An aspiring friend is a friend indeed: school peers and college aspirations in Brazil*

Jessica Gagete-Miranda †

Abstract

Aspirations are a fundamental determinant of one's effort and investments. Due to its consequences for individuals' future outcomes, understanding the process of aspirations formation helps to inform public policies. This work asks whether peers play a role in such a process. I use novel data on Brazilian students' networks, matched with administrative data, and investigate whether students' college aspirations spill over to their friends. The employed methodology acknowledges that social cliques are formed endogenously and addresses this challenge by modeling friendship formation based on similarities in predetermined characteristics. Using the predicted friendship links, I explore network structures and use predicted friends of friends' characteristics as instruments for friends' aspirations. The results show evidence of positive, significant, and quite large peer effects on aspirations - an extra friend aspiring to go to college increases, on average, 11.25% the likelihood that a student will also aspire to it. In a discussion about the possible mechanisms, I verify the existence of peer effects on certain social norms in the school, as well as on class attendance and school effort. However, peers' performance and socioemotional skills do not have an impact on student's performance and socioemotional skills, respectively. Finally, I investigate if friends' aspirations impact students' future outcomes. While friends' aspirations do not influence students' future proficiency, an extra aspiring friend decreases, on average, 4.75 percentage points the likelihood of dropping out of secondary school.

Keywords: Peer effects, Social Networks, Education, Aspiration, Human Capital Accumulation

JEL Codes: I24, I25, I29

---

*I am very grateful to Marcos A. Rangel, Paolo Pinotti, Eliana La Ferrara, and Massimo Anelli for their guidance over the course of this project. I also thank the valuable comments of Samuel Berliski, Eric Bettinger, Teodora Boneva, Matias Busso, Sebastian Calonico, Jane Cooley Fruehwirth, Armin Falk, Garance Genicot, Aureo de Paula, Michele Pellizzari, and seminar participants at the 2018 Summer School on Socioeconomic Inequality, the 9^{th} Alp-Pop Conference, the 2019 Workshop on Poverty and Inequality (X RIDGE Forum), the 2019 Empirics and Methods in Economics Conference, the 2019 APPAM Fall Research Conference, the 2020 PacDev Conference, the 2020 MWIEDC conference, the 5th IZA Workshop in the Economics of Education, and seminars at Bocconi University, the Sanford School of Public Policy at Duke, and the Duke Network Analysis Center. The usual caveats apply.

†Bocconi University, Via Sarfatti, 25 Milano, 20136, Italy; jessica.gagete@phd.unibocconi.it.

1 Introduction

The capacity to aspire to a better standard of living is an important driver of one's effort and investments (Appadurai, 2004; Ray, 2006). The recent literature on development economics considers the lack of such a capacity - or aspirations failures - as a psychological constraint that might trap people into poverty (Dalton, Ghosal, & Mani, 2016; Genicot & Ray, 2017). Understanding the determinants of aspirations is hence a powerful tool for policymaking. Such determinants might help to explain, for instance, relevant patterns of behavior, such as under-investment in education (Kearney & Levine, 2014; La Ferrara, 2019).

The theoretical literature argues that aspirations emerge in social contexts, through individuals' comparisons with similar others (Appadurai, 2004; Bogliacino & Ortoleva, 2013; Genicot & Ray, 2017; Ray, 2006). This crucial social element of aspirations construction calls for investigations on how exactly peers influence an individual's level of aspirations. Empirical contributions have found that peers' socioeconomic status is associated with individuals' aspirations (e.g. Janzen, Magnan, Sharma, and Thompson (2017); Stutzer (2004)). However, an important question still needs further investigation: do peers' aspirations influence one's own aspiration, above and beyond socioeconomic considerations? That is, after controlling for socioeconomic status, do peers still influence individuals' aspirations, through their own level of aspiration?

This work investigates peer effects on students' college aspirations - that is, how peers' desire to pursue a college degree impacts students' desire to also pursue such a degree. To do so, I rely on a unique social networks data collected among middle school students in Brazil to address the main challenges that emerge when identifying peer effects. Differently from standard linear-in-means models, my work does not assume that all individuals in a students' reference group are equally connected or have the same influence on each other. Instead, I acknowledge that individuals in social networks are idiosyncratically connected and that homophily - i.e., the tendency to form social clicks with similar others - plays an important role in friendship formation.

Linking these network data with administrative data, I model friendship formation based on homophily in predetermined characteristics. Next, based on the model's predicted connections, the identification strategy uses predicted friends of friends' characteristics as instrumental variables for friends' aspirations. It also uses network fixed effects and a broad set of controls to eliminate other possible correlated effects.

College aspirations are quite a relevant measure of aspirations in the educational scenario of developing countries. On the one hand, these countries have a high earnings premium of tertiary education, compared to other OECD and partner countries. On the other hand, they have low percentages of adults attaining such a level of education. Brazil is a good example: someone with a bachelor's degree in Brazil earns over 2.4 times what someone who only attained upper secondary education earns - the highest earnings premium among OECD and partner countries. Still, only 15% of the country's adult population has achieved tertiary education - well below the OECD average of 37% (OECD, 2017). Hence, aspiring to a college degree in a developing country is a good indicator of aspirations towards a good living standard.

I document that college aspirations at 9^{th} grade, the last grade of middle school in Brazil, are associated with students' future outcomes at school, such as dropout rates, truancy rates, and secondary school achievement. Then I go on to show evidence of positive, significant, and quite large peer effects on aspirations: an extra friend aspiring to a college degree increases a student's likelihood of also aspiring to it from 3.8 p.p. (5.6%) to 15.3 p.p. (22.5%), depending on the number of nominated friends. The results are fairly homogeneous for different students' characteristics, such as gender, race, and parental education. On a discussion about the possible mechanisms behind such an impact, I verify the existence of peer effects on some social norms in school, class attendance, and the amount of time that students dedicate to studying math. However, differently from previous literature on peer effects, I do not find significant peer effects on reading and math performance. I also do not find peer effects on students' socio-emotional skills.

I finally ask whether peers' aspirations influence students' future school outcomes, such as retention, dropout, class attendance, and performance. While there is no impact of peers' aspirations on students' performance or class attendance, I do find that peers' aspirations decrease the likelihood of dropping out of school by, on average, 4.75 percentage points.

This study adds to traditional contributions of the sociology literature on aspirations (Kao & Tienda, 1998; Sewell & Hauser, 1975; Sewell & Shah, 1968). Sociologists have long verified the existence of a positive correlation between peers' aspirations and one's own aspirations (see, for instance, the work of Campbell and Alexander (1965); Cohen (1983); Duncan, Haller, and Portes (1968)). Identification issues, however, have prevented these studies from establishing causal relationships. In the context of peer effects estimations, correlated effects - socioeconomic background, school quality, or homophily in friendship formation - might deliver a high correlation between a student's outcomes and her peers' outcomes even in the absence of peers' influence. Moreover, the reflection problem - the simultaneity of outcomes that emerges in groups' interactions - will most likely overestimate any existing peer effects (Manski, 1993).[1] Hence, correlational studies say little about the real impact that peers exert on one's aspirations. The identification strategy employed in this work allows me to establish a causal relationship between aspirations and peers' aspirations.

This work also contributes to the literature on peer effects (see Sacerdote (2011) for a review). Most contributions on primary and secondary schools focus on peer effects in test scores and look at different sources of peer effects - such as ability, gender or racial composition, parental characteristics, or behavior (Austen-Smith & Fryer Jr, 2005; Carrell & Hoekstra, 2010; Fruehwirth & Gagete-Miranda, 2019; Hanushek, Kain, Markman, & Rivkin, 2003; Hoxby, 2000; Lavy & Schlosser, 2011; Marotta, 2017). Other studies focus on peer effects in students' attitudes and behavior, such as substance use, school dropout, and criminal activity (Case & Katz, 1991; Gaviria & Raphael, 2001).

When it comes to the influence that peers have on students' aspirations, Jonsson and

[1] For a discussion about the challenges on the estimation of peer effects, see Angrist (2014).

Mood (2008) show that having high-achieving peers might depress the average students' desire to attend college. However, evidence on the impact that peers' aspirations have on students' aspirations is still scarce. My results conform with those found in concurrent work on secondary schools in Britain (Dickerson, Maragkou, & McIntosh, 2018), and high schools in the U.S (Norris, 2020), and my investigation complement their findings by detailing mechanisms through which friends influence students' aspirations. To the best of my knowledge, I am the first to show that peers' aspirations also impact students' outcomes in schools, such as their chances of graduating in secondary school. This finding is particularly important for developing countries, where school dropout rates are considerably high.[2]

2 Data and measure of aspirations

The primary source of data used in this work comes from a survey conducted in 2011 on students from the 9^{th} grade of state-owned middle schools of Sao Paulo (Brazil). I combine this survey with administrative data to recover information on students' socioeconomic background, performance, class attendance, and school path - that is, each school and class that students were enrolled at each year of their education. In what follows, I first provide some background information about the Brazilian school system and relevant details about the provision of public education on the State of Sao Paulo. Then, I describe the main characteristics of the survey and the construction of my measure of students' aspirations.

2.1 Institutional background

Basic education in Brazil is divided into preschool (attended by students up to the age of 6), primary school (attended by 6 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 1 to 5 - and middle school - grades 6

[2]In low- and middle-income countries, only 35% of students 35 complete upper secondary schooling (Filmer & Rogers, 2018).

to 9. The main subjects taught at this level are language (Portuguese), mathematics, social sciences, and sciences. Students are allocated to classes at the beginning of each academic year. They attend all subjects together with the same classmates for the whole year.

In the State of Sao Paulo, where my sample of schools comes from, municipal governments are the main responsible for the provision of primary school while the State government is the main responsible for the provision of secondary school. Hence, the vast majority of students have to switch schools in the transition from 5th to 6th grade. State-owned schools in Sao Paulo are usually larger than municipal-owned schools. In 2011, when the survey was applied, the average number of students enrolled in State-owned schools was 1,189, while the average number of students enrolled in municipal-owned schools was 697. Class size is also larger in State-owned schools, with an average of 39 students per class in 2011, while this figure for municipal-owned schools was 27 students per class.[3]

Due to the decentralized nature of the educational system in Sao Paulo, state-owned schools receive virtually no information of the students coming from municipal-owned schools. That is, when students enroll in a state-owned school in the 6th grade, the administration of this school does not know anything about these students' background, such as previous performance or behavior. Therefore, the allocation of students into classes in the 6th grade cannot take such characteristics into consideration. Due to the decentralized nature of the educational system in Sao Paulo, state-owned schools receive virtually no information about students coming from municipal-owned schools. Hence, when students enroll in a state-owned school in the 6th grade, this school's administration does not know anything about these students' backgrounds, such as previous performance or behavior. Therefore, the allocation of students into classes in the 6th grade cannot take such characteristics into consideration. Students are likely allocated in alphabetical order.[4] As I show in Section 3, I exploit this feature in the Sao Paulo educational system to model friendship formation.

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

[4]Anecdotal evidence from students in my sample shows that students who share the first letter of their names have a higher likelihood of being allocated into the same class in the 6th grade. I explain this in details in Section 3

2.2 Survey on students' profile and friendship ties

In 2011, students from the 9th grade of 85 state-owned schools in Sao Paulo answered a very comprehensive questionnaire about their personal profile, how happy or satisfied they were with their life, what were their study habits and educational aspirations.

One block of questions in this survey mapped students' social networks. They were asked to nominate their four best friends or colleagues in their grade (which, in most schools, comprehends more than one classroom).[5] Importantly, it is possible to link the nominated students to school rosters, and also to find their own answers to the questionnaire. As so, it is possible to map the network for all students of 9^{th} grade in each school.

Another block of questions was dedicated to understanding students' educational aspirations. One specific question of this block asked until when students would like to keep studying if this choice was *entirely* up to them. I use this question to build my measure of aspirations towards pursuing a college degree, which I call *college aspirations*. This is a binary variable that takes value equal one if students answered that they would like to keep studying until they get a college degree and zero otherwise. The framing of such a question is relevant here: since students were asked to reveal their preferences towards their educational future, disregarding any kind of constraint that they might face to achieve such a future, this measure captures students' genuine aspirations, and not only their expectations for the future.

The survey also approached other traits and beliefs that might be associated with college aspirations. First, it had a block of questions dedicated to extracting students' socioemotional profile. From it, it is possible to identify traits such as self-esteem, self-efficacy, self-control, agreeableness, rapport with peers, and locus of control.[6] Second, it asked stu-

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

[6] The instruments used to measure students' socio-emotional skills were adapted from the international literature. They were based in the woks of Rotter (1966), to measure locus of control; Tangney, Baumeister, and Boone (2004), to measure self-control; Bandura (1997) to measure self-efficacy; and Rosenberg (1986), to measure self-esteem. Some questions on the Strengths & Difficulties Questionnaires (http://www.sdqinfo.org/) were also included to build measures of agreeableness and rapport with peers.

dents which probability they attributed to finding a job in the future if they have a university degree.[7] This question measures students' perceptions about college returns, which might influence their aspirations towards pursuing a college degree. Finally, students were also asked about possible impediments for them to keep studying in the future. Two impediments, in particular, might also be related to students' college aspirations. The first is students' concern about being stigmatized as "nerd" if they put too much effort into school - I call this variable "Fear of nerd stigma". The second is peers pressuring students to find a job and start earning their own money - I call this variable "Peer pressure to work". I will use students' perceived college returns and these two impediments - which proxy for students' willingness to comply with "bad" social norms in the school - to discuss the mechanism behind my results.

Table 1 presents some descriptive statistics coming from this survey and administrative data, such as students' college aspirations, their demographic and socioeconomic characteristics, and their proficiency in Language and Math in standardized tests - known as Sao Paulo School Performance Assessment System (SARESP, in the Portuguese acronym) - applied every year to all state-owned schools in Sao Paulo. The table presents the mean and standard error for all students and also for those who aspire and those who do not aspire to a college degree. It also shows this same information for students' friends. First, looking at the sample composed of all students, we see that more than 30% of them do not aspire to a college degree. Second, comparing students who aspire to a college with students who do not, those who do want to go to college are better achieving and have, on average, better-educated parents. Finally, looking at the average characteristics of students' friends, we see that the friends of students aspiring to a college are also more likely to aspire to it - which could be an indicator of peer effects - but are also more likely of being high achieving students and of having more educated parents - which might exemplify the phenomenon of homophily, that

[7] This question was framed in the following way. First, students were asked to think of ten other students very similar to them in the school. They were then asked to indicate how many of these students would find a job if they went to college.

is, people's tendency to befriend with similar others. Homophily is an important confounder in the estimation of peer effects. Section 3 details how this article overcomes such an issue.

2.3 College aspirations and future outcomes

An important feature of this survey is the possibility to link it with administrative data and to recover students' school path - which allows one to know whether these students dropped out from school at some point or were retained in some grade -, as well as their past and future performance in SARESP. This allows me to test the association of college aspirations with students' future outcomes in school.

Associating college aspirations with the outcomes mentioned above is an important exercise to understand whether such a measure of aspirations goes in the expected direction. However, it is important to highlight that such an exercise does not allow for any causal interpretation. With this caveat in mind, I perform OLS estimations with each of those measures of school outcomes as dependent variables, and college aspirations as the independent variable, controlling for students' performance, demographics, socioeconomic status, and school fixed effects. Standard errors are clustered at the school level. Figure 1 presents the point estimate and the 95% confidence interval of these estimations.[8]

As described by the figure, college aspirations are highly associated with the likelihood of having a normal school path during secondary school (that is, of being at the 12^{th} grade in 2014), with class attendance in reading and math during 11^{th} grade (2013)[9], and with students' performance in the last year of secondary school (2014). At the same time, college aspirations are negatively correlated with the likelihood of school dropout during secondary school.

These exercises show that such a measure of aspirations has a strong predictive power

[8] Besides parents' education, I also use father's working status, house ownership, internet, and the number of lavatories in the house as measures of the socioeconomic status. I omitted these variables from the figures for the sake of clarity.

[9] It was not possible to recover the information on class attendance for the last year of secondary school, in 2014.

over several important school outcomes, which is a good indicator that it is indeed capturing students' aspirations.

3 Identification of peer effects

One faces several challenges when seeking to identify endogenous social effects through a linear-in-means model - that is, associating an individual's outcomes with the average outcome of her reference group on the attempt to infer whether the group behavior influences the behavior of individuals inside that group.

The first challenge is the reflection problem (Manski, 1993), namely, a simultaneity bias that emerges due to the fact that an individual might influence the behavior of her group and, at the same time, might be influenced by the group's behavior. In a friendship network, for instance, all friends potentially impact each other, so it is difficult to disentangle if one's behavior is the cause or the consequence of others' behavior.

The second challenge is correlated effects, where people in the same reference group tend to behave alike not because they influence one another but because they share similar unobserved characteristics, such as institutional environments or common shocks. For instance, students within a school are influenced by school quality, or maybe by a very inspiring professor.

Finally, connections or friendship links do not happen at random, 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 (Currarini, Jackson, & Pin, 2009; McPherson, Smith-Lovin, & Cook, 2001; Moody, 2001). An important implication of homophily and the endogenous formations of networks is that neither the connections nor the influence of individuals inside a reference group are equal for everyone. Even students enrolled at the same school and under the mentoring of the same teachers form different cliques to one another. This brings extra challenges to the estimation of peer effects since individuals might have unobserved characteristics correlated to both their outcomes and their links formation.

Several works on the peer effects literature have tackled these identification problems, with different strategies. Some use natural experiments in order to solve correlated effects (Cipollone & Rosolia, 2007; Sacerdote, 2001; Zimmerman, 2003), other use theoretical models of social interactions (Brock & Durlauf, 2001) or network structures (Boucher, Bramoullé, Djebbari, & Fortin, 2014; Bramoullé, Djebbari, & Fortin, 2009; Calvó-Armengol, Patacchini, & Zenou, 2009; De Giorgi, Pellizzari, & Redaelli, 2010; Liu, Patacchini, & Zenou, 2014) in order to address both correlated effects and the reflection problem.

Fewer works have fully acknowledged the implications of the endogenous formation of networks and tackled this problem accordingly. Johnsson and Moon (2017) develop a semi-parametric control function approach to deal with this issue. Goldsmith-Pinkham and Imbens (2013) model link formation assuming that individuals with similar observed and unobserved characteristics are more likely to form links, and perform a sample selection correction where network formation and the outcome are jointly determined. König, Liu, and Zenou (2018) use a three-stage least square (3SLS) strategy where, in the first stage, they model link formation based on past network structures as exclusion restrictions that affect current link formation but do not enter the outcome equation (König et al., 2018).[10] The second and third stages are similar to the ones implemented by Bramoullé et al. (2009) and De Giorgi et al. (2010) where friends' outcomes are instrumented by friends' of friends characteristics. The main difference is that, when building the instruments, the endogenous sociometric matrix is replaced by the predicted one that comes from the link formation model. In this work, I follow this 3SLS approach. Besides modeling friendship formation using students' pre-determined characteristics, I also look at their random chances of interacting due to random allocation into classes when students enroll in middle school and within classes at each

[10] In an interesting application of this methodology, (Santavirta & Sarzosa, 2019) uses individuals' predetermined characteristics to model link formation.

year. In what follows, I formalize my model of friends' influence, the identification issues, and the 3SLS estimation.

3.1 Model of friends' influence

Let a student's college aspirations be affected by her friends' mean college aspirations, her characteristics such as grades, gender, race, and family background, and the mean characteristics of her friends. More formally, suppose there is a set of students i, $i = (1...N)$, that belong to network l, $l = (1, ...L)$ [11]. 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 her own group of friends, such that $i \notin F_i$.[12] The model is given by:[13]

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

where y_{li} is the aspirations level of individual i in network l , which depends on the aspirations level of the friends directly connected to her - the endogenous social effect in Manski's notation (see Manski (1993)) -, on x_{li}, her own characteristics.[14], on the characteristics of her 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 $|\beta| < 1$.

Let G be the adjacency matrix, where element $g_{i,j} = 1/n_i$ if individual i sends a friendship

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

[12] 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 & Fafchamps, 2016; Guryan, Kroft, & Notowidigdo, 2009). The exclusion of an individual i from the pool of i's peers creates a mechanical negative relationship between i's characteristics and that of her 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).

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

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

tie to individual j, and $g_{i,j} = 0$ otherwise. Assume that $g_{i,i} = 0$ so that each individual is not part of her own reference group. The above model can then be translated into:

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

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 explicitly, if one assumes for a moment that $\boldsymbol{G}$ is orthogonal to $\boldsymbol{\upsilon_l}$, it is possible to causally estimate the reduced form of equation 2[15]:

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

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

Correlated effects would emerge if μ_l was not observed by the modeler, since $\boldsymbol{X_l}$ is only exogenous conditional on μ_l. School quality, for instance, is probably correlated with students' aspirations. Hence, students within the same school are more likely to have similar levels of college aspiration, which could bias estimations upwards. I address this problem by simply controlling the estimations by network fixed effects - in my case, the same as school fixed effects.

Nonetheless, this does not solve the endogeneity of link formation. That is, individuals do not befriend each other at random and homophily plays a great role in friendship formation, which yields $\boldsymbol{G} \not\perp \boldsymbol{\upsilon_l}$. Once again, such correlation would most likely bias estimates upwards, since more similar students have a greater probability of becoming friends and, at the same time, are more likely to have similar aspirations towards college.

As in König et al. (2018), I will tackle the reflection problem and the endogenous forma-

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

tion of friendship using a 3SLS estimation. The first stage models link formation based on homophily in predetermined characteristics. The second and third stages use the predicted friendship connections delivered by the first stage and use friends of friends' characteristics as instrumental variables for friends' aspirations (resembling Bramoullé et al. (2009)). In the remainder of this section, I describe this approach and explain how it overcomes the issues raised above. For the sake of clarity in exposition, I start by describing the last two stages of the implemented strategy, which address the reflection problem, and then I describe the first stage and show 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. [16] The authors prove that if the diameter[17] of the network is greater than or equal to 3, then the linear independence between $\boldsymbol{I}$, $\boldsymbol{G}$, $\boldsymbol{G}^2$, and $\boldsymbol{G}^3$ is guaranteed and the model is identified[18].

Therefore, in order 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 $\boldsymbol{S} = [\boldsymbol{X_l}\ \boldsymbol{GX_l}\ \boldsymbol{G^2X_l}\ \boldsymbol{G^3X_l}]$, such that the final estimates are given by $\hat{\varphi}^{2SLS} = (\tilde{\boldsymbol{X}}'\boldsymbol{P}\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X'}}\boldsymbol{Py_l}$, where $\boldsymbol{P} = \boldsymbol{S}(\boldsymbol{S'S})^{-1}\boldsymbol{S}$.

The intuition behind this strategy is that, 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

[16] 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.

[17] 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))."

[18] The counterpart for the diameter size in a model where 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

indirectly, through the friendship tie between B and C. Therefore, A's characteristics are valid instruments for B's outcomes.

3.3 Endogenous link formation

The aforementioned 2SLS strategy would ensure unbiased estimates of the endogenous and exogenous social effects if friendship links were formed at random - that is, if $\boldsymbol{G} \perp v_l$. However, as stated before, social networks are not formed at random and homophily plays a role in clique formation. König et al. (2018) deal with such an issue including a stage before the 2SLS, where they use predicted networks based on predetermined characteristics to build the IVs that identify the social effects.

The work of Graham (2017) 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 $Z_i = \{z_{1i}, ...z_{Ki}\}$. 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:

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

where $\mathbf{1(.)}$ is an indicator function, $\theta_{i(j)}$ is agent $i(j)$'s fixed effect, and U_{ij} is an idiosyncratic component ($U_{ij} = U_{ji}$ if the network is undirected and $U_{ij} \neq U_{ji}$ if the network is directed). 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 $i \neq j$.

I model such a probability using the following conditional logistic regression function:

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

where W_{ij} is the distance in predetermined dyadic characteristics. More specifically, this vector includes binary variables indicating whether students are similar in terms of gender, race, and if their first name start with the same letter. The inclusion of similarities in terms of gender and race is purely due to homophilly. The intuition behind the inclusion of name proximity is that such similarities might increase students interactions in school. Inter-classroom and within-classroom allocations are likely to be driven at least in part by students' first name alphabetical order. First, as described in section 2, the allocation of students into classrooms when they first enroll in State-owned schools is not based on students' previous performance or behavior, since the administration of State-owned schools do not have such information. Therefore, students are usually allocated in alphabetical order. Indeed, as presented in Table A.1, sharing the first letter of name is the only significant predictor of the likelihood that two students from my sample were allocated into the same classroom in 6th grade. Other similarities such as gender, race, and parental education are not good predictors of such allocation. Within-class allocation - such as the choice of students seats - is usually also done in alphabetical order. Hence, students whose name share the first letter have more chances of interaction since the day they arrive at the school.

Table 2 presents the results of such estimation. As it is possible to see, being the same gender and race, and sharing the fist letter of name are highly correlated with the likelihood of forming friendship ties.

Using the predicted links coming from this model, I replace the original adjacency matrix by 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 $\tilde{\boldsymbol{X}}$ =[$\boldsymbol{Gy_l}$ $\boldsymbol{X_l}$ $\boldsymbol{GX_l}$] is instrumented in the second stage by $\hat{\boldsymbol{S}} = [\boldsymbol{X_l}\ \boldsymbol{GX_l}\ \hat{\boldsymbol{G}}(\boldsymbol{W})^{\boldsymbol{2}}\boldsymbol{X_l}$ $\hat{\boldsymbol{G}}(\boldsymbol{W})^{\boldsymbol{3}}\boldsymbol{X_l}]$, where $\hat{\boldsymbol{G}}(\boldsymbol{W})$ is the predicted adjacency matrix from equation 5, $\hat{\boldsymbol{D}}(\boldsymbol{W})$, 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 $\hat{\boldsymbol{P}} = \hat{\boldsymbol{S}}(\hat{\boldsymbol{S}}'\hat{\boldsymbol{S}})^{-1}\hat{\boldsymbol{S}}$.

3.4 Potential threats to identification

This section discusses the validity of the identifying assumptions in the implemented methodology and potential threats that might emerge due to how students' networks were mapped in my data.

As specified in Bramoullé et al. (2009), the identification of peer effects using friends of friends as instrumental variables is only possible if there are intransitive triads in the network. That is, students within a network cannot be all friends among themselves. This would invalidate the instruments' exclusion restriction since all the friends of my friends would also be my friends. That is why one needs I, G, G^2, and G^3 to be linear independent. As shown in the previous section, Bramoullé et al. (2009) prove that a sufficient condition to guarantee such linear independence is that the diameter of a network is greater than or equal to 3. The average size of the diameters in my networks is 14.3, with a minimum size of 4 and a maximum size of 22, so the linear independence between I, G, G^2, and G^3 is secured for all schools in my sample.

A second 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. One needs this assumption to guarantee that intransitive triads in the network are indeed intransitive. In order words, if we observe that A is connected to B, and B is connected to C, but C is not connected to A, we need to be sure that the absence of connection between A and C is not due to missing or censored data. Such an assumption is also crucial for the model of friendship formation proposed by Graham (2017), since one should be able to identify all connections in a network in order to model them fully.

In that sense, my data might suffer from a ceiling effect, since students could nominate only four of their friends. If a student had a fifth or sixth friend in that grade, these connections do not show up in my data. Figure A.1 presents the out-degree distribution, that is, the distribution of the number of friends that each student nominated. It shows that around 20% of students might be suffering from this ceiling effect since they nominated all four friends, and it is not possible to know whether there were more friends they would like to nominate. However, it is reassuring to see that this is not the majority of students - around 60% of students nominated either one, two, or three friends, so they were not censored in any way[19]. Moreover, 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 - show that, if anything, censoring the number of friends bias the results *downwards.* Still, in section 4.1, I present some robustness checks to address potential issues with censored networks.

Another potential threat to identification is that students' aspirations towards going to college might be directly affected not only by their friends but also by other colleagues. A high achieving colleague might either be a good role model, which could increase students' aspirations, or be seen as a competitor, which could hinder such aspirations.[20] If this colleague is a friend of a friend, the exclusion restriction of the instruments might be threatened. Controlling for classroom fixed effects on top of school fixed effects alleviates such a problem since students' ranking and competitive dynamics within the classroom will be held constant.

[19]Figure A.1 also shows that around 20% of students did not nominate any friend. This proportion is in the same order as the one in Add-Health data (Niño, Cai, & Ignatow, 2016). Exercises - not shown - either controlling for isolated students or excluding them from the estimation show very similar results.

[20]Several contributions show that role models are crucial in determining individuals' aspirations (Beaman, Duflo, Pande, & Topalova, 2012; Bernard, Dercon, Orkin, Taffesse, et al., 2014; Lybbert & Wydick, 2016; Macours & Vakis, 2009), while other papers in the education literature illustrate how competition and comparisons might affect students' outcomes and behavior (Azmat & Iriberri, 2010; Bursztyn, Egorov, & Jensen, 2019; Bursztyn & Jensen, 2015; Jonsson & Mood, 2008).

Therefore, I also include classroom fixed effects in my estimations.

Besides controlling for classroom FE, I also perform a simulation where I randomly reshuffle students' friends and estimate peer effects using these new connections instead of students' real friends. Since such estimation will capture the impact that colleagues who are not friends have in students' aspirations, it should render smaller coefficients, non-significant in most of the time. Indeed, I find that less than 10% of the coefficients coming from such a simulation are equal to or greater than the results considering the connections with real friends.

Two other discussions are necessary here. First, the aspiration equation 2 excludes the homophily variables W_{ij} included in the model of friendship formation 5. While the nature of these variables is different since they vary at the dyad level instead of at the individual level (such as the variables included in model 2), one may wonder whether identification remains robust to the inclusion of variables that are functions of both own and peers characteristics in the main equation. The hypothesis underlying such exclusion is that students' aspirations are not influenced by how similar they are to their friends. While this is a reasonable assumption, it might be that the extent to which students are similar to their friends in terms of gender or race helps them to form their identity in school (Akerlof & Kranton, 2000, 2002), which might impact their aspirations. Section 4.1 presents an exercise that tests the robustness of my model to the inclusion of interactions of own and peers variables to the set of controls in equation 2.

Second, the main underlying event behind the predictive power of name similarities in friendship formation is students' alphabetical allocation in classes during the 6th grade. A concern here is that something special happens at the beginning of the middle school cycle that influences aspirations later on. For instance, early experiences in the new school, such as being allocated to a particularly motivating teacher or to a particularly disruptive group of peers could have long term effects on aspirations. I test in Section 4.1 how sensitive estimations are to different specifications of my model of link formations.

4 Results

Table 3 presents results of the main estimations.[21] I use different instruments for friends' aspirations to test for the robustness of the results. In columns (1) and (2), friends' aspirations are instrumented by $\hat{G}(W)^2X$, that is, by predicted friends of friends' characteristics. Columns (3) and (4) use $\hat{G}(W)^3X$ - that is, third-order connections[22] - as instruments. Finally, columns (5) and (6) present both $\hat{G}(W)^2X$ and $\hat{G}(W)^3X$ as instruments. Columns (2), (4), and (6) include controls for classroom fixed effects to account for the presence of role models, or competitive dynamics in the classroom, as discussed in section 3.4.

First, it is important to highlight that the instruments are quite strong, ranging from a joint significance (F statistic) of 25.98 to 40.80. Second, looking at the estimations, we see that peer effects on aspirations are positive, significant, and quite sizable. Column (2), my favorite estimation due to the presence of classroom fixed effects and where instruments are strongest, has a coefficient of 0.153, which means that if a student passes from having no nominated friends who aspire to a college degree to having all nominated friends who aspire to it, her probability of aspiring to a college degree increases by 15.3 percentage points (p.p.).

Perhaps passing from having no nominated friends aspiring to go to college to having all friends aspiring to it is a too extreme way of interpreting the results. A better way of interpreting them is to think about the marginal impact of having an extra *aspiring friend* - that is, an extra friend aspiring to go to college. Such an effect will depend on the number of nominated friends. As described in section 2, each student could nominate up to four best friends or colleagues. If a student nominates all four friends, the marginal impact of an extra aspiring friend is about 3.8 p.p..[23] If a student nominates three friends, the marginal impact

[21]For comparative purposes, Table A.2, 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 proposed by König et al. (2018) and used throughout this work. We can see that the OLS estimation is actually smaller than the 2SLS estimation, which may be due to measurement error or to the exclusion bias, discussed before. Importantly, however, the results decrease quite considerably when we compare 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.

[22]Third order connections are the friends of friends of friends.

[23]If a student has four friends, the average of peers' aspirations increases by 0.25 every time an extra friend

of an extra aspiring friend is 5.1 p.p.. If a student nominates two friends, the marginal impact is 7.65 p.p.. Finally, if a student nominates only one friend, the marginal impact will be 15.3 p.p. - naturally, the same as passing from having no friend aspiring to college to have all friends aspiring to it.

As shown in Table 1, the average number of nominated friends is about 2, so the impact of an extra aspiring friend for the average student is 7.65 p.p. increase in her likelihood of aspiring to go to college. Since on average 68% of students aspire to a college degree, this translates into an increase of 11.25% on the average aspirations. Such an impact is far from being negligible. It is 82% larger than the impact of a student's father having a secondary school or more, and about 30% larger than the impact of moving students one standard-deviation in the reading performance distribution.

4.1 Robustness check

As discussed in section 3.4, a potential threat for identification is the fact that some students in the data did not nominate all of their friends. If this is the case, the model of network formation might not be correctly estimated, and some excluded instruments used in the estimation of peer effects might be endogenous. I've shown that this is not the case for most of my sample since only 20% if students nominated the maximum allowed number of friends. Yet, I check whether possible ceiling effects in the nomination of friends could be driving the results. Column (1) of Table A.3 presents 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 nominated three friends or less. In this restricted sample, it is possible to map all students' connections with more precision, without incurring the risk of having missing links. The results are remarkably similar to the ones of Table 3.

Columns (2), (3), and (4) of Table A.3 present other robustness checks. The estimation in column (2) adds interactions between students' characteristics, such as gender and race,

aspires to go to college. If we multiply this increase by the coefficient of peers' aspirations - which is 0.153 - we get to the marginal effect of 0.038, or 3.8 p.p..

and their friends' average characteristics in the set of controls. The intuition behind such an inclusion is to check whether the main estimations are robust to the inclusion in the main model of homophily variables used to estimate link formation. Again, the results are fairly similar to the ones presented in Table 3.

Columns (3) and (4) test the robustness of the estimations for different specifications in the model of link formation. More specifically, column (3) presents estimations where the predicted adjacency matrix $\hat{G'}^2 X$ comes from a model that does not consider name similarities among students to predict friendship links. Column (4), in turn, presents estimations where the predicted adjacency matrix $\hat{G''}^2 X$ comes from a model that only considers name similarities among students to predict friendship links. In both cases, the point estimates are larger than the ones in table 3, and the instruments seem to be weaker. However, the results are quite similar when comparing columns (4) and (5), which is quite reassuring since it indicates that my results are not being driven only due to specific events that happened to students at the beginning of middle school, as discussed in section 3.4.

4.2 Heterogeneous impacts

Table 4 presents estimations considering heterogeneous characteristics of students regarding some of their demographics and socioeconomic status. Each of the variables in the columns of the table is interacted with friends' aspirations. Hence, column (1) shows heterogeneous exercises for boys and girls, column (2) presents these exercises for non-white and white students, and columns (3) and (4) show the results for students with less/more educated parents (mother in column (3) and father in column (4)).

As we can see, it does not look like a specific group of students is more or less impacted by their friends' aspirations. Such an influence seems to be very homogeneous across different students.

5 Discussion about possible mechanisms

There are at least three mechanisms that could be driving the results. The first is information diffusion. Students might exchange facts and impressions about college returns (both pecuniary and non-pecuniary), as well as about how to get into college - such as application process, fellowships, etc. Such a set of information might help them form their expectations about the benefits of attending a college and its feasibility.[24] A second mechanism behind the results is conformity to social norms. Students might either be influenced by their friends to comply with social norms that hinder their aspirations or see college aspirations itself as a social norm to which they decide to comply. As shown in (Bursztyn & Jensen, 2017), there is a burgeoning literature on how the presence of social norms and social pressure change individuals' behavior.

Finally, a third mechanism is that friends who aspire to go to college put more effort into their school activities and influence students to increase their effort as well. This, in turn, can make students revise their own aspirations and expectations. High achieving friends might also be different in their socio-emotional skills.[25] If they influence their friends' socio-emotional skills, this could also lead eventually to a shift in the aspirations of these friends.

Unfortunately, I cannot access all kinds of information that students have about college returns or how to get into college. However, it is possible to use their perceived college returns to get a sense of whether they are exchanging information regarding college. If students consider such returns when forming their aspirations and, at the same time, inform each other about these returns, then information diffusion might be a mechanism in place. It is also possible to investigate whether "bad" social norms - such as the fear of being stigmatized as a nerd or peer pressure to work - are diffused among friends. If students decide to comply with such norms, they will most likely lower their aspirations, so the

[24] The works of Jensen (2010), Belfield, Boneva, Rauh, and Shaw (2019) and Peter, Spiess, and Zambre (2018), for instance, show how students' perceptions on college returns influence their aspirations, as well as informing students about possibilities to pursue a college degree.

[25] There is an important stream of literature showing how socio-emotional skills are related to performance. For a review, see the work of Almlund, Duckworth, Heckman, and Kautz (2011).

spread of such norms could be a mechanism for peer effects on aspirations.

To test that, Table 5 presents exercises that estimate peer effects on perceived college returns, on the fear of nerd stigma, and on peer pressure to work. The methodology implemented in these estimations is the same as the one described in section 3. The difference is that now the dependent variable and the endogenous social effect are not college aspirations and peers' college aspirations, respectively, but each variable in the table columns.

The table shows that, while friends do not seem to impact perceived college returns, they do seem to have an influence on students willingness to comply with social norms: for the average student with two friends, an extra friend who sees the fear of a nerd stigma as an impediment to keep studying increases the likelihood of that student feeling the same way in about 8.3 p.p.. The impact of an extra friend who sees peer pressure to work as an impediment to keep studying is about 12.05 p.p. for the average student. Therefore, friends seem to impact more the spread of social norms than of information - or at least information about pecuniary college returns.

I also look at proxies for students' effort in school - such as whether they study math more than 30 minutes per day and whether they had more than 95% of class attendance in both reading and math in that school year - and at students performance in reading and math. While these exercises are interesting in their own right - especially those looking at peer effects on students' performance - they might also help inform whether peer effects on aspirations are coming indirectly through peers' impact on school attitudes and outcomes.

Table 6 shows these exercises. We see that friends do seem to impact students' effort: as shown in column (1), if students spend more time studying Math, this also influences their friends to spend time in this activity. In column (2), we can see that friends' attendance to classes also influences students' own attendance. However, the results in columns (3) and (4) indicate that friends do not impact performance. This last result is quite interesting since it adds to the long literature on peer effects on students' performance, which usually finds at least modest peer effects in primary and secondary education (see Sacerdote, 2011, for a review).

Table 7 finally presents estimation of peer effects on socio-emotional skills. These results should be read carefully since the instruments are very weak. This might be what is driving the large and significant results in column (3). Overall, it looks like there are no peer effects on socio-emotional skills.

6 Peers' aspirations and future outcomes

Once peers' impact on students' aspirations is verified, I investigate whether such influence spillovers to students' outcomes in school.

I have shown in section 2.3 how students' aspirations are associated with school outcomes, such as the likelihood of dropping out of school and having a normal school path. As highlighted before, such associations cannot have a causal interpretation. However, it is possible to use the previous methodology to infer the causal impact that friends' aspirations have on school outcomes. There are several reasons for such an effect to emerge. First, as shown in the main exercises, friends' aspirations influence students' own aspirations, which might change their future outcomes. Second, even after considering students' own aspirations, having aspiring friends might help the studying environment and increase students' performance - since these friends are more invested themselves in school activities -, and might decrease the presence of social norms which curb the willingness to keep studying. Moreover, if aspiring friends tend to go further in their studies, this might prevent students from dropping out or being retained, simply because they want to be around their friends.

Table 8 presents the results of five estimations that measure how peers' aspirations influence students' future outcomes in school. Column (1) shows estimations on students' dropout during secondary school; column (2) shows estimations on the likelihood of having a normal school path (that is, being at the 12^{th} grade in 2014) for those who did not drop out from school; column (3) shows estimations for a binary variable that indicates whether the student attended to at least 95% of classes in both reading and math in 2013; and columns (4) and (5) present estimations on students performance in reading and math tests that they took in the last year of secondary school. One can see that even though peers' aspirations does not have an impact on students' future performance, normal school path, or class attendance, it does decrease their likelihood of dropping out of school: for the average student with two friends, an extra aspiring friend decreases the likelihood of dropping out of school by 4.75 p.p..

7 Conclusion

This work overcomes primary challenges concerning the estimation of peer effects and investigates the influence that friends' aspirations have on one's own aspirations and future school outcomes. I first explore students' similarities in pre-determined characteristics 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' aspirations. This identification strategy overcomes both the problem of endogenous formation of friendships and the reflection problem, largely discussed in the literature of peer effects estimation.

Results show that an extra friend aspiring to go to college not only increases students' own aspirations towards going to college but also decreases students' likelihood of dropping out of school. This brings valuable insights into educational policymaking in developing countries. First, peer effects on aspirations might be a mechanism explaining peer effects on school dropout, as shown by Evans, Oates, and Schwab (1992) and Cipollone and Rosolia (2007), for instance. Aspiring students are less likely to drop out of school. At the same time, they also decrease their friends' likelihood of dropping out. Hence, part of the results of peer effects on school dropout could be coming trough peer effects on aspirations.

Second, I find that diffusion of information does not seem to be the mechanism of peer effects on students' aspirations. What does seem to matter as a mechanism here is social norms, and the need students feel to conform to them. Increasing the number of students who aspire to a college degree might lead to a change in certain harmful social norms, such as the stigmatization of students who study hard, which in turn will allow for the realization of students' true educational potential.

Finally, several works show how some educational interventions increase students' aspirations (Carlana, La Ferrara, & Pinotti, 2015; Chiapa, Garrido, & Prina, 2012; Ross, 2017). My results highlight that any impact coming from these interventions spillovers to peers, which should be considered in cost-benefit analysis.

Future works should focus on peer effects in aspirations for contexts different from education attainment. For instance, opportunities in the labor market have been shown to increase career aspirations, especially for women (Jensen, 2012). However, peer effects in such a setting might be different from the one found in this work since now one should also consider the presence of competitions for jobs and work hierarchical relations.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753.

Akerlof, G. A., & Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of economic literature*, *40*(4), 1167–1201.

Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. D. (2011). *Personality psychology and economics* (Tech. Rep.). National Bureau of Economic Research.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108.

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

Austen-Smith, D., & Fryer Jr, R. G. (2005). An economic analysis of "acting white". *The Quarterly Journal of Economics*, *120*(2), 551–583.

Azmat, G., & 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.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* Macmillan.

Beaman, L., Duflo, E., Pande, R., & 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., & Shaw, J. (2019). What drives enrolment gaps in further education? the role of beliefs in sequential schooling decisions. *Economica*.

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.

Bogliacino, F., & Ortoleva, P. (2013). The behavior of others as a reference point. *Columbia Business School Research Paper*(13-55).

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

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

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

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

Bursztyn, L., & 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., & Fafchamps, M. (2016). *Exclusion bias in the estimation of peer effects* (Tech. Rep.). National Bureau of Economic Research.

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

Campbell, E. Q., & Alexander, C. N. (1965). Structural effects and interpersonal relationships. *American Journal of Sociology*, *71*(3), 284–289.

Carlana, M., La Ferrara, E., & Pinotti, I. P. (2015). Shaping educational careers of immigrant children: Motivation, cognitive skills and teachersâ beliefs.

Carrell, S. E., & Hoekstra, M. L. (2010). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, *2*(1), 211–28.

Case, A. C., & Katz, L. F. (1991). *The company you keep: The effects of family and* *neighborhood on disadvantaged youths* (Tech. Rep.). National Bureau of Economic Research.

Chiapa, C., Garrido, J. L., & Prina, S. (2012). The effect of social programs and exposure to professionals on the educational aspirations of the poor. *Economics of Education Review*, *31*(5), 778–798.

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

Cohen, J. (1983). Peer influence on college aspirations with initial aspirations controlled. *American Sociological Review*, 728–734.

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

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

De Giorgi, G., Pellizzari, M., & 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., & McIntosh, S. (2018). The causal effect of secondary school peers on educational aspirations. *CVER Discussion Paper Series-ISSN 2398-7553*.

Duncan, O. D., Haller, A. O., & Portes, A. (1968). Peer influences on aspirations: A reinterpretation. *American Journal of Sociology*, *74*(2), 119–137.

Evans, W. N., Oates, W. E., & Schwab, R. M. (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy*, *100*(5), 966–991.

Filmer, D., & Rogers, H. (2018). Learning to realize educationâs promise. *World Development Report. The World Bank*.

Fruehwirth, J. C., & Gagete-Miranda, J. (2019). Your peers' parents: Spillovers from parental education. *Economics of Education review*.

Gaviria, A., & Raphael, S. (2001). School-based peer effects and juvenile behavior. *Review* *of Economics and Statistics*, *83*(2), 257–268.

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

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

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

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

Guryan, J., Kroft, K., & 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.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of applied econometrics*, *18*(5), 527–544.

Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation* (Tech. Rep.). National Bureau of Economic Research.

Janzen, S. A., Magnan, N., Sharma, S., & 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.

Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? experimental evidence from india. *The Quarterly Journal of Economics*, *127*(2), 753–792.

Johnsson, I., & Moon, H. R. (2017). Estimation of peer effects in endogenous social networks: control function approach. *USC-INET Research Paper*(17-25).

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

Kao, G., & Tienda, M. (1998). Educational aspirations of minority youth. *American journal* *of education*, *106*(3), 349–384.

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

König, M. D., Liu, X., & 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.

Lavy, V., & Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, *3*(2), 1–33.

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

Lybbert, T. J., & Wydick, B. (2016). *Hope as aspirations, agency, and pathways: poverty dynamics and microfinance in oaxaca, mexico* (Tech. Rep.). National Bureau of Economic Research.

Macours, K., & Vakis, R. (2009). *Changing households' investments and aspirations through social interactions: evidence from a randomized transfer program*. The World Bank.

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

Marotta, L. (2017). Peer effects in early schooling: Evidence from brazilian primary schools. *International Journal of Educational Research*, *82*, 110–123.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, *27*(1), 415–444.

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

Niño, M. D., Cai, T., & 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. (2017). *Education at a glance 2017: Oecd indicators.* OECD Publishing, Paris.

Peter, F., Spiess, C. K., & Zambre, V. (2018). Informing students about college: An efficient way to decrease the socio-economic gap in enrollment: Evidence from a randomized field experiment.

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

Rosenberg, M. (1986). *Conceiving the self.* RE Krieger.

Ross, P. H. (2017). *The aspirations gap and human capital investment: Evidence from indian adolescents* (Tech. Rep.). mimeo, Boston University.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, *80*(1), 1.

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* (Vol. 3, pp. 249–277). Elsevier.

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

Sewell, W. H., & Hauser, R. M. (1975). Education, occupation, and earnings. achievement in the early career.

Sewell, W. H., & Shah, V. P. (1968). Social class, parental encouragement, and educational aspirations. *American journal of Sociology*, *73*(5), 559–572.

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

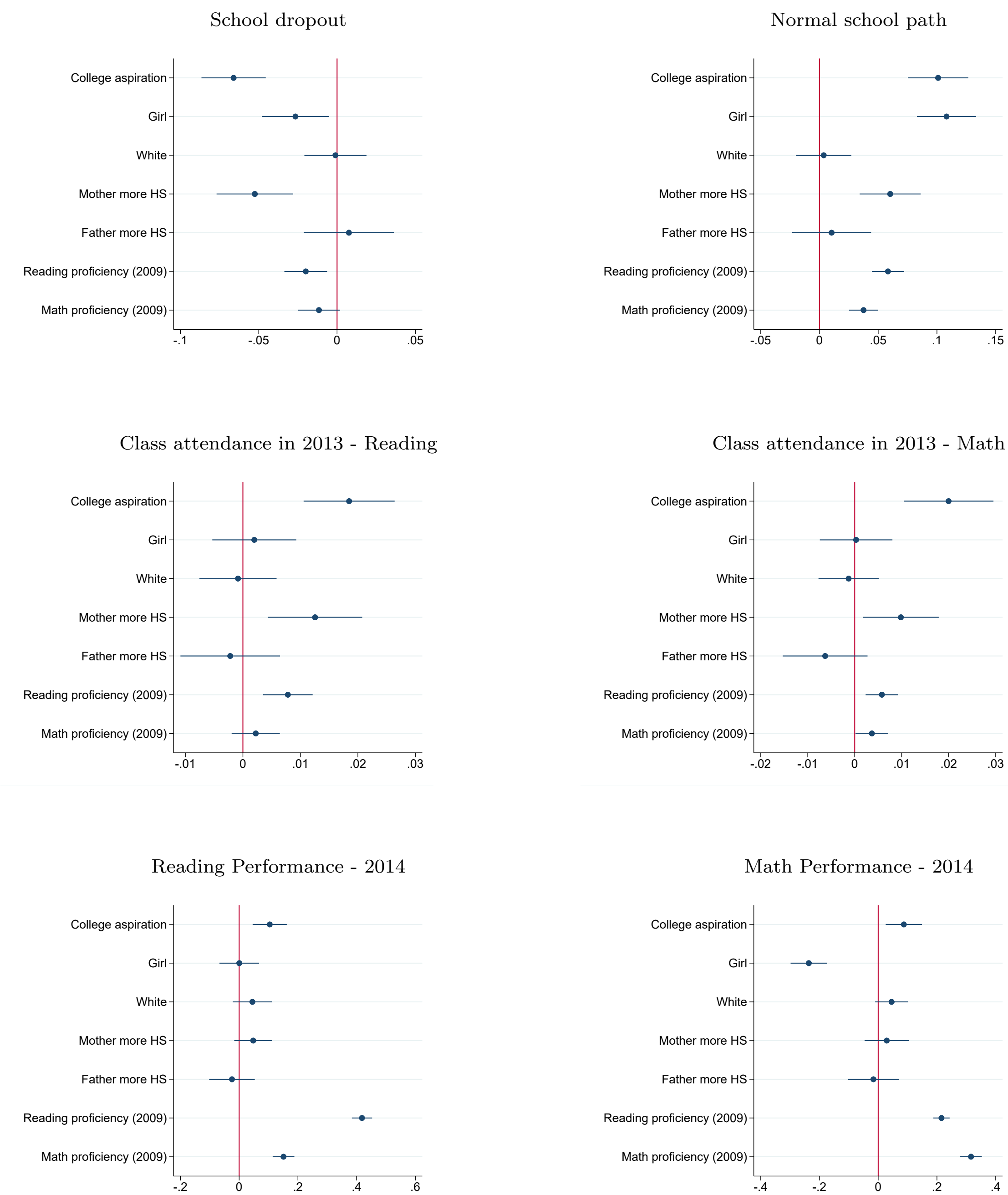
Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of personality*, *72*(2), 271–324.

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

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

8 Tables & Figures

Figure 1: College aspirations and future outcomes in school



Note: (i) results from OLS estimations with school fixed effects; (ii) all estimations also control for: home-ownership, internet at home, and number of lavatories at home; (iii) Reported 99% confidence intervals are based on standard errors clustered at the school level.

Table 1: Descriptive Statistics

	All		Coll. aspiration=1		Coll. aspiration=0	
	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error
Own characteristics						
College aspirations	0.68	0.46	1.00	0.00	0.00	0.00
Girl	0.49	0.50	0.56	0.50	0.36	0.48
White	0.33	0.47	0.35	0.48	0.30	0.46
Mother education: more than HS	0.24	0.43	0.26	0.44	0.21	0.41
Father education: more than HS	0.22	0.41	0.24	0.43	0.18	0.38
Reading proficiency (2009)	0.00	1.00	0.14	0.99	-0.30	0.94
Math proficiency (2009)	-0.00	1.00	0.10	1.00	-0.23	0.97
Nominated friends	2.02	1.41	2.16	1.38	1.72	1.42
Friends' characteristics						
College aspirations	0.59	0.42	0.64	0.40	0.47	0.43
Girl	0.43	0.45	0.48	0.45	0.31	0.43
White	0.27	0.34	0.30	0.34	0.22	0.32
Mother education: more than HS	0.21	0.30	0.22	0.31	0.17	0.29
Father education: more than HS	0.19	0.29	0.21	0.30	0.15	0.27
Math proficiency (2009)	0.08	0.66	0.12	0.67	-0.02	0.65
Reading proficiency (2009)	0.10	0.67	0.15	0.68	-0.01	0.63
Nominated friends	1.93	1.35	2.07	1.31	1.63	1.38
Observations	6076		4157		1919	
Number of schools	85					

Note: (i) "College aspiration" is a binary variable that takes value equal 1 if the student indicates that he/she wants to keep studying up to college; Math and Language proficiency are normalized with Mean=0 and SD=1; (ii) "Nominated friends" is the number friends in the 9^{th} grade nominated by the student.

Table 2: Probability of Forming a Friendship Link

	(1) Raw	(2) Odds Ratio
$\mathbf{1}[\boldsymbol{x_i} = \boldsymbol{x_j}]$		
Gender	1.493***	4.452***
	(0.049)	(0.219)
Race-white	0.132***	1.141***
	(0.024)	(0.027)
Race-black	0.158***	1.171***
	(0.045)	(0.053)
First letter of name	0.360***	1.433***
	(0.051)	(0.073)
x_j **characteristics**		
Girl	0.164***	1.179***
	(0.035)	(0.041)
Race-White	0.057**	1.059**
	(0.024)	(0.025)
Race-Black	0.100**	1.105**
	(0.041)	(0.045)
N (potential links)	524,724	524,724

Note: (i) This table shows the estimation of a conditional logistic regression model to predict the likelihood that a student i will send a friendship tie to another student j in the 9^{th} of the same school; "Raw" is the raw coefficient coming from the model; the estimation controls for i's fixed effects; (ii) Standard errors clustered at school level shown in parenthesis; (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table 3: Peer effects on aspirations (N=6,076)

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: College Aspirations					
Friends' college aspirations	0.157***	0.153***	0.162***	0.156**	0.165***	0.158***
	(0.052)	(0.054)	(0.061)	(0.061)	(0.045)	(0.049)
Student's characteristics						
Girl	0.125***	0.120***	0.125***	0.120***	0.125***	0.120***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
White	0.009	0.005	0.009	0.005	0.009	0.005
	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)	(0.013)
Mother education: more than HS	-0.015	-0.019	-0.015	-0.019	-0.015	-0.019
	(0.015)	(0.016)	(0.015)	(0.016)	(0.015)	(0.016)
Father education: more than HS	0.046***	0.042***	0.045***	0.042***	0.045***	0.042***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Reading proficiency (2009)	0.056***	0.059***	0.056***	0.059***	0.056***	0.059***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Math proficiency (2009)	0.024***	0.024***	0.024***	0.024***	0.024***	0.023***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Father works	0.062***	0.060***	0.062***	0.060***	0.062***	0.060***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Friends' characteristics						
Girl	-0.006	0.002	-0.007	0.001	-0.008	0.000
	(0.020)	(0.020)	(0.021)	(0.021)	(0.019)	(0.020)
White	0.024	0.015	0.024	0.014	0.024	0.014
	(0.020)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)
Mother education: more than HS	-0.032	-0.025	-0.032	-0.025	-0.032	-0.025
	(0.019)	(0.021)	(0.019)	(0.021)	(0.019)	(0.021)
Father education: more than HS	0.035	0.029	0.034	0.029	0.034	0.029
	(0.021)	(0.023)	(0.021)	(0.023)	(0.021)	(0.023)
Reading proficiency (2009)	0.003	0.007	0.003	0.006	0.003	0.006
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Math proficiency (2009)	0.028**	0.029**	0.027**	0.029**	0.027**	0.029**
	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)	(0.013)
Father works	-0.069***	-0.071**	-0.070**	-0.072**	-0.071***	-0.072***
	(0.026)	(0.027)	(0.028)	(0.029)	(0.024)	(0.026)
Instruments	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^3X$	$\hat{G}^3X$	$\hat{G}^2X$, $\hat{G}^3X$	$\hat{G}^2X$, $\hat{G}^3X$
IVs' joint significance	36.734	40.801	30.349	31.902	25.989	29.154
Control for school FE	Yes	Yes	Yes	Yes	Yes	Yes
Control for classroom FE	No	Yes	No	Yes	No	Yes

Note: (i) This table shows estimations of models like the one described in equation 2, where friends' college aspirations are instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$), or by the predicted friends-of-friends-of-friends' characteristics ($\hat{G}^3X$), or by both ($\hat{G}^2X$, $\hat{G}^3X$); (iii) Standard errors clustered at school level shown in parenthesis; (iv) All regressions include controls for the following characteristics (for both students and their friends): gender, race, math and reading performance in 2009, parents' education, father working status, homeownership, internet at home, and number of lavatories at home; (v) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table 4: Heterogeneous impacts - (N=6,076)

	(1)	(2)	(3)	(4)
	Dependent variable: college aspirations			
	Boys	Non-white	Mother less HS	Father less HS
Friends' aspirations	0.129*	0.112*	0.138**	0.132**
	(0.068)	(0.060)	(0.062)	(0.065)
Friends' aspiration x Variable in column	0.051	0.048	0.018	0.029
	(0.051)	(0.039)	(0.055)	(0.049)
Joint significance of Friends' aspirations				
P-value	0.003	0.017	0.021	0.017
Instruments	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$
IVs' joint significance	24.903	25.716	29.954	25.436
Control for own characteristics	Yes	Yes	Yes	Yes
Control for friends' characteristics	Yes	Yes	Yes	Yes

Note: (i) This table shows estimations of models like the one described in equation 2, where friends' college aspirations are instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) Friends' aspirations are interacted with each characteristic in the columns and such interaction is instrumented by $\hat{G}^2X$ interacted with this characteristic as well; (iii) Standard errors clustered at school level shown in parenthesis; (iv) All regressions include school and classroom FE and the following characteristics (for both students and their friends): gender, race, math and reading performance in 2009, parents' education, father working status, homeownership, internet at home, and number of lavatories at home; (v) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table 5: Peer effects on college return and on compliance with social norms (N=6,076)

	(1)	(2)	(3)
Dependent variable:	Perceived college returns	Fear of nerd stigma	Peer pressure to work
Endogenous Social Effects	-0.008	0.166*	0.241***
	(0.039)	(0.089)	(0.081)
Instruments	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$
IVs' joint significance	109.204	14.269	19.417
Mean Dep. Var.	0.612	0.256	0.298
Control for own characteristics	Yes	Yes	Yes
Control for friends' characteristics	Yes	Yes	Yes

Note: (i) This table shows estimations of models as the one described in equation 2, where the endogenous social effects (that is, friends' perceived college return in column (1), friends' fear of nerd stigma in column (2), and friends' feeling peer pressure to work in column (3)) are instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) Standard errors clustered at school level shown in parenthesis; (iii) All regressions include school and classroom FE and the following characteristics (for both students and their friends): gender, race, math and reading performance in 2009, parents' education, father working status, homeownership, internet at home, and number of lavatories at home; (iv) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table 6: Peer effects on school effort and school performance

Dependent variable:	(1) Studies +30min of math/day	(2) More than 95% of attendance (2011)	(3) Reading performance (2011)	(4) Math performance (2011)
Endogenous Social Effects	0.138*	0.459**	-0.166	-0.368
	(0.081)	(0.187)	(0.238)	(0.492)
Instruments	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$
IVs' joint significance	17.891	2.950	3.769	1.635
Mean Dep. Var.	0.396	0.104	-0.000	-0.000
Control for own characteristics	Yes	Yes	Yes	Yes
Control for friends' characteristics	Yes	Yes	Yes	Yes

Note: (i) This table shows estimations of models like the one described in equation 2, where the endogenous social effects (that is, friends' study +30min of math/day in column (1), friends' attended to more than 95% of classes in 2011 in column (2), and friends' performance in reading and math in 2011 in column (3) and (4), respectively) are instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) Standard errors clustered at school level shown in parenthesis; (iii) All regressions include school and classroom FE and the following characteristics (for both students and their friends): gender, race, math and reading performance in 2009, parents' education, father working status, homeownership, internet at home, and number of lavatories at home; (iv) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table 7: Peer effects on socio-emotional skills

Dependent variable:	(1) Self-esteem	(2) Self-efficacy	(3) Self-control	(4) Agreeableness	(5) Rapport with peers	(6) Locus of control
Endogenous Social Effects	-0.104	0.300	1.269***	-0.008	0.083	-0.509
	(0.396)	(0.274)	(0.409)	(0.230)	(0.220)	(0.316)
Instruments	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$
IVs' joint significance	1.336	1.275	1.259	2.429	2.458	1.989
Control for own characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Control for friends' characteristics	Yes	Yes	Yes	Yes	Yes	Yes

Note: (i) This table shows estimations of models like the one described in equation 2, where the endogenous social effects (that is, friends' socio-emotional skills) are instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) Standard errors clustered at school level shown in parenthesis; (iii) All regressions include school and classroom FE and the following characteristics (for both students and their friends): gender, race, math and reading performance in 2009, parents' education, father working status, homeownership, internet at home, and number of lavatories at home; (iv) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table 8: Friends' aspirations and students' future outcomes

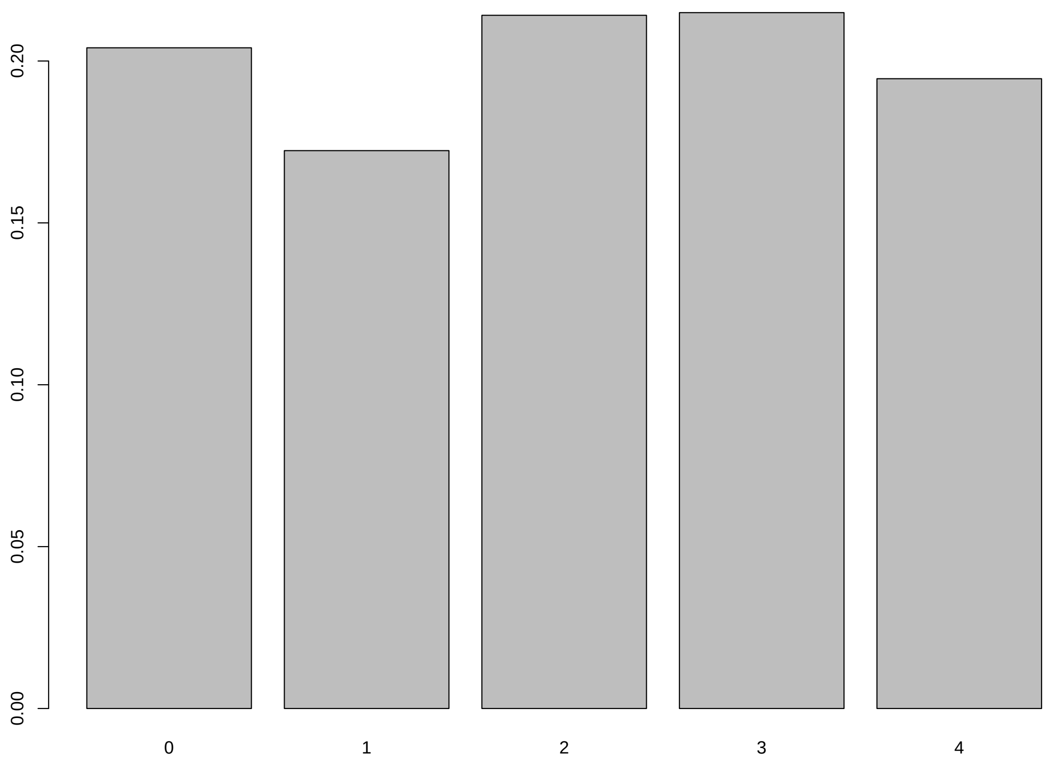
	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Dropout	Normal school path	Attendance 2013	Reading 2014	Math 2014
Friends' college aspiration	-0.095**	0.037	0.024	0.171	0.199
	(0.043)	(0.056)	(0.040)	(0.183)	(0.171)
Instruments	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^2X$
N	6076	5053	4857	3200	3200
IVs' joint significance	40.801	40.416	39.850	48.458	48.458
Mean Dep. Var.	0.168	0.766	0.121	-0.004	-0.011
Control for own characteristics	Yes	Yes	Yes	Yes	Yes
Control for friends' characteristics	Yes	Yes	Yes	Yes	Yes

Note: (i) This table shows estimations of models like the one described in equation 2, where friends' college aspirations are instrumented by the predicted friends-of-friends' characteristics ($\hat{G}^2X$); (ii) The sample size changes depending on the estimation due to students dropout and to whether students took the standardized tests in 2014; (iii) Standard errors clustered at school level shown in parenthesis; (iv) All regressions include school and classroom FE and the following characteristics (for both students and their friends): gender, race, math and reading performance in 2009, parents' education, father working status, homeownership, internet at home, and number of lavatories at home; (v) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

A Appendix

A.1 Additional tables and figures

Figure A.1: Out degree distribution



Note: Each student was asked to nominate at most four of their best friends or colleagues at school. This graph shows the distribution of the number of nominated friends by each student.

Table A.1: 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 letter of name	0.009***	0.008***
	(0.003)	(0.003)
Gender	0.002	0.002
	(0.002)	(0.002)
Race	-0.002	-0.000
	(0.002)	(0.001)
Father finished HS	0.002	0.003
	(0.003)	(0.002)
Father has college degree	-0.000	0.006
	(0.006)	(0.005)
Mother finished HS	0.001	-0.001
	(0.003)	(0.003)
Mother has college degree	-0.006	-0.006
	(0.005)	(0.004)
N	640,826	640,826
Control for school FE	No	Yes

Note: (i) This table presents estimations of 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 6th grade. The observational unit of this table are dyads of students. The dependent variable is a binary variable that takes value one if student i was enrolled in the same classroom of student j in the 6th 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 school level shown in parenthesis; (iii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$.

Table A.2: Peer effects on aspirations - comparing OLS, 2SLS, and 3SLS

	(1)	(2)	(3)	(4)
	Dependent Variable: College aspirations			
Friends' aspirations College aspiration	0.069^{***}	0.201^{**}	0.157^{***}	0.153^{***}
	(0.022)	(0.085)	(0.052)	(0.054)
Model	OLS	IV: G^2X	IV: $\hat{G}^2X$	IV: $\hat{G}^2X$
IVs' joint significance		24.548	36.734	40.801
Control for own characteristics	Yes	Yes	Yes	Yes
Control for friends' characteristics	Yes	Yes	Yes	Yes
Control for school FE	Yes	Yes	Yes	Yes
Control for classroom FE	No	No	No	Yes

Note: (i) Standard errors clustered at school level shown in parenthesis;; (ii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (iii) All regressions include the following controls (both for students and their friends): gender, race, math and reading proficiency, father working status, internet at home, and number of lavatories at home.

Table A.3: Robustness Check

	(1)	(2)	(3)	(4)
	Dependent Variable: College aspirations			
Friends' aspirations	0.150^{**}	0.159^{***}	0.395^{***}	0.323^{***}
	(0.059)	(0.056)	(0.093)	(0.103)
Instrument	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G'}^2X$	$\hat{G''}^2X$
N	4893	6075	6075	6075
IVs' joint significance	36.101	40.588	16.505	20.501
Maximum out-degree≤ 3	Yes	No	No	No
Control for homophily	No	Yes	No	No

Note: (i) "Maximum out-degree≤ 3" means that only the sub-sample of students who nominated at most three friends was considered; (ii) "Control for homophily" means that the estimation controls for interactions between student's characteristics such as gender and race with the average characteristics of their friends; (iii) $\hat{G'}^2X$ is the adjacency matrix coming from a model of link formation that does not include name similarities in the set of controls; (iv) $\hat{G'}^2X$ is the adjacency matrix coming from a model of link formation that only includes name similarities in the set of controls; (v) Standard errors clustered at school level shown in parenthesis; (vi) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (vii) All regressions include school and classroom FE and the following controls: gender, race, math and reading proficiency, father working status, home ownership, internet at home, and number of lavatories at home.