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

Jessica G. Miranda†

2019

Abstract

Aspiration is 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. The present work asks whether peers play a role in such a process. I use novel data on Brazilian students' network, matched with administrative data, and investigate whether students' college aspiration spills over to their friends. The employed methodology acknowledges that social cliques are formed endogenously and addresses this challenge by modeling friendship formation based on homophily in predetermined characteristics and on students' as-good-as-random chances of interaction. Using the predicted adjacency matrix, I explore network structures and use friends of friends' characteristics as instruments for friends' aspiration. The results show evidence of positive, significant, and quite large peer effects on aspiration, which are driven mostly by boys, and/or students with less educated fathers. Peers' aspirations also influence students' likelihood of finishing high school. Compliance with social norms seems to play a role in explaining the impact.

*I thank the valuable comments of Marcos A. Rangel, Paolo Pinotti, Eliana La Ferrara, Massimo Anelli, Jane Cooley Fruehwirth, Matias Busso, Eric Bettinger, and Armin Falk. I am also thankful for the feedback from participants of the Summer School on Socioeconomic Inequality in Bonn (2018), of the 9^{th} Alp-Pop Conference, of the Workshop on Poverty and Inequality (X RIDGE Forum), and of seminars at Bocconi University, the Sanford School of Public Policy at Duke, and the Duke Network Analysis Center.

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

1 Introduction

An important feature of human nature is the ability that people (some more than others) have to forego small luxuries in the present and to make costly investments for the future. Sociologists, psychologists, and economists have investigated for some time now what are the drivers of such behavior, especially among the poor, for whom investment costs are usually higher (Banerjee and Duflo, 2011). What motivates, for instance, a student to keep going to school even when many of her colleagues drop-out, may be working and contributing to their families' income? Although there are several answers to this question, many of them touch an important characteristic of people's personality: their capacity to aspire to a better standard of living (Appadurai, 2004). Indeed, as long as individuals do not perceive their aspirations as something unattainable, the more they aspire to, the harder they work and the more they invest in the present (Dalton et al., 2016; Genicot and Ray, 2017).

The contributions to the theoretical literature argue that individual aspirations emerge in social contexts, through individuals' comparisons with similar others (Appadurai, 2004; Falk and Knell, 2004; Ray, 1998, 2006; Genicot and Ray, 2017). This important social element of aspirations construction calls for investigations on how exactly peers influence an individual's level of aspiration. Some have found that peers' socioeconomic status is associated with individuals' aspirations (Stutzer, 2004; Macours and Vakis, 2009; Janzen et al., 2017; Tanguy et al., 2014). However, an important question still needs further investigation: do peers' aspirations influence one's own aspiration, over and above socioeconomic considerations? That is, after controlling for socioeconomic status, do peers still influence individuals' aspirations, through their own level of aspiration?

The present work investigates peer effects on students' college aspirations - that is, students willingness to pursue a college degree. In order to do so, I rely on a unique social networks data collected from middle school students in Brazil to address the main challenges that emerge in the identification of peer effects. Differently from standard linear-in-mean models, this work does not assume that all individuals in a students' reference group are equally connected or have the same influence on each other. Instead, it acknowledges that individuals in social networks are idiosyncratically connected to each other, taking this into consideration in the identification strategy (as in König et al. (2018) and Santavirta and Sarzosa (2019)).

An advantage of my data is the possibility to link it with administrative data. I explore this to model friendship formation based on homophily in predetermined characteristics and on the as-good-as-random allocation of students into classes when first enrolling at the school. Next, based on the predicted adjacency matrix, the identification strategy uses friends of friends' characteristics as instrumental variables for friends' aspiration (as in Bramoullé et al. (2009) and De Giorgi et al. (2010)). It also uses network fixed effects and a broad set of controls to eliminate other possible correlated effects.

College aspiration is a quite relevant measure of aspiration in the educational scenario of developing countries. On one hand, these countries have 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 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 earning premium among OECD and partner countries. Still, only 15% of the adult population in the country has attained 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.

An important discussion is whether college aspirations lead to aspirations frustrations and consequently to aspirations failures. Genicot and Ray (2017) show in a theoretical model of how individuals' incentives to invest grow with their aspiration level, but only up to a certain point. If aspirations are too distant from one's current outcome, the anticipation that this aspiration will be frustrated makes investment insensitive to it. *Aspirations ratio* - the ratio of aspirations to starting wealth -, therefore, plays a central role in one's futureoriented behavior, influencing her willingness to make costly investments in the present. This "inverse U-shape" relationship between aspirations and investment has been empirically demonstrated in some works (Pasquier-Doumer and Brandon, 2015; Ross, 2017; Janzen et al., 2017). Other works have shown how adjusting individuals' aspirations to their true potential might lead to better outcomes. For instance, decreasing low-achieving students aspirations decrease their likelihood of dropping out of school (Kearney and Levine, 2014; Goux et al., 2014).

Interestingly, pursuing a college degree does not seem something unattainable in developing countries (Graham et al., 2018), especially under the disclosure of information about college funding and scholarships (Bonilla et al., 2016). The survey used in the present study was collected in 2011. Since 2005, low-income students can apply for up to 100% scholarships to enroll in private universities in Brazil. Moreover, in 2010 the federal government expanded students loans for virtually all low-income students in the country[1].

I document that college aspiration is positively associated with students' effort in school - measured by their report on how many hours they usually study, and on whether they take part in study groups - and with their probability of finishing school or passing grades in high school.

I find evidence of positive, significant, and quite large peer effects on aspiration: if a student passes from having no friends aspiring to a college degree to having all her nominated friends aspiring to it, her probability of also aspiring to a college degree increases from 12 p.p to 14 p.p., depending on the specification.. Heterogeneous exercises show that boys and students' from less educated fathers are those more influenced by peers. Compliance to social norms seems to explain at least part of these results.

I finally explore more tangible impacts of peers' aspiration, investigating whether it influences students future outcomes in school, such as retention and dropout. I find that peers' aspiration increases the likelihood of finishing high school by around 11 p.p..

[1]More on the access to higher education in Brazil can be found at Silveira (2018).

This study adds to traditional works in the sociology literature (Sewell and Shah, 1968; Sewell and Hauser, 1975; Kao and Tienda, 1998). Sociologists have long verified the existence of a positive correlation between peers' aspiration and one's own aspiration (see, for instance, the work of Campbell and Alexander (1965); Duncan et al. (1968); Cohen (1983)). Identification issues, however, have prevented these works from establishing causal relationships. In the context of peer effects estimations, correlated effects - socioeconomic background, school quality, or homophily in friendship formation - might deliver 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). Hence, correlational studies say little about the true impact that peers' exert in one's aspiration.

This work also contributes to the literature on peer effects (see Sacerdote (2011) for a review). Most of the works 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 (Hoxby, 2000; Lavy and Schlosser, 2011; Austen-Smith and Fryer Jr, 2005; Carrell and Hoekstra, 2010; Marotta, 2017; Fruehwirth and Miranda, 2019). Other studies focus on peer effects in students' attitudes and behavior, such as substance use, school dropout, and criminal activity (Case and Katz, 1991; Gaviria and Raphael, 2001). Few works, however, investigate peers' influence on students aspirations. Two exceptions are the works of Jonsson and Mood (2008) and Norris (2017). The first shows that having high-achieving peers might depress the average students' desire to attend college. The second shows that peers' attitudes about school influences one's own attitudes [2]. Both studies, however, focus on high school, when very low-aspiring students might have already dropped out.

Finally, this work contributes to the discussion about college choice by low-income stu-

[2] Even though Norris (2017)'s measure of attitude about school has a component of desire to go to college, it also has other components, such as how much students' fell they are part of their current school. Hence, this is not a pure measure of students' aspiration.

dents. Some works have shown the importance of information and encouragement for low-income students to choose a selective college (Hoxby and Avery, 2012; Hoxby et al., 2013). This is also evident in this paper since low-income students are the most influenced by their peers in the decision to go to college.

2 Data and measure of aspiration

The data used in this work come from a survey conducted in 2011 on students from the final grade of 85 state-owned middle schools of Sao Paulo (Brazil). The students answered a comprehensive questionnaire about their personal profile, how happy or satisfied they were with their life, what were their study habits, and what were their expectations. One block of questions mapped students' social networks. They were asked to nominate their four best friends or classmates in their grade (which, in most schools, comprehends more than one classroom). Importantly, it is possible to link the nominated students to school records, 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 important block of questions was dedicated to understanding students' expectations towards their future. One specific question of this block asked until when students would like to keep studying if this choice was entirely up to them, and whether they would like to go to college and pursue a higher education degree or not. I use this question to build my measure of aspiration towards pursuing a college degree, which I call *college aspiration*. 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.

The survey also approached other traits and beliefs that might be associated with college aspiration. First, it had a block of questions asking about students personality. From it, it is possible to identify students soft skills such as "locus of control" - the extent to which individuals attribute current experiences to decisions and attitudes they have taken in the past -, "self-efficacy" - individuals beliefs about their capacity to establish and achieve goals -, "ambition" - which proxies for long-run aspirations such as career success -, and "pro-social behavior" - the tendency to act in a cooperative and unselfish manner. Second, it asked students which probability they attributed for them to find a job in the future if they have a university degree. This question measures students perceptions about college returns, which might influence their aspiration 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' willingness to go to college: (1) their concern about being stigmatized as "nerd" if they put too much effort into school - I call this variable "Fear of nerd stigma"; and (2) the fact that their friends pressure them 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 brings some descriptive statistics coming from this survey and from administrative data, such as students' college aspiration, their demographic and socioeconomic characteristics, and their proficiency in Language and Math in a diagnostic exam - known as Sao Paulo School Performance Assessment System (SARESP, in the Portuguese abbreviation) - applied every year to all state-owned schools in Sao Paulo. The table brings the mean and standard error for all students and also for those who aspire and those who do not aspire to a college degree. Furthermore, it also brings information about students' friends. First, looking at the sample composed of all students, we see that more than 30% of them does not aspire to a college degree. Second, comparing students who aspire to a college with students who do not, it is possible to see that 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 higher achieving students and of having more educated parents - which might exemplify the phenomenon of homophily, that is, people's tendency to befriend with similar others. Homophily is an important confounder in the estimation of peer effects. Section 3 explains how this work overcomes such an issue.

2.1 College aspiration and prediction of students effort and future outcomes

This survey has two great features that allow for tests on the association of college aspirations with both students current effort and future outcomes, which is an important exercise to understand whether such a measure of aspiration goes in the expected direction. First, students' had to indicate how long they studied Math on weekdays and during pre-test periods, and also whether they took part in Math study groups. These answers might be used as proxies for students' effort in school. Second, it is possible to link this survey with administrative data and to recover students' school path - that is, in which grade they were enrolled at each year - before and after they answered the survey. With this information, I calculate the probability that students at the 9^{th} grade in 2011 will finish high school, and their probability of passing grades during high school.

I perform OLS estimations with each of those measures of students' effort and school outcomes as dependent variables, and college aspiration as independent variable, also controlling for students' performance, demographics, socioeconomic status, and school fixed effects. Figures 1 and 2 bring the point estimate and the 95% confidence interval of these estimations[3].

One can see in Figure 1 that college aspiration is the most important predictor of all measures of students' effort in school. It increases the likelihood a student will study for more than 30 min/day during weekdays and pre-test days by roughly 12 p.p. and 11 p.p.,

[3]Besides parents' education and working status, I also use own house, 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, but all their coefficients were either insignificant or very small.

respectively, and it increases the likelihood of taking part in study groups by around 5 p.p.

As described by Figure 2, college aspiration is also relevant for the likelihood of graduating in high school, increasing such likelihood in about 6 p.p. It is also an important predictor of the likelihood of passing grades in high school: it increases the probability of passing grades from 3 p.p. to 6.5 p.p., depending on the grade.

These exercises clearly cannot identify more than simply associations, but they are an important validation of the measure of aspiration used in this work since it correlates as expected with other important measures of students outcomes in school.

3 Identification of peer effects

There are several challenges that one faces 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 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 are 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 and/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 important 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 (McPherson et al., 2001; Moody, 2001; Currarini et al., 2009). 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 (Sacerdote, 2001; Zimmerman, 2003; Cipollone and Rosolia, 2007), other use theoretical models of social interactions (Brock and Durlauf, 2001) or network structures (Calvó-Armengol et al., 2009; Boucher et al., 2014; Liu et al., 2014; Bramoullé et al., 2009; De Giorgi et al., 2010) in order to address both correlated effects and the reflection problem.

To the best of my knowledge, however, few works so far have fully acknowledge the implications of 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 determined jointly. König et al. (2018) and Santavirta and Sarzosa (2019) use a three stage least square (3SLS) strategy where, in the first stage, they model links formation based either on past network structures (König et al., 2018) or predetermined individual characteristics (Santavirta and Sarzosa, 2019) as exclusion restrictions that affect current link formation but do not enter the outcome equation. 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. The present work follows this 3SLS approach. In what follows, I formalize the structural model, the identification issues, and the 3SLS estimation.

3.1 Structural model

Let a student's college aspiration be affected by the mean college aspiration of her friends, her own characteristics such as grades, gender, race, and family background, and by the mean characteristics of her friends. More formally, suppose there is a set of students i, $i = (1...N)$, that belong to network l, $l = (1, ...L)$ [4]. 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 $i \notin F_i$. The structural model is given by[5]

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

where y_{li} is the aspiration level of individual i in network l , which depends on the aspiration level of the friends directly connected to her - the endogenous social effect -, on x_{li}, her own characteristics[6], on the characteristics of her friends - the exogenous social effects - , 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 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:

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

[5] This model reasembles 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.

[6] 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.

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

It is easy to see that the reflection problem emerges because the outcome variable y is present in 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[7]:

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

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.

Nonetheles, 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 on 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 of have similar aspirations towards college.

As in König et al. (2018) and Santavirta and Sarzosa (2019), I will tackle the reflection problem and the endogenous formation of friendship using a 3SLS estimation. The first stage models link formation based on homophily in predetermined characteristics. The second and

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

third stages use the predicted friendship connections delivered by the first stage, and uses friends of friends characteristics as instrumental variables for friends aspirations (reasembling Bramoullé et al. (2009)). In the remaining 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 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 ($\boldsymbol{G^2 X_l}$, $\boldsymbol{G^3 X_l}$, ...) as excluded instruments for $\boldsymbol{Gy}$ and, as so, to identify all the parameters of the structural model 2 [8]. The authors prove that if the diameter[9] 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 structural model is identified[10].

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 $\boldsymbol{\tilde{X}}$ =[$\boldsymbol{Gy_l}$ $\boldsymbol{X_l}$ $\boldsymbol{GX_l}$] 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 $\hat{\varphi}^{2SLS}$ = $\boldsymbol{(\tilde{X}'P\tilde{X})^{-1}\tilde{X'}Py_l}$, where $\boldsymbol{P = S(S'S)^{-1}S}$.

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

[8]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.

[9]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))."

[10]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 good 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 cliques formation. König et al. (2018) and Santavirta and Sarzosa (2019) 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

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

for all $i \neq j$.

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

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

where W_{ij} is the distance in predetermined dyadic characteristics. More specifically, I use individuals similarities on gender and race. I also include binary variables indicating whether individuals i and j were enrolled at the same class when they first enrolled at the state-school in the 6^{th} grade - before 6^{th} grade students were enrolled in municipality schools and their first allocation into classes when arriving at state-schools in the 6^{th} is as good as random. The intuition behind the inclusion of this variable is that, conditional on individuals' own characteristics, sharing the same class when they first arrive at their new school should increase their likelihood of being friends while not *directly* impacting their outcomes (in this case, their aspiration levels). Therefore, this variable can be used as excluded instruments for this first stage of my estimation.

Table A.1 brings the results of such estimation. As it is possible to see, sharing the same class in the first year of secondary school (class in 2008) 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 $\boldsymbol{\tilde{X}}$ =[$\boldsymbol{Gy_l}$ $\boldsymbol{X_l}$ $\boldsymbol{GX_l}$] is instrumented in the second stage by $\boldsymbol{\hat{S}}$ = [$\boldsymbol{X_l}$ $\boldsymbol{GX_l}$ $\boldsymbol{\hat{G}(W)^2X_l}$ $\boldsymbol{\hat{G}(W)^3X_l}$], where $\boldsymbol{\hat{G}(W)}$ is the predicted adjacency matrix from equation 5, $\boldsymbol{\hat{D}(W)}$, 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 some of the identifying assumptions of 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 exclusion restriction of the instruments since all the friends of my friends would also be my friends. This 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) proofs 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. We need 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 important 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 fully model them.

In that sense, my data might suffer from a ceiling effect, since students were only able to nominate 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 3 brings the out-degree distribution, that is, the distribution of the number of friends that each student nominated. Looking at the figure it is possible to see that around 20% of students might be suffering from this ceiling effect since they nominated 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 - more than 50% of students nominated either one, two or three friends so they were not censored in any way. In section 4.1 I present some robustness checks to address potential issues with censored networks[11].

4 Results

Table 2 brings results of the aforementioned estimations. I use different instruments for peers' aspiration to test for the robustness of the results. In columns (1) and (2), peers' aspiration is instrumented by $\hat{G}(W)^2X$, that is, by friends of friends characteristics. Columns (3) and (4) use $\hat{G}(W)^3X$ - that is, third order connections[12] - as instruments. Finally, columns (5) and (6) bring both $\hat{G}(W)^2X$ and $\hat{G}(W)^3X$ as IVs. For comparative purposes, columns (2), (4) and (6) include controls for both students' and friends' soft skills. If homophily was the factor driving such results, the inclusion of soft skills would decrease the coefficient of peers' aspiration, since homophily in personality also drives many link formation. However, if anything, the inclusion of soft skills increase the impact of peers' aspiration, which brings an additional robustness check to the identification strategy.

Looking at the estimations, it is possible to see that peer effects on aspiration are positive, significant, and quite sizable. Column (1), for instance, has a coefficient of 0.124, which means that if a student passes from having no friends who aspire to a college degree to having all friends who aspire to it, her probability of aspiring to a college degree increases by 12.4 p.p.

[11]Figure 3 also shows that around 20% of students did not nominate any friend. This number is surprisingly high, but exercises - not shown - either controlling for isolated students or excluding them from the estimation show very similar results.

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

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 impact will depend on the number of nominated friends. As described in section 2, each student could nominated up to four best friends or colleagues. If a student nominates all fours friends, the marginal impact of an extra aspiring friend is about 3.1 p.p.[13]. If a student nominates three friends, the marginal impact of an extra aspiring friend is 4.1 p.p.. If a student nominates two friends, the marginal impact is 6.2 p.p.. Finally, if a student nominates only one friend, the marginal impact will be 12.4 p.p. - naturally, the same as passing from having no friend aspiring to college to have all friends aspiring to it.

In order to have some perspective about the size of peer effects in this context, consider that the impact of having an extra aspiring friend in a students' own aspiration is of the same magnitude of increasing this students' reading proficiency in one standard deviation.

4.1 Robustness check

As discussed in section 3.4, the main 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 actually be endogenous. In particular, if a friend of a friend is my friend, we should be worried about homophily being driving part of the results shown in Table 2. I've already shown that this is not the case for the majority of my sample since more than 50% of students nominated between one and three friends. I have also shown that the estimations actually increase when controlling for soft skills, which are potential drivers of homophily in friendship formation.

[13] If a student has four friends, the average of peers' aspiration increases by 0.25 every time an extra friend aspires to go to college. if we multiply this increase by the coefficient of peers' aspiration - which is about 0.12 - we get to the marginal effect of 0.031, or 3.1 p.p.

Yet, Tables A.2 and A.3 bring another two exercises that ensure the robustness of the results. Table A.2 brings some falsification exercises where I investigate the existence of peer effects (the endogenous social effects) in students' socioeconomic status, where the impact of peers should not exist. Columns (1) and (2) investigate whether peers impact students parents' education, that is, whether the fact that peers' mother or father have more than high school influences students' mother or father to also have more than high school. Columns (3) and (4) analyze whether the fact that peers live in their own house - in opposition of living in a rented or borrowed house - or have internet at home influence students to also live in their own house or have internet at home. These variables are clearly either pre-determined, as in the case of parental education, or very unlikely to be influenced by school peers - as in the case of having your own house or internet at home. Therefore, the presence of peer effects in these variables would indicate that the employed methodology is not completely ruling out the presence of homophily. However, as shown in the table, this is not the case: none of these exercises delivered significant results of peer effects. Column (5) of Table A.2 brings again the estimation of peer effects on college aspiration without controlling for the variables analyzed in column (1) to (4), in order to check whether the main results were not being driven only by the inclusion of these controls.

Table A.3 brings another robustness check, where I re-analyzed my results in the sub-sample of students who were not censored by the limit in friendship nomination - that is, students who nominated only three friends or less. In this restricted sample, it is possible to map all students' connection with more precision, without incurring the risk of having missing links. The results show that, if anything, censoring the data is biasing the results *downwards*, since the magnitude of estimations in Table A.3 is actually bigger than the ones in Table 4. This conclusion is similar to the one of Griffith (2019), who use data from Add Health and other smaller survey to investigate the direction of the bias when censoring network data.

4.2 Heterogeneous impacts

Table 3 brings estimations considering heterogeneous characteristics of students regarding some of their socio-economic status. Each of the variables in the columns of the table is interacted with peers' aspiration. Hence, column (1) brings heterogeneous exercises for boys and girls, column (2) brings these exercises for non-white and white students, and columns (3) and (4) bring the results for students with less/more educated parents (mother in column (3) and father in column (4)).

As shown in the table, boys and students from less educated fathers seem to be more susceptible to peer effects. This might be because girls and students from more educated fathers have higher levels of aspiration, as shown in Table 4, so the marginal impact of a friend who aspires to go to college is likely smaller for them. This makes sense especially for the case of fathers education: students whose parents have a higher level of education probably form their aspirations at home and, as so, are not so influenced by their friends.

5 Discussion about possible mechanisms

Information diffusion and conformity to social norms may play an important role in peers' influence in college aspiration. On one hand, 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. On the other hand, students might either be influenced by their friends to comply with social norms that hinder their aspiration or see college aspiration itself as a social norm to which they decide to comply. As shown in (Bursztyn and Jensen, 2017), there is a burgeoning literature on how the presence of social norms and social pressure change individuals' behavior.

Unfortunately, I cannot access all kinds of information that students have about college returns or about how to get into college. However, it is possible to get a sense of whether they are exchanging information using their perceived college returns. 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 spread of such norms could be a mechanism for peer effects on aspiration.

In order to test that, Table 4 brings 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 is not college aspiration and peers' college aspiration, respectively, but each variable in the columns of the table.

The table shows that, while peers do not seem to impact perceived college returns or the fear of nerd stigma, they do seem to have an influence on students feeling pressured to work.

I also use the information on the fear of nerd stigma and on peer pressure to work to test whether college aspiration might be itself a social norm that students want to comply with. That is, students might start aspiring to a college degree in order to be part of, or be accepted by, a group of friends that already aspires to it. If this is true, one should observe greater peer effects for students who are more responsive to the presence of social norms. In order to test for that, I test for peer effects in college aspiration in two different groups of students - one that does not seem to conform easily to social norms, and other that seems to be more responsive to it. Table 5 brings such estimation for the two measures of "conformity to social norms" that I have. Even though peer effects on college aspiration are positive and significant for all estimations, it is much larger for the group of students willing to comply with social norms. This is an indicator that college aspiration might as well be a social norm these students are responding to.

6 Peers' aspirations and future outcomes

Once the impact of peers on students' aspiration is verified, it is valid to investigate whether such influence spillovers to students' outcomes in school.

I have shown in section 2.1 how students' aspiration predicts school outcomes such as the likelihood to dropping out of school and of passing grades during high school. However, such predictions cannot have a causal interpretation - maybe both students' aspiration and their school outcomes are coming from the impact of highly motivating parents, for instance. However, it is possible to use the previous methodology in order to infer the causal impact that peers' aspirations have on students' outcomes in school. There are several reasons for such an impact. First, as shown in my main exercises, peers' aspiration influence students' own aspiration, which might change their future outcomes. Second, even after considering students' own aspiration, having aspiring friends might help the studying environment - since these friends are more invested in school activities -, might decrease students' fear of being stigmatized as a nerd, or the peer pressure too work, and might prevent students from dropping out or being retained, simply because they now want to be with their friends.

Table 6 brings the results of three estimations that measure how peers' aspiration influence students' outcomes in school. Column (1) brings estimations on the likelihood of students passing grade at the end of 2011, the year when they answered the survey; column (2) brings estimations on the likelihood of students passing all grades in high school; and column (3) brings estimations on the likelihood of students dropping out of school. One can see that even though peers' aspiration does not have an impact on the probability that students will pass grades in school, it has a positive, significant, and quite large impact on the likelihood that they will finish high school.

7 Conclusion

The results presented in this work bring valuable insights for educational policymaking in developing countries. First and foremost, I show that higher levels of college aspirations are related to higher effort in school and that there are large and significant peer effects on aspiration. These results raise important policy implications. They show that peers' aspiration impact students' own aspiration which, in turn, increases their effort in school. Therefore, peer effects in aspiration might be an important mechanism through which peers affect other school outcomes. Several works, for instance, have largely documented peer effects in student's performance (Sacerdote, 2001; Angrist and Lavy, 1999; Hanushek et al., 2003; Zimmerman, 2003; Calvó-Armengol et al., 2009; Fruehwirth, 2013; Boucher et al., 2014) or dropouts (Evans et al., 1992; Cipollone and Rosolia, 2007) and they usually find positive peer spillovers on these outcomes. Part of this impact, however, could be coming from the influence that peers have on one's level of aspiration which, in turn, might increase her effort in school. I consider this hypothesis and perform estimations showing that, indeed, peers' aspiration has a positive impact on students probability of finishing high school.

Moreover, other works also show how some educational interventions increase students' aspiration (Carlana et al., 2015; Ross, 2017; Chiapa et al., 2012). My results show that any impact coming from these interventions spillovers to peers, which should be considered in cost-benefit analysis.

Second, the impact is driven mainly by more vulnerable students, which shows that they are the most susceptible to peers' influence.

Future works should focus on peer effects in aspiration for contexts different from education attainment. Opportunities in the labor market, for instance, has 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

Angrist, J. D. and Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, 114(2):533–575.

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

Austen-Smith, D. and Fryer Jr, R. G. (2005). An economic analysis of âacting whiteâ. *The Quarterly Journal of Economics*, 120(2):551–583.

Banerjee, A. V. and Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs.

Bonilla, L., Bottan, N. L., and Ham, A. (2016). Information policies and higher education choices experimental evidence from colombia.

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.

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

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

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.

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.

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

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

Carrell, S. E. and 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. and Katz, L. F. (1991). The company you keep: The effects of family and neighborhood on disadvantaged youths. Technical report, National Bureau of Economic Research.

Chiapa, C., Garrido, J. L., and 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. and 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*, pages 728–734.

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

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.

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

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

Falk, A. and Knell, M. (2004). Choosing the joneses: endogenous goals and reference standards. *The Scandinavian Journal of Economics*, 106(3):417–435.

Fruehwirth, J. C. (2013). Identifying peer achievement spillovers: Implications for desegregation and the achievement gap. *Quantitative Economics*, 4(1):85–124.

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

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

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. (2014). Adjusting your dreams? the effect of school and peers on dropout behaviour.

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

Graham, C., Pozuelo, J. R., et al. (2018). Does hope lead to better futures? evidence from a survey of the life choices of young adults in peru. Technical report.

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

Hanushek, E. A., Kain, J. F., Markman, J. M., and 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. Technical report, National Bureau of Economic Research.

Hoxby, C., Turner, S., et al. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, (12-014).

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.

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

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.

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

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*, (0).

Lavy, V. and 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., 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. (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., and 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.

Norris, J. (2017). Family and peer effects on schooling attitudes, performance, and attainment.

OECD (2017). *Education at a Glance 2017: OECD Indicators*. OECD Publishing, Paris.

Pasquier-Doumer, L. and Brandon, F. R. (2015). Aspiration failure: a poverty trap for indigenous children in peru? *World Development*, 72:208–223.

Ray, D. (1998). *Development economics*. Princeton University Press.

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

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

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.

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

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

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

Silveira, L. C. T. d. (2018). Widening access to higher education for low-income students: a brazilian case study (1990s-2016). *Revista Brasileira de Educação*, 23.

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

Tanguy, B., Dercon, S., Orkin, K., and Taffesse, A. (2014). The future in mind: Aspirations and forward-looking behaviour in rural ethiopia.

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

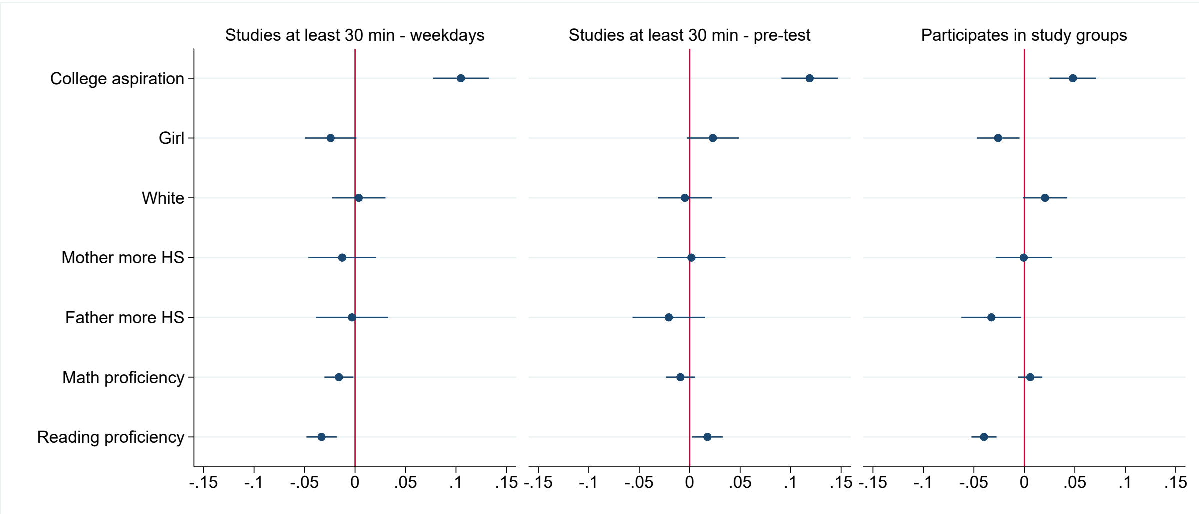
8 Tables & Figures

Table 1: Descriptive Statistics

	All		Coll. aspiration=1		Coll. aspiration=0	
	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error
Own characteristics						
College aspiration	0.68	0.46	1.00	0.00	0.00	0.00
Girl	0.49	0.50	0.55	0.50	0.35	0.48
White	0.33	0.47	0.35	0.48	0.30	0.46
Mother education: more than HS	0.22	0.41	0.23	0.42	0.19	0.39
Father education: more than HS	0.19	0.39	0.21	0.41	0.15	0.36
Math proficiency	0.00	1.00	0.10	1.00	-0.21	0.96
Reading proficiency	0.00	1.00	0.15	1.00	-0.34	0.92
Friends' characteristics						
College aspiration	0.59	0.42	0.64	0.40	0.47	0.43
Girl	0.42	0.45	0.48	0.45	0.31	0.42
White	0.27	0.34	0.30	0.34	0.22	0.32
Mother education: more than HS	0.19	0.29	0.20	0.29	0.16	0.27
Father education: more than HS	0.17	0.28	0.18	0.28	0.13	0.25
Math proficiency	0.08	0.67	0.11	0.68	0.01	0.63
Reading proficiency	0.11	0.68	0.17	0.69	-0.00	0.66
Observations	6076		4157		1919	

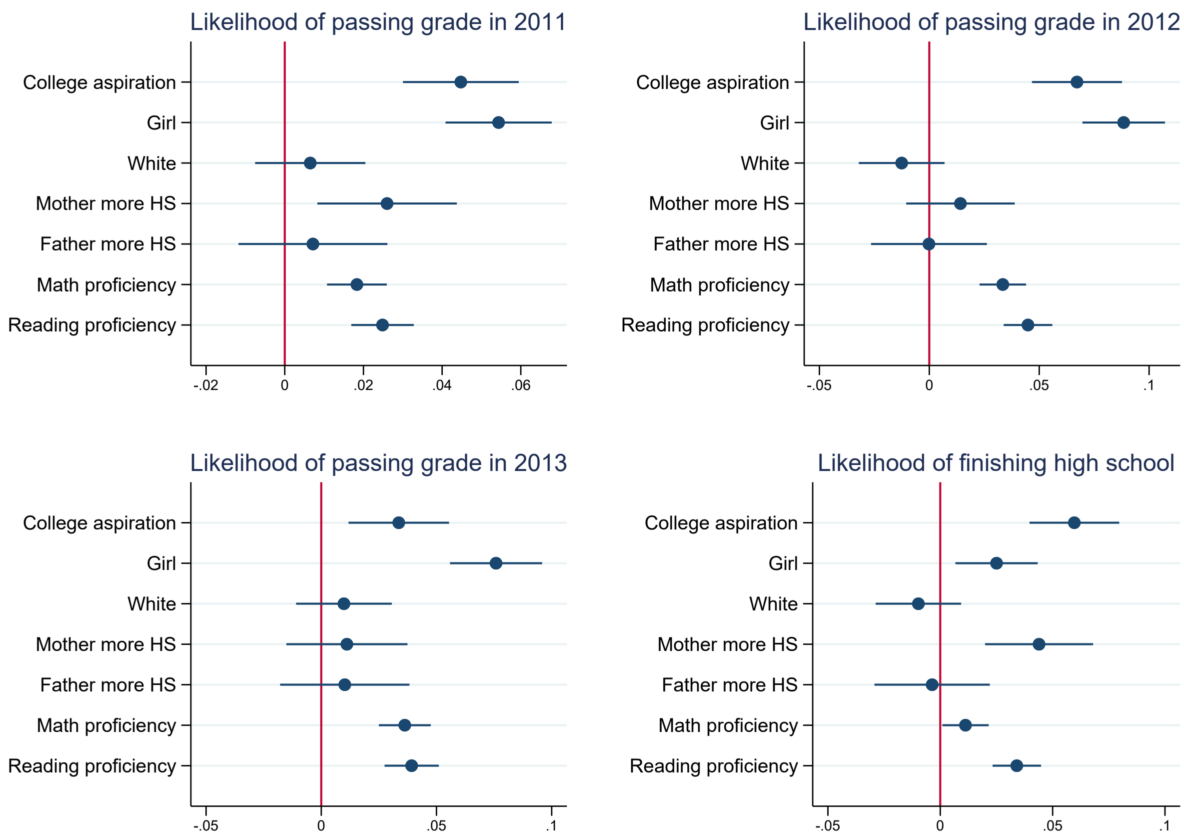
Note: "College aspiration" is a binary variable that takes value equal 1 if the student indicates that he/she wants to keep studying up to college; Math and Language proficiency are normalized with Mean=0 and SD=1.

Figure 1: College aspiration and effort in school



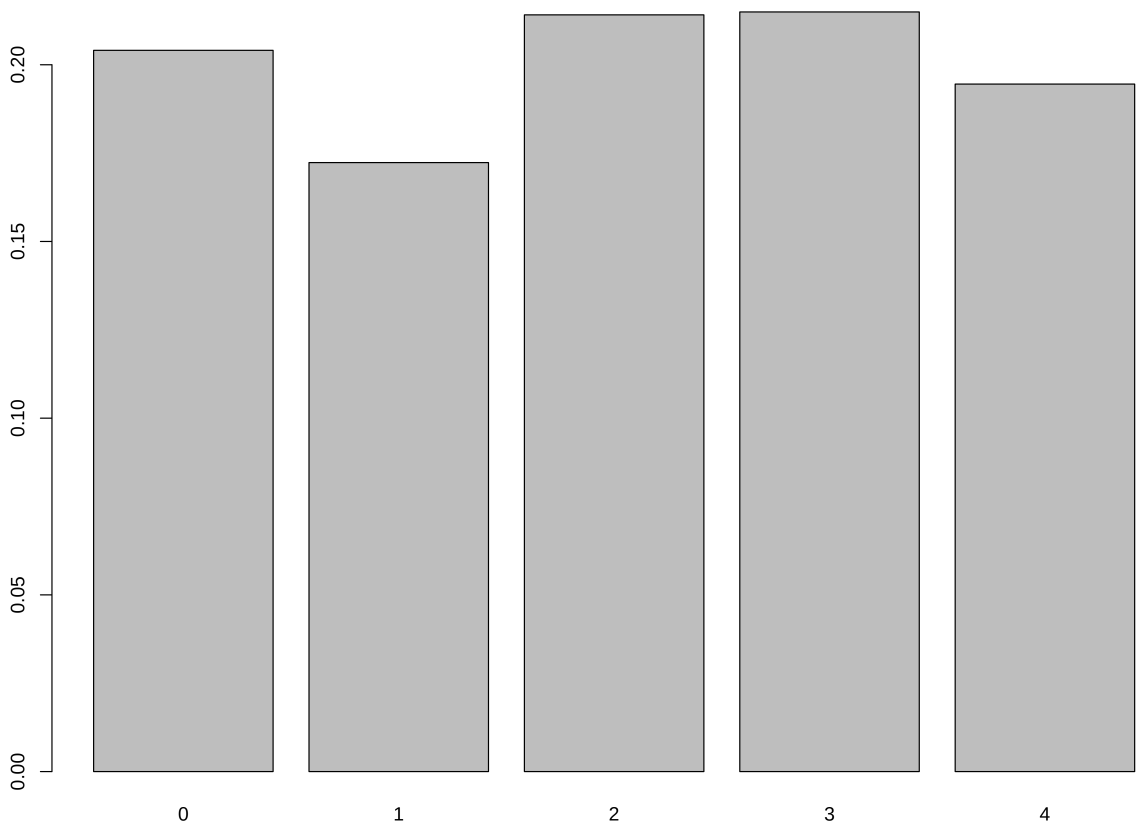
Note: results from OLS estimations where the dependent variables are binary variables that take values equal 1 if the student studies Math more than 30min/day during weekdays; more than 30min/day during pre-test days; and if the student takes part in Math study groups (bottom-right figure); Math and Reading proficiency are normalized with mean zero and standard-deviation one.

Figure 2: College aspiration and school outcomes



Note: results from OLS estimations where the dependent variables are binary variables that take values equal 1 if the student dropped school at any time after 2011 (upper-left figure); passed 9^{th} grade in 2011 (upper-right figure); passed 10^{th} grade in 2012 (lower-left figure); and passed 11^{th} grade in 2013 (bottom-right figure); Math and Reading proficiency are normalized with mean zero and standard-deviation one.

Figure 3: 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 2: Peer effects on Aspiration (N=6,076)

	(1)	(2)	(3)	(4)	(5)	(6)
Endogenous Social Effects						
Peers' aspiration	0.124***	0.134***	0.130***	0.142***	0.142***	0.149***
	(0.047)	(0.046)	(0.047)	(0.046)	(0.047)	(0.046)
Own characteristics						
Girl	0.120***	0.117***	0.120***	0.117***	0.121***	0.117***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
White	0.007	0.002	0.007	0.002	0.007	0.002
	(0.013)	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)
Mother education: more than HS	-0.019	-0.019	-0.019	-0.019	-0.019	-0.019
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Father education: more than HS	0.032**	0.029*	0.032**	0.029*	0.031**	0.028*
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Math proficiency	0.017**	0.015**	0.017**	0.015**	0.017**	0.015**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Reading proficiency	0.070***	0.060***	0.070***	0.060***	0.070***	0.060***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Exogenous Social Effects						
Girl	-0.005	-0.009	-0.007	-0.010	-0.009	-0.012
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
White	0.032	0.029	0.031	0.028	0.030	0.027
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Mother education: more than HS	-0.017	-0.014	-0.017	-0.014	-0.017	-0.014
	(0.022)	(0.021)	(0.022)	(0.021)	(0.022)	(0.021)
Father education: more than HS	0.040*	0.041*	0.039*	0.041*	0.039	0.040*
	(0.024)	(0.023)	(0.024)	(0.023)	(0.024)	(0.023)
Math proficiency	0.016	0.010	0.015	0.010	0.015	0.010
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Reading proficiency	0.010	0.009	0.009	0.009	0.009	0.008
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Model	$\hat{G}^2X$	$\hat{G}^2X$	$\hat{G}^3X$	$\hat{G}^3X$	$\hat{G}^2X$, $\hat{G}^3X$	$\hat{G}^2X$, $\hat{G}^3X$
IVs' joint significance	51.195	46.919	52.639	48.186	27.984	25.828
Control for Non-Cog Skills	No	Yes	No	Yes	No	Yes

Note: (i) Standard errors clustered at class level; (ii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (iii) Math and Reading proficiency are normalized with mean zero and standard-deviation one; (iv) All regressions include school FE and the following SES indicators: father working status, home ownership, internet at home, and number of lavatories at home

Table 3: Heterogeneous impacts (N=6,076)

Dependent variable: college aspiration				
	(1)	(2)	(3)	(4)
	Boys	Non-white	Mother less HS	Father less HS
Peers' aspiration	0.044	0.086	0.081	0.051
	(0.055)	(0.055)	(0.064)	(0.063)
Peers' aspiration x Variable in column	0.109**	0.053	0.054	0.095*
	(0.043)	(0.040)	(0.052)	(0.054)
Joint significance of Peers' Aspiration				
P-value	0.001	0.013	0.017	0.005
Model	$\hat{G}^3 X$	$\hat{G}^3 X$	$\hat{G}^3 X$	$\hat{G}^3 X$
IVs' joint significance	29.723	27.766	27.611	27.504

Note: (i) Standard errors clustered at class level; (ii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (iii) Math and Reading proficiency are normalized with mean zero and standard-deviation one; (iv) All regressions include school FE and the following SES indicators: father working status, home ownership, internet at home, and number of lavatories at home

Table 4: Peer effects on college return & compliance to social norms (N=6,076)

	(1) Perceived coll. returns	(2) Fear of nerd stigma	(3) Peer pressure to work
Endogenous Social Effects	-0.018	0.116	0.160**
	(0.030)	(0.074)	(0.068)
Model	$\hat{G}^3X$	$\hat{G}^3X$	$\hat{G}^3X$
IVs' joint significance	128.148	19.917	25.286
Control for Non-Cog Skills	No	No	No
Control for SES	Yes	Yes	Yes

Note: (i) Standard errors clustered at class level; (ii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (iii) Math and Reading proficiency are normalized with mean zero and standard-deviation one; (iv) All regressions include school FE and the following SES indicators: father working status, home ownership, internet at home, and number of lavatories at home; "Perceived college returns" is students' perceived likelihood of finding a job in case they go to college; "Fear of nerd stigma" is a binary variable that takes value equal one if students indicate the fear of being stigmatized as "nerd" as a possible impediment for them to keep studying; "Peer pressure to work" is a binary variable that takes value equal one if students indicate peer pressure to find a job as a possible impediment for them to keep studying

Table 5: Peer effects on college aspiration - heterogeneity by social norms

	Dependent variable: college aspiration			
	Fear of nerd stigma		Peer pressure to work	
	No	Yes	No	Yes
	(1)	(2)	(3)	(4)
College aspiration	0.133**	0.200*	0.098*	0.266***
	(0.056)	(0.107)	(0.054)	(0.098)
Model	$\hat{G}^3X$	$\hat{G}^3X$	$\hat{G}^3X$	$\hat{G}^3X$
N	4521	1552	4263	1811
Mean Dep. Var.	0.721	0.578	0.733	0.570
IVs' joint significance	33.717	15.418	34.299	18.745

Note: (i) Standard errors clustered at class level; (ii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (iii) Math and Reading proficiency are normalized with mean zero and standard-deviation one; (iv) All regressions include school FE and the following SES indicators: father working status, home ownership, internet at home, and number of lavatories at home; "Fear of nerd stigma" is a binary variable that takes value equal one if students indicate the fear of being stigmatized as "nerd" as a possible impediment for them to keep studying; "Peer pressure to work" is a binary variable that takes value equal one if students indicate peer pressure to find a job as a possible impediment for them to keep studying

Table 6: Friends' aspiration and students' future outcomes (N=6,076)

	(1) Pass grade in 2011	(2) Pass all HS grades	(3) Finish high school
Peers' aspiration	-0.013	0.040	0.111**
	(0.038)	(0.050)	(0.044)
Model	$\hat{G}^3X$	$\hat{G}^3X$	$\hat{G}^3X$
IVs' joint significance	38.048	38.048	38.048
Control for Non-Cog Skills	No	No	No
Control for SES	Yes	Yes	Yes

Note: (i) Standard errors clustered at class level; (ii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (iii) Math and Reading proficiency are normalized with mean zero and standard-deviation one; (iv) All regressions include school FE and the following SES indicators: father working status, home ownership, internet at home, and number of lavatories at home;

A Appendix tables

Table A.1: Probability of Forming a Friendship Link

	Raw	OR
$\mathbf{1}[\boldsymbol{x_i} = \boldsymbol{x_j}]$		
Gender	1.490***	4.437***
	(0.023)	(0.104)
Race-white	0.130***	1.139***
	(0.021)	(0.024)
Race-black	0.162***	1.176***
	(0.039)	(0.046)
Class in 2008	1.347***	3.846***
	(0.023)	(0.090)
x_j **characteristics**		
Girl	0.155***	1.167***
	(0.023)	(0.027)
Race-White	0.053**	1.055**
	(0.021)	(0.023)
Race-Black	0.120***	1.127***
	(0.040)	(0.045)
N (potential links)	524,724	524,724

Note: (i) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (ii) Class in 2008 is the class where students were allocated when enrolling in the first grade of secondary education, when they switch from municipal to state-owned school. The allocation into these first classes is made at random; (iii) estimates control for sender fixed effects.

Table A.2: Falsification exercise

	(1) Mother more than HS	(2) Father more than HS	(3) Own house	(4) Internet	(5) College Aspiration
Endogenous Social Effects	0.197	0.310	0.086	0.137	0.125***
	(0.298)	(0.345)	(0.171)	(0.293)	(0.047)
Model	$\hat{G}^3X$	$\hat{G}^3X$	$\hat{G}^3X$	$\hat{G}^3X$	$\hat{G}^3X$
IVs' joint significance	3.415	1.589	6.102	2.335	71.827
Control for Non-Cog Skills	No	No	No	No	No
Control for SES	Yes	Yes	Yes	Yes	SES

Note: (i) Standard errors clustered at class level; (ii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (iii) Math and Reading proficiency are normalized with mean zero and standard-deviation one; (iv) All regressions include school FE and the following SES indicators: father working status, home ownership, internet at home, and number of lavatories at home

Table A.3: Maximum out-degree=3

	(1)	(2)
Peers' aspiration	0.189***	0.197***
	(0.068)	(0.065)
Model	$\hat{G}^3X$	$\hat{G}^3X$
IVs' joint significance	32.525	30.304
Control for Non-Cog Skills	No	Yes

Note: (i) Standard errors clustered at class level; (ii) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.001$; (iii) Math and Reading proficiency are normalized with mean zero and standard-deviation one; (iv) All regressions include school FE and the following SES indicators: father working status, home ownership, internet at home, and number of lavatories at home