher (14.52 percent) if we consider the likelihood of graduating from high school without being retained in any grade during that cycle. I also find that friends' influence is greater for low-SES students, perhaps those with lower support from family or school and who need to rely the most on their friends.

To answer the second question – on the mechanisms behind friends' influence on high school completion–, I exploit rich information on students' aspirations, perceptions about schooling returns, fear of stigma, and school effort. These variables represent possible psychological and informational factors that enter the students' human capital production function. I first investigate endogenous social effects in these variables and find that aspirations and school effort spread throughout students' social networks. Interestingly, friends do not seem to share information about schooling returns, nor do friends influence students' performance or fear of being stigmatized as "nerds".

Next, I perform a mediation analysis to understand whether these proxies of psychological and informational factors explain friends' influence on human capital investments. I show that once these variables are accounted for, the impact of friends' high school completion on students' high school completion vanishes. Moreover, students' and friends' aspirations represent the largest share of the total mediated effect, indicating that aspirations are a crucial mechanism behind friends' influence on human capital investments.

The present paper contributes to the literature on peer effects on human capital accumulation. Traditional contributions to the literature on peer effects focus on outcomes such as school or college performance (see Sacerdote (2011) and Sacerdote (2014) for comprehensive reviews). Results in this setting are context-dependent and vary considerably, with prominent peer effects in some cases and absent or negligible peer effects in others. A more recent stream of this literature has focused on peers' influence on schooling decisions and has found larger and more consistent effects.[1] Many papers in this second stream of the literature focus on the impact of older siblings or neighbors on human capital investment decisions (Aguirre and Matta, 2021; Altmejd et al., 2021; Barrios-Fernández, 2022; Dahl et al., 2020; Dustan, 2018,1; Joensen and Nielsen, 2018; Qureshi, 2018). Other contributions focus on the impact of peers' characteristics or peers whose incentives to accumulate human capital suddenly increased (Abramitzky et al., 2021; Anelli and Peri, 2019; Anelli et al., 2022; Ballis, 2020; Bobonis and Finan, 2009; Brenøe and Zölitz, 2020; Cipollone and Rosolia, 2007; Cools et al., 2021; Feld and Zölitz, 2022; Pagani and Pica, 2021; Zölitz and Feld, 2021).

This paper contributes to this literature in two ways. First, it adds another piece of evidence on how peers are influential on students' schooling decisions. This time, however, it focuses on *friends*, a source of influence on human capital investments overlooked by the literature on the topic so far. Friends' have been shown to impact individuals' productivity (e.g. Bandiera et al., 2010) and performance (e.g. Calvó-Armengol et al., 2009; Fletcher et al., 2020). However, to the best of my knowledge, this is the first paper to show that they also impact schooling decisions. Second, and most importantly, this paper exploits rich information on several students' characteristics to investigate the mechanisms behind friends' influence. The studies that have found positive peer effects on schooling decisions

[1] However, Feld and Zölitz (2022) and Anelli and Peri (2019) did not find strong peers influence even in this setting.

were not able, so far, to comprehend the channels leading to this effect. In particular, many of these studies explicitly say that they cannot disentangle information diffusion from changes in preferences (e.g. Altmejd et al., 2021; Barrios-Fernández, 2022). This paper shows that while aspirations and effort in the school spread through students' networks, information about perceived schooling returns does not. While I cannot test spillovers in other types of preferences or information, the findings in this paper suggest that changes in preferences are more relevant than the spread of information.

The mechanisms' analysis of this paper also sheds light on the literature investigating the determinants of aspirations and their role in individuals' outcomes (see Fruttero et al., 2021; Mani and Riley, 2019, for comprehensive reviews). Theoretical contributions discuss how the capacity to aspire to a better standard of living is an essential driver of individuals' efforts and investments (Appadurai, 2004; Ray, 2006) and consider the lack of such a capacity – or aspirations failures – as a psychological constraint that might trap people in poverty (Dalton et al., 2016; Genicot and Ray, 2017). When it comes to the determinants of aspirations, a large body of empirical contributions has found that peers' socioeconomic status is associated with individuals' aspirations (e.g. Galiani et al., 2021; Janzen et al., 2017; Stutzer, 2004). Fewer papers, however, have focused on how peers' aspirations influence one's own aspirations (but see Boucher et al., 2018; Dickerson et al., 2018; Mora and Oreopoulos, 2011; Norris, 2020, for notable exceptions). This paper estimates how friends' college aspirations influence students' college aspirations. Moreover, to the best of my knowledge, this is the first paper to provide evidence that the aspirations of both students and their friends are important predictors of students' future schooling decisions and the main mechanisms behind friends' influence on such decisions. Given how challenging it is to identify the role of aspirations on individuals' outcomes, these findings represent an essential step in this direction.

Finally, this paper contributes to flourishing literature that has combined estimations of peer effects on networks and models of network formation to address endogenous link formation on the estimation of peer effects (see Bramoullé et al., 2020, for a comprehensive survey about this topic). Specifically, my identification strategy relates closely to papers that, first, model link formation exploiting individuals' pre-determined characteristics and, second, use the predicted links from such a model to estimate peer effects (Fletcher et al., 2020; Santavirta and Sarzosa, 2019). Besides leveraging pre-determined characteristics, I indirectly exploit individuals' random chances to interact due to quasi-experimental allocations of students into classes and show that such interactions are important predictors of friendship formation.

2 Data and background

The primary data source used in this work is a survey conducted on students enrolled in the ninth grade of state-operated middle schools in Sao Paulo, Brazil, in 2011. I combine this survey with administrative data to recover information on students' socioeconomic background, performance, and school path – each school, grade, and class in which students were enrolled throughout their education. Information on students' school paths allows me to verify whether they finished high school or not. I build two different measures for high school completion. The first – "HS completion" – is a binary variable indicating whether students graduated from high school at any time in the future. The second – "HS completion without retention" – is a binary variable indicating whether students graduated from high school without being retained in any grade during that cycle.

In what follows, I first provide some background information about the Brazilian educational system and relevant details about the provision of public education in the State of Sao Paulo. Then, I describe the main characteristics of the survey and present some descriptive statistics.

2.1 Institutional background

Basic education in Brazil is divided into preschool (attended by students up to the age of six), primary school (attended by six- to 14-year-olds), and secondary or high school (attended by 15- to 17-year-olds). Primary school is the only mandatory level of education in Brazil. It is subdivided into two levels: elementary school – grades one to five – and middle school – grades six to nine. The main subjects taught in primary school are the Portuguese language, mathematics, social sciences, and sciences. Students are assigned to classes at the beginning of each academic year. They study all subjects together with the same classmates for the entire year.

In most Brazilian states, and particularly in the State of Sao Paulo, municipal governments are generally responsible for elementary schools, while the state government is typically responsible for middle and high schools. Hence, the vast majority of students must change schools in the transition from fifth to sixth grade. State-operated schools in Sao Paulo are usually larger than municipal-operated schools. In 2011, when the relevant survey was conducted, the average number of students enrolled in a state-operated school was 1,189, while the average number of students enrolled in a municipal-operated school was 697. Class size is also larger in state-operated schools, with an average of 39 students per class in 2011; this figure was 27 students per class for municipal-operated schools in that same year.[2]

Because of the decentralized nature of the educational system in Sao Paulo, state-operated schools receive virtually no information about students from municipal-operated schools in their transition from fifth to sixth grade. Hence, when students enroll in a state-operated school in the sixth grade, that school's administration does not have information about these students' backgrounds, such as previous performance or behavior. Since the assignment of students in sixth-grade classes cannot take their characteristics into consideration, they are often assigned according to the alphabetical order of their names.[3] As Section 3 shows, I

[2]Source: 2011 school census (http://portal.inep.gov.br/censo-escolar).

[3]Evidence from students in my sample corroborates this pattern. I show in section 3 that students who share their first-name initials have a higher likelihood of being assigned to the same class in the sixth grade.

exploit this feature in the Sao Paulo educational system to model friendship formation.

2.2 Survey on students' profiles and friendship ties

In 2011, students in the ninth grade of selected state-operated schools in Sao Paulo answered a comprehensive questionnaire about their personal profiles, happiness or satisfaction with their lives, study habits, perceptions about schooling returns, and educational aspirations.[4]

One block of questions in the survey mapped students' social networks. They were asked to name their four closest friends or colleagues in their grade (which, in most schools, comprehends more than one classroom).[5] Importantly, it is possible to link the named students to school rosters and also to locate their own answers to the questionnaire. The survey sampled all students in the ninth grade of each selected school, and administrative records are available for all students in state-operated schools, which makes it possible to map the network and characteristics of all ninth-graders in each school.

Besides leveraging information on students' networks to identify friends' influence on high school completion, I exploit other questions in the survey to investigate the mechanisms behind such influence, as discussed in section 5. These questions aimed to extract essential inputs that students consider when deciding on their schooling investments. First, one question asked how many years the students would like to keep studying if this choice were *entirely* up to them. I use this question to create a measure of aspirations toward pursuing a college degree, called "college aspirations".[6] This is a binary variable that takes a value equal to one if students answered that they would like to keep studying until they get a college degree and zero otherwise.

[4]This survey was conducted by professors from the University of Sao Paulo with funding from the Inter-American Development Bank.

[5]On average, there were four classrooms in each school.

[6]The framing of such a question is relevant: Since students were asked to reveal their preferences about their educational future regardless of any constraint they might face, this measure is more likely to capture students' genuine aspirations and not merely their expectations for the future. Of course, completely disentangling aspirations from expectations is challenging, and an in-depth discussion about such a difference is not in the scope of this work (see Fruttero et al., 2021, for a comprehensive discussion on aspirations' measures.)

Second, another question asked students to indicate their probability of finding jobs if they finish high school, vis-a-vis dropping out of school before obtaining such a level of education.[7] I call this variable "perceived HS returns" since it measures students' perception of the labor market returns from finishing high school.

Third, students were asked about possible impediments to future educational pursuits. One impediment, in particular, might both be impacted by students' networks and related to their schooling decisions: the students' concern about being stigmatized as "nerds" if they put too much effort into studying. I call this variable the "fear of nerd stigma." Such an impediment might proxy for students' willingness to comply with rather harmful social norms in the school that are potentially detrimental to human capital accumulation.

Finally, to proxy for the effort students put into school, a question of the survey asked them how long per day they study math outside school hours. I create a binary variable indicating whether they study at least half an hour per day, and I call this variable "30+ min math study/day".

Table 1 presents some descriptive statistics derived from this survey and administrative data. The table shows the mean and standard deviation for all students and separately for those who concluded high school (at any time in the future) and those who did not. First, the sample composed of all students reveals that only about 79% of them went on in their studies to complete high school. Such a figure is in line with the average high school completion in Brazil and shows a worrisome pattern of low investments in human capital in the country (OECD, 2019). Second, comparing students who graduated from high school with students who did not, we see that those with high school completion are, on average, higher achievers and have better-educated parents. Their aspirations and perceptions about college return are also higher, and their willingness to comply with harmful social norms in the school is lower. School effort, in turn, is no different among the two groups of students.

[7] This question was framed in the following way. First, students were asked to think of 10 other students very similar to them in the school. They were then asked to indicate how many of these students would find a job depending on how far they kept studying.

Table 1: Descriptive Statistics

	All		HS completion =1		HS completion=0	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
HS completion	0.79	0.41	1.00	0.00	0.00	0.00
HS completion w/out retention	0.64	0.48	0.81	0.39	0.00	0.00
Girl	0.49	0.50	0.53	0.50	0.38	0.48
White	0.33	0.47	0.35	0.48	0.30	0.46
Mother education: more than HS	0.24	0.43	0.27	0.44	0.15	0.36
Father education: more than HS	0.22	0.41	0.24	0.43	0.16	0.37
Father works	0.73	0.44	0.75	0.43	0.69	0.46
Reading proficiency (2009)	-0.00	1.00	0.10	1.01	-0.38	0.88
Math proficiency (2009)	-0.00	1.00	0.08	1.01	-0.31	0.89
Reading proficiency (2011)	-0.00	1.00	0.12	1.00	-0.44	0.87
Math proficiency (2011)	0.00	1.00	0.10	1.01	-0.36	0.87
College aspiration	0.68	0.46	0.73	0.45	0.53	0.50
Perceived HS returns	0.50	0.27	0.51	0.27	0.47	0.26
Fear of nerd stigma	0.26	0.44	0.23	0.42	0.35	0.48
30+ min math study/day	0.40	0.49	0.40	0.49	0.39	0.49
Named friends	2.02	1.41	2.14	1.39	1.59	1.40
Observations	6075		4771		1304	
Number of schools	85					

Note: (i) Math and Reading proficiency are standardized with Mean=0 and SD=1; (ii)"College aspiration" is a binary variable indicating that the students would like to have a college degree, "Perceived HS returns" is the likelihood students' attribute for them to find a job if they finish high school vis-a-vis drooping out from school before it, "Fear of nerd stigma" is a binary variable indicating that the student would like to put more effort into studying but does not due to the fear of being stigmatized as a nerd, "30+ min math study/day" is a binary variable indicating that the student studies math outside of school hours at least 30 min per day, "Named friends" is the number of friends in the ninth grade named by the student.

3 Identification of peer effects

Researchers face several challenges when identifying endogenous social effects – that is, the impact that peers' outcomes have on one's outcomes – through a linear-in-means model. The first challenge is the reflection problem (Manski, 1993): a simultaneity bias that emerges because an individual might influence the behavior of their group and, at the same time, be influenced by the group's behavior. For instance, in a friendship network, all friends potentially impact each other, so it is hard to discern if one's behavior is the cause or consequence of the group's behavior.

The second challenge, also discussed by Manski (1993), is correlated effects, whereby people in the same reference group tend to behave alike not because they influence one another but because they share similar unobserved characteristics. For instance, students within a school are all influenced by school quality or an inspiring teacher.

Finally, a specific instance of correlated effects is that connections or friendship links do not happen randomly, which makes reference groups themselves endogenous. Several works have shown the central role of homophily in friendship formation. That is, the likelihood that two people will interact with one another is higher if they share similar characteristics, like race or SES (Alan et al., 2020; Currarini et al., 2009; Mayer and Puller, 2008; Moody, 2001; Weinberg, 2007). An important implication of homophily and the endogenous formation of networks is that neither the connections nor the influence of individuals inside a reference group is equal for everyone. Even students enrolled at the same school and receiving instruction from the same teachers form different cliques. This brings extra challenges to estimating peer effects since individuals might have unobserved characteristics correlated to both their outcomes and their link formation.

Several works in the peer-effects literature use different strategies to tackle these identification problems. Some use natural experiments to solve correlated effects (Cipollone and Rosolia, 2007; Sacerdote, 2001; Zimmerman, 2003), others use theoretical models of social interactions (Brock and Durlauf, 2001) or network structures (Boucher et al., 2014; Bramoullé et al., 2009; Calvó-Armengol et al., 2009; De Giorgi et al., 2010; Liu et al., 2014) to address both correlated effects and the reflection problem.

Fewer works have relaxed the assumption of strictly exogenous networks and fully acknowledged the implications of the endogenous formation of links. Flourishing literature, however, has combined estimations of peer effects on networks and models of network formation – leveraging contributions such as the model proposed by Graham (2017) – to address endogenous link formation on the estimation of peer effects.[8] A stream of this literature adopts a control function approach that models link formation and controls for it in the estimation of peer effects. (Goldsmith-Pinkham and Imbens, 2013; Griffith, 2016; Hsieh and Lee, 2016; Johnsson and Moon, 2021; Qu and Lee, 2015)

My paper relates more closely to another stream of this literature that models link formation exploiting individuals' pre-determined characteristics and uses the predicted links from such a model to estimate peer effects. Fletcher et al. (2020), for instance, estimate a non-parametric model of within-grade and school friendship links based on similarities in the characteristics of students' mothers and use this model to predict the number of a student's friends with a four-year college-educated mother. They then estimate the impact of friends' maternal education on students' academic performance. Santavirta and Sarzosa (2019) employ a similar estimation but use individuals' similarities in characteristics at the time of birth to model link formation. Other contributions outside the literature on the economics of education have also implemented similar strategies. König et al. (2018), for instance, model link formation based on past network structures as exclusion restrictions that affect current link formation but do not enter the network effects outcome equation. The authors then use the predicted links coming from this model to employ an instrumental variables strategy to estimate technology spillovers in firms' networks.

In this paper, I first model friendship formation based on students' unobserved social ability – or degree heterogeneity – and their exogenous chances of interacting due to as-

[8] See Bramoullé et al. (2020) for a comprehensive survey about this topic.

good-as-random assignments to classes when they enroll in middle school. Next, I perform estimations similar to the ones implemented by Bramoullé et al. (2009) and De Giorgi et al. (2010) in which friends' outcomes are instrumented by friends' of friends characteristics. The main difference is that, when building the instruments, I replace the endogenous sociometric matrix with the predicted one from the link formation model.

3.1 Model of friends' influence

Let a student's high school completion be affected by the average high school completion of their friends, their characteristics, such as previous performance, gender, race, and family background, and the average characteristics of their friends. More formally, suppose there is a set of students i, $i = (1...N)$, that belongs to network l, $l = (1, ...L)$[9]. Each student may have a group of friends F_i of size n_i or may be isolated, where $F_i = \varnothing$. Assume that each student i is not included in their own group of friends, such that $i \notin F_i$.[10] The model is given by:[11]

$$
\begin{aligned}
y_{li} = \beta \frac{\sum_{j \in F_i} y_{lj}}{n_i} + \gamma x_{li} + \eta \frac{\sum_{j \in F_i} x_{lj}}{n_i} + \mu_l + \upsilon_{li} \\
E(\upsilon_{li} | \boldsymbol{X_l}, \mu_l) = 0
\end{aligned}
\tag{1}
$$

where y_{li} is the high school completion status of individual i in network l , which depends on the average high school completion of the friends directly connected to them[12] – the

[9]In this study, each network is formed by all students in 9^{th} grade of each school.

[10]The exclusion of individuals from their own reference group might lead to yet another source of bias, namely the exclusion bias, that causes an underestimation of peer effects (Caeyers and Fafchamps, 2016; Guryan et al., 2009). The exclusion of an individual i from the pool of i's peers creates a negative mechanical relationship between i's characteristics and that of their peers, especially in small samples. The identification strategy adopted in this work – that follows the works of Bramoullé et al. (2009) and De Giorgi et al. (2010) – also addresses this source of bias. For more details, see the work of Caeyers and Fafchamps (2016).

[11]This model reassembles the one described in Bramoullé et al. (2009) and is a special case of the model described in Manski (1993), in which an individual reference group is the friends linked to them.

[12]Boucher and Bramoullé (2021) show that linear models of peer effects, traditionally used to study continuous outcomes, can also be used for binary outcomes. In particular, they show that the identification results of Bramoullé et al. (2009) apply when outcomes are binary.

endogenous social effects in Manski's notation (see Manski (1993)) –, on x_{li}, their own characteristics[13], on the average characteristics of their friends – the exogenous social effects in Manski's notation – and on network unobserved fixed effects, μ_l. The only restriction imposed to parameters in this model is that $|\beta| < 1$, so that its reduced form is identifiable.

Let G be the adjacency matrix, where element $g_{i,j} = 1/n_i$ if individual i sends a friendship tie to individual j, and $g_{i,j} = 0$ otherwise. Assume that $g_{i,i} = 0$ so that each individual is not part of their own reference group. The above model can then be translated to:

$$\begin{aligned} \boldsymbol{y_l} = \beta \boldsymbol{G} \boldsymbol{y_l} + \gamma \boldsymbol{X_l} + \eta \boldsymbol{G} \boldsymbol{X_l} + \mu_l + \boldsymbol{\upsilon_l} \\ E(\upsilon_l | \boldsymbol{X_l}, \mu_l) = 0 \end{aligned} \tag{2}$$

It is easy to see that the reflection problem emerges because the outcome variable y is present on both sides of the equation. To be more explicit, if one assumes that $\boldsymbol{G}$ is orthogonal to $\boldsymbol{\upsilon_l}$ (I will relax this assumption next), it is possible to causally estimate the reduced form of equation 2[14]:

$$\boldsymbol{y_l} = (\boldsymbol{I} - \beta \boldsymbol{G})^{-1} (\gamma \boldsymbol{I} + \eta \boldsymbol{G}) \boldsymbol{X_l} + (\boldsymbol{I} - \beta \boldsymbol{G})^{-1} \mu_l + (\boldsymbol{I} - \beta \boldsymbol{G})^{-1} \boldsymbol{\upsilon} \tag{3}$$

However, such estimation will yield only unbiased estimates of $(\boldsymbol{I} - \beta \boldsymbol{G})^{-1} \eta$, which will not disentangle the endogenous social effects (β) from the exogenous social effects (η).

Correlated effects would emerge if the modeler did not observe μ_l, since $\boldsymbol{X_l}$ is only exogenous conditional on μ_l. School quality, for instance, is probably correlated with students' schooling decisions. Hence, students within the same school are more likely to have similar levels of high school completion, which could bias estimations upwards. This problem is addressed by simply controlling the estimations by network fixed effects – in this case, the same as school fixed effects (as discussed in section 3.4, the most stringent specifications of

[13] For the sake of notation clarity, there is only one exogenous characteristic included in equation 1 In the next equation, the model is generalized to more characteristics.

[14] Given the restriction on β, $\boldsymbol{I} - \beta \boldsymbol{G}$ is invertible.

the model also control for classroom fixed effects).

Nonetheless, this does not solve the endogeneity of link formation. That is, individuals do not befriend one another at random, and homophily plays a significant role in friendship formation, which yields $\boldsymbol{G} \not\perp \boldsymbol{v_l}$. Once again, such correlation would most likely bias estimates upward since more similar students have a greater probability of becoming friends and, at the same time, are more likely to have similar schooling paths.

I will address the reflection problem and the endogenous formation of friendship with the use of a three-stage estimation. The first stage models link formation based on students' unobserved degree heterogeneity and their exogenous chances of interacting. The second and third stages use the predicted friendship connections delivered by the first stage and use predicted friends of friends' characteristics as instrumental variables for friends' outcomes (resembling Bramoullé et al. (2009)). The remainder of this section describes this approach and explains how it overcomes the issues mentioned above. For the sake of clarity in exposition, the last two stages of the implemented strategy, which address the reflection problem, are described first. The first stage is then described, along with an explanation of how it overcomes the endogenous formation of networks.

3.2 The reflection problem

Through a series expansion of equation 3 and assuming $\beta\gamma + \eta \neq 0$, Bramoullé et al. (2009) show that if $\boldsymbol{I}$, $\boldsymbol{G}$, $\boldsymbol{G}^2$, and $\boldsymbol{G}^3$ are linear independent, it is possible to use ($\boldsymbol{G}^2\boldsymbol{X_l}$, $\boldsymbol{G}^3\boldsymbol{X_l}$, ...) as excluded instruments for $\boldsymbol{Gy}$ and, as so, to identify all the parameters of model 2.[15] The authors prove that if the diameter of the network is greater than or equal to 3, then the

[15]If correlated effects were not an issue and μ_l could be excluded from the model, this condition would be less restrictive. As a matter of fact, one would need only $\boldsymbol{I}$, $\boldsymbol{G}$, $\boldsymbol{G}^2$ to be linear independent in order for the model to be identified.

linear independence between $\boldsymbol{I}$, $\boldsymbol{G}$, $\boldsymbol{G}^2$, and $\boldsymbol{G}^3$ is guaranteed and the model is identified.[16][17]

Therefore, to identify the parameters φ =(β, η, γ), it is possible to follow a 2SLS estimation, where the matrix of explanatory variables $\tilde{\boldsymbol{X}}$ =[$\boldsymbol{Gy_l}$ $\boldsymbol{X_l}$ $\boldsymbol{GX_l}$] is instrumented in the second stage by $\boldsymbol{S}$ = [$\boldsymbol{X_l}$ $\boldsymbol{GX_l}$ $\boldsymbol{G^2X_l}$], such that the final estimates are given by $\hat{\varphi}^{2SLS} = (\boldsymbol{\tilde{X}'P\tilde{X}})^{-1}\boldsymbol{X'\tilde{P}y_l}$, where $\boldsymbol{P} = \boldsymbol{S(S'S)^{-1}S}$.

In other words, unless the network is fully connected, there will always be an individual A in the network whose characteristics will directly affect the outcome of another individual B but will affect the outcome of a third individual C only indirectly, through the friendship tie between B and C. Therefore, A's characteristics are valid instruments for B's outcomes (see section 3.4 for a discussion on possible limitations of such strategy in the present setting).

3.3 Endogenous link formation

The 2SLS strategy mentioned above would ensure unbiased estimates of the endogenous and exogenous social effects if friendship links were formed at random – that is, if $\boldsymbol{G} \perp \upsilon_l$. However, as stated before, social networks are not formed at random, and homophily plays a role in clique formation. I deal with such an issue by including a stage before the 2SLS, in which I use predicted networks based on exogenous interaction opportunities to build the IVs that identify the social effects.

My model of link formation leverages the fact that inter-classroom and within-classroom assignments are driven at least in part by students' first-name alphabetical order. First, as described in section 2, the assignment of students to classrooms when they first enroll in state-operated schools is not based on students' previous performance or behavior, since the school administrators do not have such information. Therefore, students are usually

[16] As in Bramoullé et al. (2009)[pg 47], "define the distance between two students i and j in the network as the number of friendship links connecting i and j in the shortest chain of students $i_1 ... i_l$ such that i_1 is a friend of i, i_2 is a friend of i_1, ...and j is a friend of i_l.(...) Define the *diameter* of the network as the maximal friendship distance between any two students in the network (see Wasserman and Faust (1994))."

[17] The counterpart for the diameter size in a model in which correlated effects are absent is the presence of *intransitive triads* – that is, when we have a set of three individual i, j, and k such that i is connected to j and j is connected to k but i is not connected to k - in at least some networks.

assigned in alphabetical order. Indeed, as presented in Table A.2, sharing the first-name initial is the only significant predictor of the likelihood that two students were assigned to the same classroom in sixth grade. Other similarities, such as gender, race, and parental education, are not good predictors of such assignments.[18] Second, within-class allocation– such as the choice of students' seats – is usually also done in alphabetical order. Hence, if two students share their first name initials, a variable that is arguably exogenous, they are more likely to meet and interact in the school, which increases their chances of becoming friends.

I formalize this idea by estimating a dyadic regression based on Graham (2017), which models network formation considering individuals' unobserved degree heterogeneity – that is, the fact that individuals have different, not observed, social abilities – and homophily – that is, the fact that individuals are more likely to send a friendship tie to someone similar to them. I adapt this model and consider that the friendship connection $D_{i,j}$ between two agents i and j, depends on their unobserved degree heterogeneity and on whether they share their first name initials. Clearly, the mechanism behind friendship formation, in this case, is not homophily but the exogenous increase in students' likelihood to meet and interact in the school – one might think of it as a decrease in the search cost for a friend. In Appendix A.1, I present estimations employing a model closer to the one proposed by Graham (2017), where homophily in pre-determined characteristics is the driver of friendship formation.

If we consider in the dyadic regression w_{ij} as a binary variable that takes value equal to one if i and j have the same first name initials, and zero otherwise, then agent i will send a friendship tie to agent j if the total surplus of doing so is positive:

$$D_{i,j} = \mathbf{1}(\varphi w_{ij} + \theta_i + \theta_j + U_{ij} \geq 0) \tag{4}$$

[18]Table A.3 presents balance tests where students' characteristics are regressed against the leave-out-mean characteristics of their peers in 6^{th} grade. I implement Guryan et al. (2009)'s correction to address the negative mechanical correlation created when performing such an exercise. The only relevant positive association is that girls seem more likely to study with other girls. However, as I control for students' gender in my main exercises, this should not be a concern.

where $\mathbf{1(.)}$ is an indicator function, $\theta_{i(j)}$ is agent $i(j)$'s fixed effect – to control for agents' unobserved degree heterogeneity–, and U_{ij} is an idiosyncratic component. Hence, if we assume that U_{ij} is a standard logistic random variable that is independently and identically distributed across dyads, the conditional likelihood of observing network $\boldsymbol{D} = \boldsymbol{d}$ is

$$Pr(\boldsymbol{D} = \boldsymbol{d}|\boldsymbol{w}, \boldsymbol{\theta}) = \prod_{i \neq j} Pr(D_{ij} = d|w_{ij}, \theta_i, \theta_j)$$

with

$$Pr(D_{ij=d}|\boldsymbol{w}, \boldsymbol{\theta}) = \left[\frac{1}{1 + exp(\varphi w_{ij} + \theta_i + \theta_j)}\right]^{1-d} \left[\frac{exp(\varphi w_{ij} + \theta_i + \theta_j)}{1 + exp(\varphi w_{ij} + \theta_i + \theta_j}\right]^{d}$$

for all $i \neq j$.

Such a probability is modeled using the following conditional logistic regression function:

$$Pr(D_{ij=d}|\boldsymbol{Z}, \boldsymbol{\theta}) = \frac{exp(\varphi w_{ij} + \theta_i + \theta_j)}{1 + exp(\varphi w_{ij} + \theta_i + \theta_j)} \tag{5}$$

Table 2 presents the results of such estimation. Column (1) presents the raw estimation, and column (2) presents the odds ratio. Analyzing the table, we can see that students' exogenous chances of interacting in the school due to allocations according to alphabetical order increase their likelihood of becoming friends.

Table 2: Probability of Forming a Friendship Link

	(1)	(2)
	Raw estimation	Odds-ratio
$1[x_i = x_j]$		
First name initial	0.407***	1.503***
	(0.053)	(0.080)
Constant	-3.970***	0.019***
	(0.035)	(0.001)
N (potential links)	524724	524724

Note: (i) This table shows the results of a conditional logistic regression model that predicts the likelihood that a student i will send a friendship tie to another student j in the ninth grade of the same school; the estimation controls for i's and j's fixed effects; (ii) Standard errors clustered at the school level shown in parenthesis; (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Using this model's predicted links, I replace the original adjacency matrix with the predicted adjacency matrix when building the instruments used to identify model 2. Therefore, in the final estimation of the parameters φ =(β, η, γ), the matrix of explanatory variables $\tilde{\boldsymbol{X}}$ =[$\boldsymbol{G y_l}$ $\boldsymbol{X_l}$ $\boldsymbol{G X_l}$] is instrumented in the second stage by $\hat{\boldsymbol{S}} = [\boldsymbol{X_l}\ \boldsymbol{G X_l}\ \hat{\boldsymbol{G}}(\boldsymbol{W})^{\boldsymbol{2}}\boldsymbol{X_l}$], where $\hat{\boldsymbol{G}}(\boldsymbol{W})$ is the predicted adjacency matrix from equation 8, $\hat{\boldsymbol{D}}(\boldsymbol{W})$, row normalized so that each row sums to one. The final estimates are, therefore, given by $\hat{\varphi}^{3SLS} = (\tilde{\boldsymbol{X}}'\hat{\boldsymbol{P}}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\hat{\boldsymbol{P}}\boldsymbol{y_l}$, where $\hat{\boldsymbol{P}} = \hat{\boldsymbol{S}}(\hat{\boldsymbol{S}}'\hat{\boldsymbol{S}})^{-1}\hat{\boldsymbol{S}}$.

Importantly, I only use the predicted links from my friendship formation model as instruments. As I am interested in the impacts of students' true (and not predicted) friends, the predicted friends of friends' characteristics instrument the true friends' outcomes, not the predicted ones.

3.4 Potential threats to identification and inference

This section discusses the validity of identifying assumptions in the implemented methodology and threats that might emerge from mapping students' networks. It also describes a procedure to adjust the standard errors of the model to deal with the fact that the network structure is predicted in the construction of the models' instrument variables.

Possible violation of the exclusion restriction

The exclusion restriction in the identification of model 2 is that the predicted friends of friends who are not the student's friends do not directly influence the student. In that sense, another potential threat to identification is that students' schooling decisions might be directly affected not only by their friends but also by other colleagues. A high-achieving colleague might be a role model, which could increase students' desire to invest more in their schooling, or a competitor, which could hinder such investments.[19] If this colleague is predicted to be a friend's friend, the exclusion restriction of the instruments might be threatened. Controlling for classroom fixed effects alleviates such a problem. Even though it does not address possible heterogeneous impacts that a specific colleague can have on a particular student, including classroom fixed effects in the model ensures that students' ranking and competitive dynamics within the classroom are held constant. Therefore, I also include classroom fixed effects in my most stringent estimations to check robustness.

Mapping of students' networks

An important assumption of Bramoullé et al. (2009) is that networks are fully mapped. That is, we should be able to identify all connections made by all individuals within a network. This assumption is necessary to guarantee that intransitive triads in the network are indeed intransitive. In order words, if A is connected to B, and B is connected to C, but

[19]Several contributions show that role models are crucial in determining individuals' aspirations (Beaman et al., 2012; Bernard et al., 2014; Macours and Vakis, 2014), while other papers in the education literature illustrate how competition or social image concerns might affect students' outcomes and behavior (Azmat and Iriberri, 2010; Bursztyn et al., 2019; Bursztyn and Jensen, 2015; Jonsson and Mood, 2008).

C is not connected to A, the absence of connection between A and C must not be due to missing or censored data. Such an assumption is also relevant for the model of friendship formation proposed by Graham (2017); all connections in a network must be identifiable to be fully modeled.

In that sense, the data used in this work may suffer from a ceiling effect since students could name only four of their friends. If a student had a fifth or sixth friend in that grade, these connections do not show up. Figure A.1 presents the out-degree distribution – that is, the distribution of the number of friends that each student named. While we can see that about 80% of students named at most three friends and hence were not censored in any way, the figure also shows that about 20 percent of students might be suffering from this ceiling effect since they named all four friends, and it is not possible to know if there were more friends they would like to name.[20] Although this proportion is considerable, the work of Griffith (2019) – who uses data from Add Health and other smaller survey to investigate the direction of the bias when censoring network data – shows that, if anything, censoring the number of friends yields a *downward* bias in the results. Still, in section 4.1, some robustness checks are offered to address potential issues with censored networks.

Specifications of the model of friendship formation

A final relevant discussion is in place. The main underlying event behind the predictive power of name similarities in friendship formation is students' alphabetical assignment in classes in the sixth grade. A concern here is that something special happens at the beginning of the middle school cycle that influences schooling decisions later on. For instance, early experiences in the new school, such as being assigned to a particularly motivating teacher or a particularly disruptive group of peers, could have long-term effects on students' outcomes. Another concern with using similarities in first-name initials is that such initials could be correlated with students' socioeconomic status. While there is evidence that first names are

[20] Figure A.1 also shows that about 20 percent of students did not named any friend. This proportion is in the same order as the one in Add-Health data (Niño et al., 2016). Exercises – not shown – either controlling for isolated students or excluding them from the estimation show very similar results.

correlated with individuals' socioeconomic background (e.g. Olivetti and Paserman, 2015), this should be less of a concern when looking at first name initials. Table A.4 in the appendix indeed shows that students' gender, race, parental education, and father's working status do not vary greatly depending on students' first name initials. In any case, it is important to check the robustness of the model for estimations that do not consider such variable in the friendship formation model. In appendix A.1, I check the sensitivity of estimations to different specifications of the model of link formations. One specification considers a model of friendship formation closer to the one proposed by Graham (2017), where the drivers of a link between two students are their unobserved degree heterogeneity and similarities in pre-determined characteristics, such as gender, race, and week of birth. Another specification considers both similarities in pre-determined characteristics and names similarities to predict friendship formation to check robustness.

Inference – standard errors adjustment

Finally, since this is a three-stage estimation where a network predicted in the first stage is used to build the instruments in the next two stages, the estimations' standard errors should be adjusted. I present bootstrapped standard errors with 50 replications in the most stringent estimations to account for this fact.[21]

[21] Since the bootstrap involves estimating the model of friendship formation and calculating $\hat{\boldsymbol{G}}(\boldsymbol{W})^2\boldsymbol{X_l}$ for each replication, such procedure is very computational demanding. Since, as shown, the bootstrapped standard errors tend to be larger than the non-bootstrapped ones, performing such a procedure only on the most stringent estimations is already a good indicator of whether the estimations remain significant.

4 Friends' influence on students' high school completion

Table 3 presents results of the main estimations of model 2.[22] Columns (1) and (2) present estimations for students' likelihood to finish high school, no matter when, and columns (3) and (4) present estimations for students' likelihood to finish high school without being retained in any grade during that school cycle. Columns (1) and (3) control for schools' fixed effects, and columns (2) and (4) control for classroom fixed effects.

The estimations indicate that peer effects on high school completion are positive, significant, and quite sizable. Let us consider the most stringent estimations that control for classroom fixed effects. Column 2 shows that if a student passes from having no named friends who finish high school to having all named friends who do, this student's probability of finishing high school increases by 10.4 percentage points. Column 4 shows that if we consider the likelihood of finishing high school with no retention, such an influence is even more prominent, with an impact of 18.5 percentage points. Importantly, the instruments are very strong, and we do not reject the hypothesis that they are indeed exogenous. First, the Kleibergen-Paap rk LM statistic is considerably high, lending to a rejection of the underidentification test in all estimations, which indicates the IVs' relevance. Second, such relevance is also supported by the IVs' join significance (F-statistic), which ranges from 165.72 to 380.46.[23] Finally, we also do not reject the Hansen test of over-identification restrictions, which indicates the exogeneity of the instruments.

[22]For comparative purposes, Table A.5, in the appendix, presents the results from an OLS estimation, the 2SLS estimation proposed by Bramoullé et al. (2009) and De Giorgi et al. (2010), and the 3SLS used throughout this work. The OLS estimation is actually smaller than the 2SLS estimation, which may be due to measurement error or to the exclusion bias discussed earlier. Importantly, however, the results decrease considerably when comparing the 2SLS estimation with the 3SLS one. This indicates that homophily might indeed bias the results upwards and shows the importance of properly correcting it.

[23]In a recent contribution about inference using instrumental variables, Lee et al. (2021) show that the commonly employed rule of thumb for the first-stage F statistic to be greater than 10 to guarantee strong instruments might be problematic. The authors demonstrate that both the value of the first-stage F statistic and the second-stage t statistic should be considered when evaluating the IVs' strength. In Table 3 of the paper, the authors present the critical values for t at the 5 percent significance level associated with a given F statistic. In all estimations of Table 3, the t-statistics are greater than the critical t, given the F-statistics associated with them. Hence, the instruments employed here are strong even under this more stringent condition.

Table 3: Friends' influence on schooling decisions

	HS completion		HS comp. w/out retention	
	(1)	(2)	(3)	(4)
Endogenous social effect	0.113***	0.104***	0.181***	0.185***
	(0.034)	(0.036)	(0.049)	(0.052)
		[0.055]		[0.099]
Own characteristics				
Girl	0.053***	0.055***	0.110***	0.110***
	(0.014)	(0.014)	(0.017)	(0.017)
White	0.006	0.006	0.002	0.004
	(0.011)	(0.011)	(0.013)	(0.014)
Mother education: more than HS	0.041***	0.039***	0.057***	0.058***
	(0.011)	(0.011)	(0.014)	(0.014)
Father education: more than HS	0.010	0.010	0.014	0.017
	(0.013)	(0.014)	(0.015)	(0.015)
Reading proficiency (2009)	0.044***	0.042***	0.061***	0.059***
	(0.006)	(0.007)	(0.007)	(0.007)
Math proficiency (2009)	0.023***	0.023***	0.038***	0.037***
	(0.006)	(0.007)	(0.008)	(0.008)
Father works	0.031**	0.032**	0.050***	0.048***
	(0.013)	(0.013)	(0.013)	(0.014)
Friends' characteristics				
Girl	0.003	0.012	-0.022	-0.018
	(0.019)	(0.019)	(0.024)	(0.025)
White	0.018	0.017	0.038**	0.042**
	(0.017)	(0.018)	(0.019)	(0.020)
Mother education: more than HS	-0.040**	-0.036*	-0.045**	-0.045*
	(0.019)	(0.021)	(0.022)	(0.023)
Father education: more than HS	0.028	0.033	0.031	0.042*
	(0.020)	(0.021)	(0.023)	(0.025)
Reading proficiency (2009)	0.013	0.008	0.016	0.013
	(0.011)	(0.011)	(0.012)	(0.012)
Math proficiency (2009)	0.005	0.009	-0.006	-0.003
	(0.010)	(0.011)	(0.012)	(0.013)
Father works	-0.005	-0.006	-0.020	-0.028
	(0.019)	(0.020)	(0.023)	(0.024)
N	6075	6075	6075	6075
Mean Dep. Var.	0.785	0.785	0.637	0.637
R2	0.055	0.047	0.089	0.076
Kleibergen-Paap rk LM statistic	211.000	198.885	175.718	165.322
P-val underidentification test	0.000	0.000	0.000	0.000
IVs' joint significance	379.623	380.464	165.714	165.995
Hansen J statistic	5.887	3.159	10.230	8.117
P-val overidentification test	0.436	0.789	0.115	0.230
Control for school FE	✓		✓	
Control for classroom FE		✓		✓

Note: (i) This table shows estimations of models like the one described in equation 2, in which friends' high school completion is instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) "Endogenous social effect" is the effect that friends' outcomes have on students' outcomes. In columns (1) and (2), this is the average of friends who finished HS, while in columns (3) and (4), this is the average of friends who finished HS without retention; (iii) Standard errors clustered at the classroom level are shown in parenthesis, bootstrapped standard errors clustered at the classroom level are shown in brackets; (iv) To avoid overcrowding, bootstrapped standard errors are shown only in the endogenous social effect coefficients; (v) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Perhaps passing from having no friends with high school completion to having all friends with high school completion is a too-extreme interpretation of the results. A better interpretation might be to consider the marginal impact of an extra friend with such a level of education. This effect will depend on the number of named friends. As shown in Table 1, students named on average two friends. Hence, we need to divide the estimated coefficient by two to retrieve the impact of an extra friend finishing high school for the average student. Regarding the likelihood of finishing high school at any time, such an impact would be 5.20 p.p., translating into a 6.62 percent increase in the likelihood that a student will finish high school. If we focus on finishing high school with no retention, such an impact would be 9.25 p.p., translating into a 14.52 percent increase in the likelihood that a student will finish high school without being retained.

To understand the importance of such an impact, we can compare it with other inputs entering the model of school completion shown in Table 3. For instance, the impact of an extra friend finishing high school (without retention) is 1.3 (1.6) times larger than the impact of having a mother with a high school diploma or more. This impact is 2.3 (2.5) times larger than moving students in one standard deviation in their math performance.

These results are also larger than those found in other studies examining peer effects on high school completion. Abramitzky et al. (2021), for instance, shows that a reform that increased the returns to schooling for kibbutzim students in Israel led to spillovers on their non-kibbutz schoolmates. After the reform, the exposition to an additional kibbutz schoolmate increased by 1.2 percentage points the likelihood that a non-kibbutz student would finish high school.[24] Another example is Ballis (2020), which exploits spillover from the Deferred Action for Childhood Arrivals (DACA) that significantly increased the returns to schooling for undocumented youth in the United States. She finds that the exposition to an additional DACA-eligible schoolmate increases by 4.5 percentage points the likelihood

[24] The impact of an additional kibbutz schoolmate was calculated based on Table A2, showing the sample size, and Table 3, showing the effect of the share of post-reform kibbutz students on high school completion.

that a US-born student will finish high school.[25] The fact that the results here are larger than the ones looking at general peer effects is expected since friends are socially closer to students than other schoolmates and, as such, might have a stronger influence on them.

Table 4 shows that friends' influence is stronger for low-SES students. The columns of the table show heterogeneous results regarding gender (columns (1) and (2)), race (columns (3) and (4)), and maternal education (columns (5) and (6)). Panel A shows results for finishing high school, while Panel B shows results for finishing high school without retention. We can see in Panel B that, while friends have roughly the same influence on boys and girls, they are more influential for black students and students whose mothers did not finish high school. Such heterogeneity might emerge because low-SES students receive less support from their parents and possibly even from their teachers,[26] which makes the influence of their friends more critical for their outcomes in school.

Table 4: Heterogeneous impacts

	HS completion			HS completion w/out retention		
	(1) Girl	(2) Black	(3) Mother less HS	(4) Girl	(5) Black	(6) Mother less HS
Endogenous social effects	0.123***	0.110***	0.098**	0.190***	0.153***	0.090*
	(0.026)	(0.022)	(0.039)	(0.033)	(0.026)	(0.048)
Endog. social effects x Var. in column	-0.030	0.004	0.017	-0.061	0.066*	0.094*
	(0.037)	(0.024)	(0.043)	(0.046)	(0.034)	(0.052)
N	6075	6075	6075	6075	6075	6075
Mean Dep. Var.	0.785	0.785	0.785	0.637	0.637	0.637
R2	0.046	0.045	0.045	0.077	0.078	0.076
Kleibergen-Paap rk LM statistic	200.232	227.273	224.771	195.508	218.519	205.056
P-val underidentification test	0.000	0.000	0.000	0.000	0.000	0.000
IVs' joint significance	473.401	524.093	523.594	268.030	277.254	272.569
Hansen J statistic	4.462	7.055	6.046	11.284	14.696	11.705
P-val overidentification test	0.974	0.854	0.914	0.505	0.258	0.470
Control for classroom FE	✓	✓	✓	✓	✓	✓

Note: (i) This table shows derivations of the model described in equation 2, in which the endogenous social effect and an interaction of the endogenous social effect with the variables described in the table's columns are instrumented by $\hat{G}^2X$, as well as interactions between these instruments and the variables described in the table's columns; (ii) "Endogenous social effect" is the effect that friends' outcomes have on students' outcomes. In columns (1) to (3), this is the impact of the average of friends who finished HS, while in columns (4) to (6), this is the impact of the average of friends who finished HS without retention; (iii) All estimations include the same controls as in Table 3; (iv) Standard errors clustered at the classroom level are shown in parenthesis, bootstrapped standard errors clustered at the classroom level are shown in brackets; (v) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

[25] The impact of being exposed to one additional schoolmate was calculated based on the results from Panel D of Table 5 and considering the average cohort size and percentage of DACA-eligible students, shown in Table 2.

[26] The study by Papageorge et al. (2016) shows how teachers expect less from black students, which turns out to be a self-fulfilling prophecy regarding their future outcomes.

4.1 Robustness checks

Tables A.6, A.7, and A.8, in the appendix, present several exercises that address possible concerns with the identification strategy proposed by this paper. First, I include contextual effects in my model of friends' influence – namely, the impacts of friends' characteristics, or $\boldsymbol{\eta}$ in equation 2. Given the endogenous formation of links and the fact that only $\boldsymbol{Gy_l}$ is instrumented by $\hat{G}^2X$, the inclusion of $\boldsymbol{Gy_X}$ in the model might raise concerns of endogeneity. Hence, columns (1) and (3) of table A.6 present estimations without the inclusion of friends' characteristics. The results are pretty stable for such a change in the estimates.

Second, as discussed in section 3.4, a potential threat for identification is the fact that some students in the data did not name all of their friends due to space restrictions. If this is the case, the model of friendship formation might not be correctly estimated, and some excluded instruments used in the estimation of peer effects might be endogenous. Even though the contribution by Griffith (2019) has shown that, if anything, such a problem biases the estimations upwards, I check whether possible ceiling effects in the naming of friends could be driving the results. Columns (2) and (4) of table A.6 present an estimation of the results in the sub-sample of students who were not censored by the limit in friendship nomination – that is, students who named three friends or fewer. In this restricted sample, it is possible to map all students' connections with more precision without incurring the risk of missing links. Again, the results are remarkably similar to the ones in Table 3.

Third, table A.7 presents some placebo exercises where I check the endogenous social effects on some students' pre-determined characteristics, such as parental education and home ownership. If endogenous social effects emerged as significant from such tests, this would indicate that my model is, at least to some extent, still capturing the homophily in friendship formation instead of causal friends' influence. As we can see, this is not the case: not only none of the coefficients in table A.7 is significant, but the size of the coefficient is also very small.

Finally, table A.8 presents estimations of friends' influence using different models of friendship formation, as explained in appendix A.1 (see section 3.4 for a discussion about the importance of such checks). Columns (1) and (3) present estimations in which the predicted adjacency matrix $\hat{G'}^2X$ comes from a model that considers students' similarities in terms of gender, race, and date of birth, and do not consider name similarities among students to predict friendship links. Columns (2) and (4), in turn, present estimations in which the predicted adjacency matrix $\hat{G''}^2X$ comes from a model that considers both name similarities and similarities in these other pre-determined characteristics among students to predict friendship links. In both cases, the point estimates are remarkably similar to the ones in table 3, and the instruments are as strong as in the main estimations.

5 Mechanisms behind friends' influence

This section discusses different channels through which friends' high school completion influences students' high school completion. I leverage some variables from the survey described in section 2 that might enter as inputs in a model of human capital accumulation decisions. The first is college aspirations, a powerful predictor of individuals' investments, including in education (Fruttero et al., 2021). The second is students' perception of high school returns. Several contributions have shown that information on schooling returns and perceptions about the consumption value of education increase students' educational investments (or, at least, investments preferences) (Belfield et al., 2020; Bleemer and Zafar, 2018; Boneva and Rauh, 2017; Jensen, 2010). The third variable is students' fear of nerd stigma and how they see such fear as an impediment to putting effort into school. This variable can proxy for students' willingness to comply with rather harmful social norms in the school, ultimately impacting their schooling decisions (Bursztyn and Jensen, 2017). The last variables are school effort – proxied by an indicator of outside school hours math study of at least 30 minutes –and reading and math proficiency.

I perform two exercises focusing on these variables. First, I investigate whether these characteristics spread through students' networks in the same way as schooling decisions. The second exercise is a mediation analysis where I test whether and how much each of these characteristics captures the main results of friends' influence on students' schooling decisions.

5.1 Friends' influence on potential inputs of schooling decision

The impact of friends' high school completion on students' high school completion may come through friends' influence on different inputs that enter a model of human capital accumulation. I investigate such a channel employing the same methodology as the one described in section 3, but this time looking at how friends' characteristics related to schooling investments influence students' such characteristics. Table 5 presents the results of this exercise. Column (1) shows estimations of how friends' college aspirations influence students' college aspirations; column (2) shows estimations of how friends' perceived high school returns influence students' perceived high school returns; column (3) shows estimations of how friends' fear of nerd stigma influences students' such fear; column (4) shows estimations of how friends' math study time influences students' study time; and, columns (5) and (6) show estimations of how friends' performance in reading and math, respectively, impacts students performance on these subjects.

The results in column (1) show that friends' college aspirations have a crucial impact on students' own college aspirations. For the average student with two friends, an extra friend aspiring to a college degree increases by 8.15 percentage points the likelihood that the student will also aspire to it. In column (2), we see that students do not seem to share information about high school returns, at least not about labor market returns from finishing high school.[27] The results in column (3) also show absent endogenous social effects on students' fear of being stigmatized as nerds. Column (4), in turn, shows that friends' school effort, proxied by their math study time, influences students' school effort. Specifically, for the av-

[27] I am unfortunately unable to test whether information about non-pecuniary returns is spread through the network.

Table 5: Friends' influence on inputs of schooling decisions

	(1) College aspiration	(2) Perceived HS returns	(3) Fear of nerd stigma	(4) 30+ min math study/day	(5) Reading proficiency	(6) Math proficiency
Endog. social effects	0.163***	-0.034	0.034	0.170**	-0.526	0.747
	(0.053)	(0.036)	(0.098)	(0.077)	(0.387)	(0.537)
	[0.086]	[0.059]	[0.160]	[0.119]	[0.624]	[0.638]
N	6075	6075	6075	6075	5833	5833
Mean Dep. Var.	0.684	0.500	0.256	0.396	-0.000	0.000
R2	0.067	0.042	0.066	-0.004	0.377	0.164
Kleibergen-Paap rk LM statistic	176.175	198.567	94.758	114.644	16.986	10.222
P-val underidentification test	0.000	0.000	0.000	0.000	0.017	0.176
IVs' joint significance	185.743	343.749	47.358	62.623	3.919	1.925
Hansen J statistic	7.138	3.647	10.509	3.277	6.352	10.505
P-val overidentification test	0.308	0.724	0.105	0.773	0.385	0.105
Control for classroom FE	✓	✓	✓	✓	✓	✓

Note: (i) This table shows estimations of models like the one described in equation 2, in which friends' outcomes are instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) "Endogenous social effect" is the effect that friends' average outcomes have on students' outcomes. For instance, in column (1) this is the impact of the average of friends' college aspirations, in column (2), this is the impact of the average of friends' perceptions about HS returns, etc.; (iii) All estimations include the same controls as in Table 3; (iv) Standard errors clustered at the classroom level are shown in parenthesis, bootstrapped standard errors clustered at the classroom level are shown in brackets; (v) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

erage student, an extra friend studying at least half an hour per day increases the likelihood that this student will also study such amount of time by 8.5 percentage points. Interestingly, the results in columns (5) and (6) show absent peer effects on students' performance.

Since aspirations and school efforts spread through students' networks, they are good candidates for the channels behind the influence of friends' schooling decisions on students' schooling decisions. Information diffusion and compliance with social norms, in turn, are unlikely to be the mechanisms in place. I next perform a mediation analysis to test how much friends' influence on students' schooling decisions can be explained by these potential channels.

5.2 Mediation analysis

I start by estimating an extended version of model 2:

$$\begin{aligned} \boldsymbol{y_l} = \beta \boldsymbol{G y_l} + \lambda \boldsymbol{A_l} + \theta \boldsymbol{B_l} + \mu_l + \boldsymbol{\upsilon_l} \\ E(\upsilon_l | \boldsymbol{A_l}, \boldsymbol{B_l}, \mu_l) = 0 \end{aligned} \tag{6}$$

where $\boldsymbol{A_l} = [\boldsymbol{X_l}, \boldsymbol{GX_l}]$ is the vector of controls included in model 2, and $\boldsymbol{B_l} = [\boldsymbol{X_l'}, \boldsymbol{GX_l'}]$ is a vector containing the new variables discussed above, namely students' and friends' aspirations, perceived high school returns, fear of nerd stigma, study time, and reading and math proficiency. Remember that all these variables were measured when students were in 9^{th} grade – the last grade of primary school in Brazil –, so they are all pre-determined with respect to high school conclusion.

Table 6 presents the results of such estimation. Columns (1) and (3) present benchmark estimations, that is, estimations of model 2 without the inclusion of further controls, as in Table 3. Columns (2) and (4), in turn, present estimations considering the new controls. Two important results stem from this table. First, the aspirations of both students and their friends are important predictors of students' high school graduation. This finding adds to previous evidence corroborating the theoretical literature discussing the important role of aspirations in determining individuals' investments and efforts.[28] Second, after including these other inputs, the influence that friends' high school conclusion has on students' high school conclusion vanishes: The estimated coefficient approaches zero and loses its significance. Hence, at least one of the new controls is an important mechanism behind friends' influence on high school graduation.

I employ Gelbach (2016)'s decomposition to understand the relevance of each of these new controls for the decrease in the coefficient of my main estimation.

Call β^{base} the coefficient of model 2, that omits $\boldsymbol{B_l}$ and β^{full} the coefficient of model 6 that includes such variables. As shown in Gelbach (2016), the total mediated effects after the inclusion of $\boldsymbol{B_l}$ in the model is the difference between $\widehat{\beta}^{base}$ and $\widehat{\beta}^{full}$, and is given by $\widehat{\delta} = \sum_k \widehat{\delta}_k$, where $\widehat{\delta}_k$ is the omitted variable bias for each variable k in vector $\boldsymbol{B_l}$. Hence, the relevance of each variable $\boldsymbol{B_{l}}_k$ for the total mediated effects is given by $\widehat{\delta}_k = \widehat{\Gamma_k}\widehat{\theta}_k$, where $\widehat{\Gamma_k}$ are estimates of a regression of $\boldsymbol{B_{l}}_k$ on $\boldsymbol{A_l}$, $\boldsymbol{Gy_l}$.Gelbach (2016) also shows that the same strategy can be employed to mediate differences in IV estimates of β due to addition of exogenous covariates, which is the case here.[29]

Figures 1 and 2 present the mediation analysis for friends' influence on high school completion and high school completion without retention, respectively. Table A.9 in the appendix also presents such analysis. The total mediated effect is the difference between $\widehat{\beta}^{base}$ and $\widehat{\beta}^{full}$, that is, how much of friends' influence is reduced once the new controls are included in the model. Each coefficient below the total mediated effect represents how much that control contributes to such a difference. We can see that while students' own aspirations and performance explain some of the decrease in the endogenous social effects, most of the decrease is due to friends' aspirations.

Overall, the above exercises show that aspirations play a crucial role in friends' influence

[28]See Dalton et al. (2016); Genicot and Ray (2017) for reference on the theoretical literature, and Fruttero et al. (2021) for a review on empirical studies showing the associations between aspirations and future outcomes.

[29]See Gelbach (2016) for more details.

Table 6: Friends' influence on schooling decisions – further controls

	HS completion		HS comp. w/out retention	
	(1)	(2)	(3)	(4)
Endog. social effects	0.104***	0.033	0.185***	-0.016
	(0.036)	(0.053)	(0.052)	(0.090)
College aspiration		0.056***		0.080***
		(0.013)		(0.015)
Perceived HS returns		0.016		0.014
		(0.021)		(0.023)
Fear of nerd stigma		-0.020		-0.007
		(0.013)		(0.015)
30+ min math study/day		0.012		0.009
		(0.011)		(0.012)
Reading proficiency (2011)		0.039***		0.050***
		(0.008)		(0.009)
Math proficiency (2011)		0.025***		0.038***
		(0.007)		(0.007)
College aspiration		0.057**		0.098***
		(0.023)		(0.029)
Perceived HS returns		-0.006		0.048
		(0.034)		(0.043)
Fear of nerd stigma		-0.003		0.016
		(0.021)		(0.023)
Reading proficiency (2011)		0.006		0.018
		(0.012)		(0.014)
Math proficiency (2011)		-0.001		-0.003
		(0.011)		(0.013)
30+ in study/day		0.022		0.017
		(0.019)		(0.022)
N	6075	6075	6075	6075
Mean Dep. Var.	0.785	0.785	0.637	0.637
R2	0.047	0.068	0.076	0.109
Kleibergen-Paap rk LM statistic	198.885	149.016	165.322	107.739
P-val underidentification test	0.000	0.000	0.000	0.000
IVs' joint significance	380.464	166.535	165.995	56.744
Hansen J statistic	3.159	3.083	8.117	6.771
P-val overidentification test	0.789	0.798	0.230	0.343
Control for classroom FE	✓	✓	✓	✓

Note: (i) This table shows estimations of models like the one described in equation 2, in which friends' high school completion is instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) "Endogenous social effect" is the effect that friends' outcomes have on students' outcomes. In columns (1) and (2), this is the average of friends who finished HS, while in columns (3) and (4), this is the average of friends who finished HS without retention; (iii) Standard errors clustered at the classroom level are shown in parenthesis; (iv) All estimations include the same controls as in Table 3; (v) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

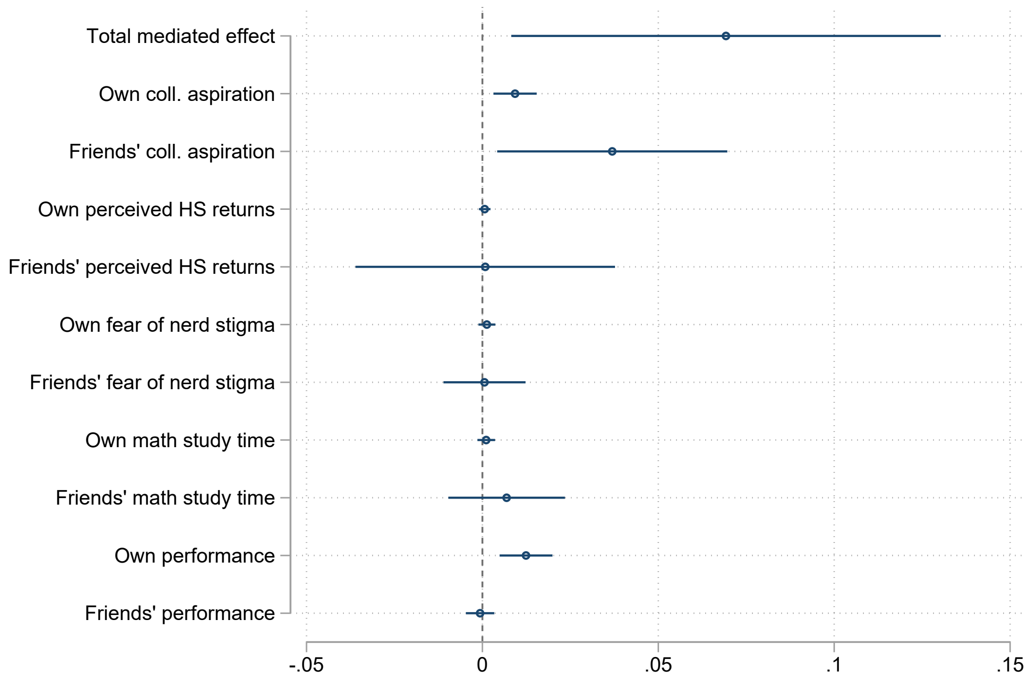

Figure 1: Mediation analysis – HS Completion

Note: The total mediated effect represents how much friends' influence is reduced once the model includes the new controls. The coefficients below the total mediated effect represent the contribution of each control in explaining the variation of friends' influence. Detailed results shown in Table A.9.

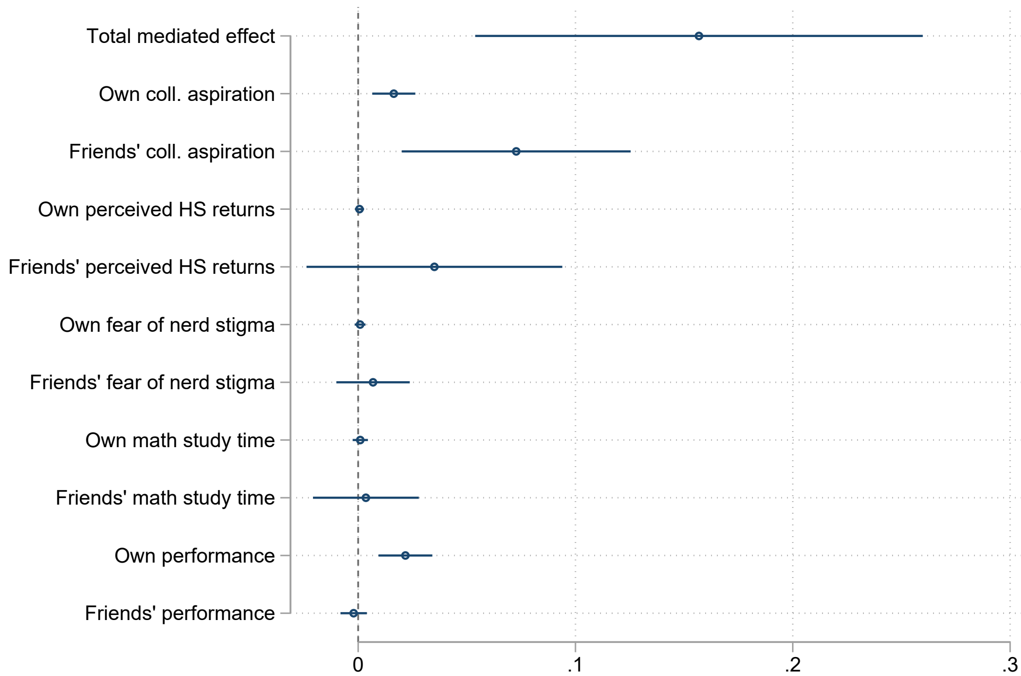

Figure 2: Mediation analysis – HS Completion w/out retention

Note: The total mediated effect represents how much friends' influence is reduced once the model includes the new controls. The coefficients below the total mediated effect represent the contribution of each control in explaining the variation of friends' influence. Detailed results shown in Table A.9.

on students' schooling decisions. Not only do college aspirations spread through students' networks, but they also capture the impact that friends' schooling decisions have on students' schooling decisions. Importantly, however, friends' aspirations play a crucial role in students' schooling decisions, above and beyond influencing students' own aspirations in the 9^{th} grade. There are two possible reasons for such a finding. First, the process of aspirations formation is probably dynamic, such that students might keep influencing each other's aspirations during high school. Second, students with higher aspirations might create a better school environment with higher engagement and lower disruptions, decreasing students' dropout likelihood. While separating these two channels is beyond the scope of this paper, uncovering the role of aspirations behind friends' influence on schooling decisions have important policy implications, discussed in the next section.

6 Conclusion

This work addresses primary challenges concerning the estimation of peer effects and investigates the influence of friends' schooling decisions on students' schooling decisions and the mechanisms behind such an impact. I first leverage students' exogenous opportunities of interaction to model friendship formation. Then, based on the predicted friendship links coming from the model, I use predicted friends of friends' characteristics as instrumental variables for friends' outcomes. This identification strategy overcomes both the endogenous formation of friendships and the reflection problem, largely discussed in the literature on peer effects estimation.

Results show that an extra friend finishing high school increases students' likelihood of finishing high school. Such an impact is concentrated among students from low-income backgrounds, possibly the ones who need to rely more heavily on the support of their friends. An in-depth investigation of the mechanisms behind friends' influence delivers some relevant patterns. First, while I do not find evidence of information diffusion or peer effects on students' fear of nerd stigma or performance, friends influence students' aspirations and their effort in the school through their own aspirations and effort, respectively. Second, the aspirations of students and their friends are crucial predictors of students' likelihood of finishing high school and are the main variables mediating the influence of friends' schooling decisions on students' schooling decisions. Friends' aspirations, in particular, play an essential role in mediating such an impact.

These findings offer a relevant policy message. While providing students information about schooling returns and raising students' aspirations might be effective in increasing their schooling (e.g Carlana et al., 2022; Jensen, 2010), the present paper finds that aspirations are more likely to spill over to other students. I also show here that aspirations are the main channel through which friends influence students' high school completion. Hence, if a policymaker wants the impact of a policy aiming at improving human capital accumulation to reach other students, raising their aspirations might be a better option. However, a note of caution is in place here: this paper is not able to address the question of whether *new* information would spread through students' networks and lead to higher peers' influence on schooling decisions. The fact that students in the networks analyzed here are not spreading their perceptions about schooling returns does not mean that they would not spread it if someone changed their priors regarding such returns. Banerjee et al. (2019) find, for instance, that giving information to central individuals in a network might be an effective policy tool due to high information diffusion.

Regardless of the possible effects of offering more information to students, this paper evidences the crucial role that aspirations play in human capital accumulation decisions. Educational policies that raise students' aspirations will likely spill over to these students' friends. As long as such policies do not raise students' aspirations so much that they fall into aspirations failures (Genicot and Ray, 2017; Goux et al., 2017; Kearney and Levine, 2014), they might help both students and their friends to go further in school.

References

Abramitzky, R., Lavy, V., and Pérez, S. (2021). The long-term spillover effects of changes in the return to schooling. *Journal of Public Economics*, 196:104369.

Aguirre, J. and Matta, J. (2021). Walking in your footsteps: Sibling spillovers in higher education choices. *Economics of Education Review*, 80:102062.

Alan, S., Duysak, E., Kubilay, E., and Mumcu, I. (2020). Social exclusion and ethnic segregation in schools: The role of teacher's ethnic prejudice. *The Review of Economics and Statistics*, pages 1–45.

Altmejd, A., Barrios-Fernández, A., Drlje, M., Goodman, J., Hurwitz, M., Kovac, D., Mulhern, C., Neilson, C., and Smith, J. (2021). O brother, where start thou? sibling spillovers on college and major choice in four countries. *The Quarterly Journal of Economics*, 136(3):1831–1886.

Anelli, M. and Peri, G. (2019). The effects of high school peers' gender on college major, college performance and income. *The Economic Journal*, 129(618):553–602.

Anelli, M., Shih, K., and Williams, K. (2022). Foreign students in college and the supply of stem graduates. *Journal of Labor Economics*.

Appadurai, A. (2004). The capacity to aspire: Culture and the terms of recognition'in vijayendra rao and michael walton (eds), culture and public action.

Azmat, G. and Iriberri, N. (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics*, 94(7-8):435–452.

Bailey, M. J. and Dynarski, S. M. (2011). Gains and gaps: Changing inequality in us college entry and completion. Technical report, National Bureau of Economic Research.

Ballis, B. (2020). Does peer motivation impact educational investments? evidence from daca.

Bandiera, O., Barankay, I., and Rasul, I. (2010). Social incentives in the workplace. *The review of economic studies*, 77(2):417–458.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies*, 86(6):2453–2490.

Barrios-Fernández, A. (2022). Neighbors' effects on university enrollment. *American Economic Journal: Applied Economics*.

Beaman, L., Duflo, E., Pande, R., and Topalova, P. (2012). Female leadership raises aspirations and educational attainment for girls: A policy experiment in india. *science*, 335(6068):582–586.

Belfield, C., Boneva, T., Rauh, C., and Shaw, J. (2020). What drives enrolment gaps in further education? the role of beliefs in sequential schooling decisions. *Economica*, 87(346):490–529.

Bernard, T., Dercon, S., Orkin, K., Taffesse, A., et al. (2014). *The future in mind: Aspirations and forward-looking behaviour in rural Ethiopia*. Centre for Economic Policy Research London.

Bleemer, Z. and Zafar, B. (2018). Intended college attendance: Evidence from an experiment on college returns and costs. *Journal of Public Economics*, 157:184–211.

Bobonis, G. J. and Finan, F. (2009). Neighborhood peer effects in secondary school enrollment decisions. *The Review of Economics and Statistics*, 91(4):695–716.

Boneva, T. and Rauh, C. (2017). Socio-economic gaps in university enrollment: The role of perceived pecuniary and non-pecuniary returns. *Available at SSRN 3106691*.

Boucher, V. and Bramoullé, Y. (2021). Binary outcomes and linear interactions.

Boucher, V., Bramoullé, Y., Djebbari, H., and Fortin, B. (2014). Do peers affect student achievement? evidence from canada using group size variation. *Journal of Applied Econometrics*, 29(1):91–109.

Boucher, V., Dedewanou, F. A., and Dufays, A. (2018). Peer-induced beliefs regarding college participation.

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

Bramoullé, Y., Djebbari, H., and Fortin, B. (2020). Peer effects in networks: A survey. *Annual Review of Economics*, 12:603–629.

Brenøe, A. A. and Zölitz, U. (2020). Exposure to more female peers widens the gender gap in stem participation. *Journal of Labor Economics*, 38(4):1009–1054.

Brock, W. A. and Durlauf, S. N. (2001). Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260.

Bursztyn, L., Egorov, G., and Jensen, R. (2019). Cool to be smart or smart to be cool? understanding peer pressure in education. *The Review of Economic Studies*, 86(4):1487–1526.

Bursztyn, L. and Jensen, R. (2015). How does peer pressure affect educational investments? *The quarterly journal of economics*, page qjv021.

Bursztyn, L. and Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, 9:131–153.

Caeyers, B. and Fafchamps, M. (2016). Exclusion bias in the estimation of peer effects. Technical report, National Bureau of Economic Research.

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

Carlana, M., La Ferrara, E., and Pinotti, P. (2022). Goals and gaps: Educational careers of immigrant children. *Econometrica*, 90(1):1–29.

Cipollone, P. and Rosolia, A. (2007). Social interactions in high school: Lessons from an earthquake. *American Economic Review*, 97(3):948–965.

Cools, A., Fernández, R., and Patacchini, E. (2021). The asymmetric gender effects of high flyers. Technical report, mimeo.

Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.

Dahl, G. B., Rooth, D.-O., and Stenberg, A. (2020). Family spillovers in field of study. Technical report, National Bureau of Economic Research.

Dalton, P. S., Ghosal, S., and Mani, A. (2016). Poverty and aspirations failure. *The Economic Journal*, 126(590):165–188.

De Giorgi, G., Pellizzari, M., and Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2(2):241–75.

Dickerson, A., Maragkou, K., and McIntosh, S. (2018). The causal effect of secondary school peers on educational aspirations. *CVER Discussion Paper Series-ISSN 2398-7553*.

Dustan, A. (2018). Family networks and school choice. *Journal of Development Economics*, 134:372–391.

Dynarski, S., Libassi, C. J., Michelmore, K., and Owen, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *The American Economic Review*, 111:1721–1756.

Feld, J. and Zölitz, U. (2022). The effect of higher-achieving peers on major choices and labor market outcomes. *Journal of Economic Behavior & Organization*, 196:200–219.

Fletcher, J. M., Ross, S. L., and Zhang, Y. (2020). The consequences of friendships: Evidence on the effect of social relationships in school on academic achievement. *Journal of Urban Economics*, 116:103241.

Fruttero, A., Muller, N., and Calvo-Gonzalez, O. (2021). The power and roots of aspirations.

Galiani, S., Gertler, P. J., and Undurraga, R. (2021). Aspiration adaptation in resource-constrained environments. *Journal of Urban Economics*, 123:103326.

Galor, O. (2011). Inequality, human capital formation, and the process of development. In *Handbook of the Economics of Education*, volume 4, pages 441–493. Elsevier.

Galor, O. and Moav, O. (2004). From physical to human capital accumulation: Inequality and the process of development. *The Review of Economic Studies*, 71(4):1001–1026.

Gelbach, J. B. (2016). When do covariates matter? and which ones, and how much? *Journal* *of Labor Economics*, 34(2):509–543.

Genicot, G. and Ray, D. (2017). Aspirations and inequality. *Econometrica*, 85(2):489–519.

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

Goux, D., Gurgand, M., and Maurin, E. (2017). Adjusting your dreams? high school plans and dropout behaviour. *The Economic Journal*, 127(602):1025–1046.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.

Griffith, A. (2016). Random assignment with non-random peers: A structural approach to counterfactual treatment assessment. *The Review of Economics and Statistics*, pages 1–40.

Griffith, A. (2019). Name your friends, but only five? the importance of censoring in peer effects estimates using social network data.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 1(4):34–68.

Hoxby, C. M. and Avery, C. (2012). The missing" one-offs": The hidden supply of high-achieving, low income students. Technical report, National Bureau of Economic Research.

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

Janzen, S. A., Magnan, N., Sharma, S., and Thompson, W. M. (2017). Aspirations failure and formation in rural nepal. *Journal of Economic Behavior & Organization*, 139:1–25.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2):515–548.

Joensen, J. S. and Nielsen, H. S. (2018). Spillovers in education choice. *Journal of Public Economics*, 157:158–183.

Johnsson, I. and Moon, H. R. (2021). Estimation of peer effects in endogenous social networks: Control function approach. *Review of Economics and Statistics*, 103(2):328–345.

Jonsson, J. O. and Mood, C. (2008). Choice by contrast in swedish schools: How peers' achievement affects educational choice. *Social forces*, 87(2):741–765.

Kearney, M. S. and Levine, P. B. (2014). Income inequality, social mobility, and the decision to drop out of high school. Technical report, National Bureau of Economic Research.

König, M. D., Liu, X., and Zenou, Y. (2018). R&d networks: Theory, empirics and policy implications. *Review of Economics and Statistics*.

La Ferrara, E. (2019). Aspirations, social norms, and development. *Journal of the European Economic Association*, 17(6):1687–1722.

Lee, D. S., McCrary, J., Moreira, M. J., and Porter, J. R. (2021). Valid t-ratio inference for iv. Technical report, National Bureau of Economic Research.

Liu, X., Patacchini, E., and Zenou, Y. (2014). Endogenous peer effects: local aggregate or local average? *Journal of Economic Behavior & Organization*, 103:39–59.

Macours, K. and Vakis, R. (2014). Changing households' investment behaviour through social interactions with local leaders: Evidence from a randomised transfer programme. *The Economic Journal*, 124(576):607–633.

Mani, A. and Riley, E. (2019). *Social networks, role models, peer effects, and aspirations*. Number 2019/120. WIDER Working Paper.

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

Mayer, A. and Puller, S. L. (2008). The old boy (and girl) network: Social network formation on university campuses. *Journal of public economics*, 92(1-2):329–347.

Moody, J. (2001). Race, school integration, and friendship segregation in america. *American journal of Sociology*, 107(3):679–716.

Mora, T. and Oreopoulos, P. (2011). Peer effects on high school aspirations: Evidence from a sample of close and not-so-close friends. *Economics of Education Review*, 30(4):575–

581.

Niño, M. D., Cai, T., and Ignatow, G. (2016). Social isolation, drunkenness, and cigarette use among adolescents. *Addictive behaviors*, 53:94–100.

Norris, J. (2020). Peers, parents, and attitudes about school. *Journal of Human Capital*, 14(2):290–342.

OECD (2019). *Education at a Glance 2019: OECD Indicators*. OECD Publishing, Paris.

Olivetti, C. and Paserman, M. D. (2015). In the name of the son (and the daughter): Intergenerational mobility in the united states, 1850-1940. *American Economic Review*, 105(8):2695–2724.

Pagani, L. and Pica, G. (2021). A peer like me? early exposure to high achievers in math and later educational outcomes. *Early Exposure to High Achievers in Math and Later Educational Outcomes (July 2021)*.

Papageorge, N. W., Gershenson, S., and Kang, K. M. (2016). Teacher expectations matter. *Review of Economics and Statistics*, pages 1–46.

Qu, X. and Lee, L.-f. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(2):209–232.

Qureshi, J. A. (2018). Additional returns to investing in girls' education: Impact on younger sibling human capital. *The Economic Journal*, 128(616):3285–3319.

Ray, D. (2006). Aspirations, poverty, and economic change. *Understanding poverty*, 409421.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics*, 116(2):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 *Handbook of the Economics of Education*, volume 3, pages 249–277. Elsevier.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: two steps forward? *Annu. Rev. Econ.*, 6(1):253–272.

Santavirta, T. and Sarzosa, M. (2019). Effects of disruptive peers in endogeneous social networks. *Working paper*.

Stutzer, A. (2004). The role of income aspirations in individual happiness. *Journal of Economic Behavior & Organization*, 54(1):89–109.

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

Weinberg, B. A. (2007). Social interactions with endogenous associations. Technical report, National Bureau of Economic Research.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics*, 85(1):9–23.

Zölitz, U. and Feld, J. (2021). The effect of peer gender on major choice in business school. *Management Science*, 67(11):6963–6979.

A Online Appendix

A.1 Model of friendship formation based on homophily in pre-determined characteristics

The main model of friendship formation implemented in this paper uses similarities in name initials as students' pre-determined characteristics influencing their link formation. As said before, the mechanism behind friendship formation in this case is not homophily, but the exogenous increase in students' likelihood to meet and interact in the school. In this section, I present other specifications of this model, closer to the contribution by Graham (2017), who explicitly models network formation based on homophily. The main idea of this model is that the friendship connection $D_{i,j}$ between two agents i and j, depends on the distance between these two agents regarding several agent-level attributes $Z_i = \{z_{1i}, ... z_{Ki}\}$, and on their unobserved degree heterogeneity. If we consider $W_{ij} = \sum_{k=1}^{K}(|z_{ki} - z_{kj}|)$ as a measure of the total distance between i and j, then agent i will send a friendship tie to agent j if the total surplus of doing so is positive:

$$D_{i,j} = \mathbf{1}(W_{ij}'\varphi + \theta_i + \theta_j + U_{ij} \geq 0) \tag{7}$$

where $\mathbf{1(.)}$ is an indicator function, $\theta_{i(j)}$ is agent $i(j)$'s fixed effect, and U_{ij} is an idiosyncratic component. Hence, if we assume that U_{ij} is a standard logistic random variable that is independently and identically distributed across dyads, the conditional likelihood of observing network $\boldsymbol{D} = \boldsymbol{d}$ is

$$Pr(\boldsymbol{D} = \boldsymbol{d}|\boldsymbol{Z}, \boldsymbol{\theta}) = \prod_{i \neq j} Pr(D_{ij} = d|Z_i, Z_j, \theta_i, \theta_j)$$

with

$$Pr(D_{ij=d}|\boldsymbol{Z},\boldsymbol{\theta}) = \left[\frac{1}{1+exp((W'_{ij}\varphi+\theta_i+\theta_j)}\right]^{1-d}\left[\frac{exp((W'_{ij}\varphi+\theta_i+\theta_j)}{1+exp(W'_{ij}\varphi+\theta_i+\theta_j}\right]^{d}$$

for all $i \neq j$.

Such a probability is modeled using the following conditional logistic regression function:

$$Pr(D_{ij=d}|\boldsymbol{Z},\boldsymbol{\theta}) = \frac{exp((W'_{ij}\varphi+\theta_i+\theta_j)}{1+exp((W'_{ij}\varphi+\theta_i+\theta_j)} \tag{8}$$

where W_{ij} is the distance in predetermined dyadic characteristics. More specifically, this vector includes binary variables indicating if students are similar in terms of gender, race, and week of birth.

Table A.1 presents the results of such estimation. The odd-numbered columns present raw estimations, and the even-numbered columns present odds ratios. Columns (1) and (2) show the results of the estimation considering the model described above, where homophily in pre-determined characteristis is the main mechanism in place. Column (3) and (4), in turn, present estimations considering both homophily in pre-determined characterists and students' random chances of interacting at the school, captured by similarities in their first name initials. Table A.8 presents estimations of friends' influence in schooling decisions considering these other models of friendship formation, to check the robustness of the paper's main results.

Table A.1: Friendship link formation – other specifications

	(1)	(2)	(3)	(4)
	raw estimation	Odds-ration	raw estimation	Odds-ration
Gender	1.570***	4.806***	1.567***	4.794***
	(0.053)	(0.255)	(0.053)	(0.254)
Race-white	0.128***	1.136***	0.128***	1.137***
	(0.022)	(0.025)	(0.022)	(0.025)
Race-black	0.116***	1.123***	0.115***	1.122***
	(0.023)	(0.026)	(0.023)	(0.025)
Week of birth	0.015	1.015	0.013	1.013
	(0.095)	(0.096)	(0.095)	(0.096)
First letter of name			0.373***	1.453***
			(0.053)	(0.076)
Constant	-5.166***	0.006***	-5.192***	0.006***
	(0.057)	(0.000)	(0.058)	(0.000)
N	524724	524724	524724	524724

Note: (i) This table shows results of a conditional logistic regression model that predicts the likelihood that a student i will send a friendship tie to another student j in the ninth grade of the same school; the estimation controls for i's and j's fixed effects; (ii) Standard errors clustered at the school level shown in parenthesis; (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

A.2 Auxiliary tables and figures

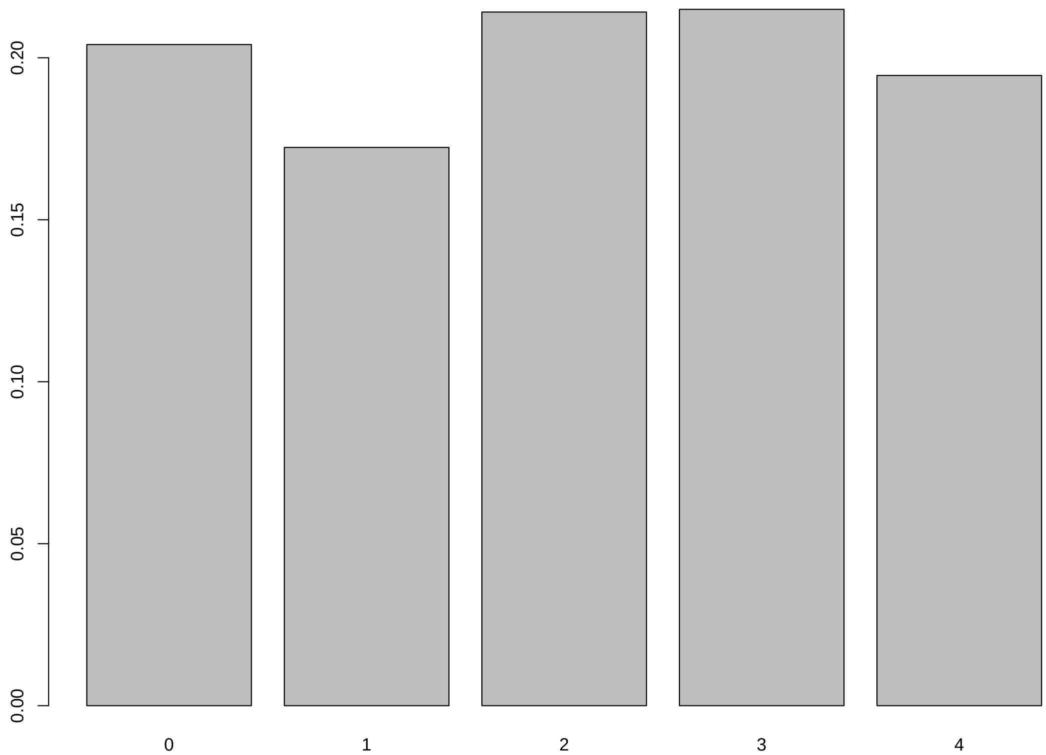

Figure A.1: Out degree distribution

Note: Each student was asked to name at most four of their best friends or colleagues at the 9^{th} grade. This graph shows the distribution of the number of named friends by each student.

Table A.2: Students' allocation into classrooms in 6th grade

	(1)	(2)
	Dependent variable: same classroom in 6th grade	
$\mathbf{1}[\boldsymbol{x_i} = \boldsymbol{x_j}]$		
First-name initial	0.881***	0.803***
	(0.264)	(0.259)
Gender	0.156	0.165
	(0.189)	(0.184)
Race	-0.192	-0.045
	(0.159)	(0.126)
Father finished HS	0.176	0.281
	(0.258)	(0.232)
Father has college degree	-0.043	0.630
	(0.587)	(0.498)
Mother finished HS	0.076	-0.059
	(0.297)	(0.264)
Mother has college degree	-0.623	-0.594
	(0.517)	(0.429)
N	640,826	640,826
Control for school FE		✓

Note: (i) This table presents estimations of conditional logistic regression models to predict the likelihood that two students i and j in the same school in 2011 were allocated into the same classroom in 2008 when they enrolled in the sixth grade. The observational unit of this table is dyads of students. The dependent variable is the likelihood that student i was enrolled in the same classroom as student j in the sixth grade. Each independent variable is a binary variable that takes value one if student i shares the same characteristic as student j; (ii) Standard errors clustered at the school level shown in parenthesis; (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table A.3: Balancing Tests – students' peers during 6th grade

	(1) Girl	(2) White	(3) Mother educ: HS	(4) Mother educ: college	(5) Father educ: HS	(6) Father educ: college
Peer outcome leave-out-mean	0.173***	0.005	-0.019	0.073*	0.027	0.001
	(0.038)	(0.042)	(0.041)	(0.041)	(0.041)	(0.039)
N	4920	4920	4920	4920	4920	4920
R2	0.039	0.059	0.048	0.028	0.041	0.047

Note: (i) All estimations control for school FE; (ii) Estimations perform the Guryan et al. (2009)'s correction to account for the negative mechanical correlation between a student's characteristics and the characteristics of the leave-out peer group; (v) The sample size is smaller than the one in the main regressions because some students' changed schools from 6^{th} to 9^{th} grade and the 6^{th}-grade peers of these students are not considered; (iv)* $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table A.5: Friends' influence on schooling decisions – comparing OLS, 2SLS, and 3SLS

	HS completion			HS completion w/out retention		
	(1)	(2)	(3)	(4)	(5)	(6)
Endogenous social effects	0.036	0.244***	0.104***	0.052**	0.315***	0.185***
	(0.023)	(0.064)	(0.036)	(0.026)	(0.089)	(0.052)
Model	OLS	IV:G^2X	IV: $\hat{G}^2X$	OLS	IV: G^2X	IV: $\hat{G}^2X$
N	6075	6075	6075	6075	6075	6075
Mean Dep. Var.	0.785	0.785	0.785	0.637	0.637	0.637
R2	0.197	0.028	0.047	0.245	0.055	0.076
Kleibergen-Paap rk LM statistic		127.019	198.885		101.661	165.322
P-val underidentification test		0.000	0.000		0.000	0.000
IVs' joint significance		75.452	380.464		49.392	165.995
Hansen J statistic		0.825	3.159		0.725	8.117
P-val overidentification test		0.991	0.789		0.994	0.230
Control for classroom FE	✓	✓	✓	✓	✓	✓

Note: (i) The "OLS" model estimates equation 2 through an OLS; the "2SLS" model estimates equation 2 using G^2X as instruments for friends' high school completion; the "3SLS" model estimates equation 2 using $\hat{G}^2X$ as instruments for friends' high school completion; (ii) Standard errors clustered at the classroom level are shown in parenthesis; (iii) All estimations include the same controls as in Table 3; (iv) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table A.4: Correlation between SES and first name initial

	(1) Girl	(2) White	(3) Mother educ: HS+	(4) Father educ:HS+	(5) Father works
First name initial=0	-0.481	-0.675	-0.655	-0.781*	-0.196
	(0.487)	(0.478)	(0.434)	(0.419)	(0.452)
First name initial=1	-0.311	-0.615	-0.614	-0.679	-0.135
	(0.483)	(0.474)	(0.430)	(0.415)	(0.448)
First name initial=2	-0.316	-0.592	-0.605	-0.671	-0.119
	(0.483)	(0.474)	(0.431)	(0.415)	(0.448)
First name initial=3	-0.365	-0.686	-0.594	-0.622	-0.189
	(0.483)	(0.474)	(0.431)	(0.415)	(0.448)
First name initial=4	-0.510	-0.671	-0.631	-0.653	-0.133
	(0.483)	(0.474)	(0.431)	(0.415)	(0.448)
First name initial=5	-0.466	-0.624	-0.619	-0.629	-0.121
	(0.483)	(0.474)	(0.431)	(0.415)	(0.448)
First name initial=6	-0.543	-0.636	-0.659	-0.640	-0.113
	(0.484)	(0.475)	(0.431)	(0.416)	(0.449)
First name initial=7	-0.467	-0.610	-0.603	-0.640	-0.163
	(0.483)	(0.474)	(0.431)	(0.415)	(0.448)
First name initial=8	-0.702	-0.605	-0.584	-0.681	-0.159
	(0.485)	(0.476)	(0.432)	(0.417)	(0.450)
First name initial=9	-0.405	-0.623	-0.553	-0.602	-0.129
	(0.484)	(0.475)	(0.431)	(0.416)	(0.449)
First name initial=10	-0.435	-0.696	-0.635	-0.657	-0.160
	(0.483)	(0.474)	(0.430)	(0.415)	(0.448)
First name initial=11	-0.255	-0.664	-0.602	-0.635	-0.128
	(0.483)	(0.474)	(0.431)	(0.415)	(0.448)
First name initial=12	-0.461	-0.637	-0.582	-0.637	-0.121
	(0.483)	(0.474)	(0.430)	(0.415)	(0.448)
First name initial=13	-0.436	-0.653	-0.662	-0.664	-0.160
	(0.483)	(0.474)	(0.430)	(0.415)	(0.448)
First name initial=14	-0.095	-0.612	-0.588	-0.647	-0.110
	(0.484)	(0.475)	(0.431)	(0.416)	(0.449)
First name initial=15	-0.819	-0.655	-0.677	-0.539	-0.100
	(0.500)	(0.491)	(0.446)	(0.430)	(0.464)
First name initial=16	-0.333	-0.623	-0.612	-0.660	-0.151
	(0.484)	(0.475)	(0.431)	(0.416)	(0.449)
First name initial=17	0.055	-0.979*	-0.678	-0.896**	-0.141
	(0.527)	(0.517)	(0.470)	(0.453)	(0.489)
First name initial=18	-0.532	-0.666	-0.623	-0.661	-0.156
	(0.483)	(0.474)	(0.431)	(0.415)	(0.448)
First name initial=19	-0.220	-0.605	-0.679	-0.667	-0.131
	(0.483)	(0.474)	(0.431)	(0.416)	(0.449)
First name initial=20	-0.101	-0.610	-0.625	-0.667	-0.128
	(0.483)	(0.474)	(0.431)	(0.415)	(0.448)
First name initial=21	-0.616	-0.557	-0.845*	-0.873**	-0.299
	(0.515)	(0.506)	(0.459)	(0.443)	(0.478)
First name initial=22	-0.529	-0.646	-0.560	-0.582	-0.248
	(0.483)	(0.474)	(0.431)	(0.416)	(0.449)
First name initial=23	-0.903*	-0.641	-0.664	-0.696*	-0.170
	(0.484)	(0.475)	(0.431)	(0.416)	(0.449)
First name initial=24	-0.359	-0.711	-0.704	-0.679	-0.207
	(0.487)	(0.478)	(0.434)	(0.419)	(0.452)
First name initial=25	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)
N	6075	6075	6075	6075	6075
R2	0.554	0.366	0.280	0.262	0.741

Note: (i) All estimations control for school FE; (ii) First name initial= 0 represents students with missing names (60 students); (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table A.6: Friends' influence on schooling decisions – Robustness checks

	HS completion		HS completion w/out retention	
	(1)	(2)	(3)	(4)
Endogenous social effect	0.111***	0.103***	0.162***	0.175***
	(0.022)	(0.037)	(0.026)	(0.055)
Own characteristics				
Girl	0.060***	0.055***	0.103***	0.111***
	(0.011)	(0.016)	(0.013)	(0.018)
White	0.008	0.010	0.006	0.008
	(0.012)	(0.013)	(0.014)	(0.015)
Mother education: more than HS	0.041***	0.036***	0.060***	0.067***
	(0.011)	(0.013)	(0.014)	(0.016)
Father education: more than HS	0.010	0.017	0.017	0.031*
	(0.014)	(0.016)	(0.015)	(0.017)
Reading proficiency (2009)	0.044***	0.048***	0.061***	0.063***
	(0.007)	(0.008)	(0.007)	(0.009)
Math proficiency (2009)	0.023***	0.021***	0.037***	0.035***
	(0.006)	(0.008)	(0.008)	(0.009)
Father works	0.031**	0.032**	0.049***	0.050***
	(0.013)	(0.014)	(0.014)	(0.015)
Friends' characteristics				
Girl		0.017		-0.016
		(0.021)		(0.027)
White		0.011		0.035
		(0.020)		(0.022)
Mother education: more than HS		-0.040*		-0.052**
		(0.022)		(0.025)
Father education: more than HS		0.035		0.052*
		(0.023)		(0.027)
Reading proficiency (2009)		0.006		0.008
		(0.012)		(0.013)
Math proficiency (2009)		0.013		-0.005
		(0.012)		(0.014)
Father works		-0.009		-0.024
		(0.021)		(0.026)
N	6075	4893	6075	4893
Mean Dep. Var.	0.785	0.768	0.637	0.613
R2	0.045	0.049	0.076	0.076
Kleibergen-Paap rk LM statistic	223.813	190.318	216.106	154.872
P-val underidentification test	0.000	0.000	0.000	0.000
IVs' joint significance	1047.228	311.694	555.552	138.575
Hansen J statistic	2.978	2.057	7.018	6.623
P-val overidentification test	0.812	0.914	0.319	0.357
Control for classroom FE	✓	✓	✓	✓
Maximum out-degree	4	3	4	3

Note: (i) This table shows estimations of models like the one described in equation 2, in which friends' high school completion is instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) "Endogenous social effect" is the effect that friends' outcomes have on students' outcomes. In columns (1) to (3), this is the impact of the average of friends who finished HS, while in columns (4) to (6), this is the impact of the average of friends who finished HS without retention; (iii) Maximum out-degree is the maximum number of friends named by a student; (iv) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table A.7: Placebo exercises

	(1) Mother educ.HS+	(2) Father educ. HS+	(3) Own house
Endogenous social effects	-0.071	0.020	-0.014
	(0.107)	(0.113)	(0.065)
N	6075	6075	6075
Mean Dep. Var.	0.684	0.684	0.684
R2	0.013	0.008	0.005
Kleibergen-Paap rk LM statistic	36.206	32.817	41.042
P-val underidentification test	0.000	0.000	0.000
IVs' joint significance	39.698	33.325	79.661
Hansen J statistic	3.082	2.405	0.607
P-val overidentification test	0.544	0.662	0.962
Control for classroom FE	✓	✓	✓

Note: (i) This table shows estimations of models like the one described in equation 2, in which friends' outcomes are instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) "Endogenous social effect" is the effect that friends' outcomes have on students' outcomes. In column (1) this is the impact of the average of friends whose mother finished HS, in column (2) this is the impact of the average of friends whose father finished HS, and in column (3) this is the impact of the average of friends who live in a owned house; (iii) All estimations include the same controls as in Table 3 (iv) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table A.8: Friends' influence on schooling decisions – other models of friendship formation

	HS completion		HS completion w/out retention	
	(1)	(2)	(3)	(4)
Endog. social effects	0.104***	0.104***	0.189***	0.188***
	(0.036)	(0.036)	(0.054)	(0.054)
Model	IV: $\hat{G'}^2X$	IV: $\hat{G''}^2X$	IV: $\hat{G'}^2X$	IV: $\hat{G''}^2X$
N	6075	6075	6075	6075
Mean Dep. Var.	0.785	0.785	0.637	0.637
R2	0.047	0.047	0.076	0.076
Kleibergen-Paap rk LM statistic	195.739	195.846	160.978	161.239
P-val underidentification test	0.000	0.000	0.000	0.000
IVs' joint significance	354.232	353.762	152.588	152.757
Hansen J statistic	3.379	3.471	5.844	5.899
P-val overidentification test	0.760	0.748	0.441	0.435
Control for classroom FE	✓	✓	✓	✓

Note: (i) This table shows estimations of models like the one described in equation 2, in which friends' outcomes are instrumented by the predicted friends-of-friends' characteristics ($\hat{G'}^2X$ or $\hat{G''}^2X$); (ii) "Endogenous social effect" is the effect that friends' outcomes have on students' outcomes. In columns (1) and (2) this is the impact of the average of friends who finished HS, in columns (3) and (4) this is the impact of the average of friends who finished HS without retention; (iii) $\hat{G'}^2X$ is the adjacency matrix coming from a model of link formation that includes students similarities in pre-determined characteristics (but not name similarities) in the set of controls; (iv) $\hat{G''}^2X$ is the adjacency matrix coming from a model of link formation that includes students similarities in pre-determined characteristics and name similarities in the set of controls; (v) Standard errors clustered at classroom level shown in parenthesis; (vi) All estimations include the same controls as in Table 3; (vii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table A.9: Mediation analysis

	(1) HS completion	(2) HS completion w/out retention
Total mediated effect	0.069*	0.157**
	(0.031)	(0.053)
Own coll. aspiration	0.009**	0.016**
	(0.003)	(0.005)
Friends' coll. aspiration	0.037*	0.073**
	(0.017)	(0.027)
Own perceived HS returns	0.001	0.001
	(0.001)	(0.001)
Friends' perceived HS returns	0.001	0.035
	(0.019)	(0.030)
Own fear of nerd stigma	0.001	0.001
	(0.001)	(0.001)
Friends' fear of nerd stigma	0.001	0.007
	(0.006)	(0.009)
Own math study time	0.001	0.001
	(0.001)	(0.002)
Friends' math study time	0.007	0.004
	(0.008)	(0.012)
Own performance	0.012**	0.022***
	(0.004)	(0.006)
Friends' performance	-0.001	-0.002
	(0.002)	(0.003)
Observations	6075	6075

Note: (i) The total mediated effect represents how much friends' influence is reduced once the model includes the new controls. The coefficients below the total mediated effect represent the contribution of each control in explaining the variation of friends' influence; (ii) All estimations include the same controls as in Table 3; (iii) Standard errors clustered at the classroom level are shown in parenthesis; (iv) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.