

PEER EFFECTS AND SCHOLASTIC ACHIEVEMENT: SPATIAL MODELS ESTIMATES USING THE STUDENT FRIENDSHIP NETWORK AT THE CLASSROOM LEVEL

ISABEL PESSOA DE ARRUDA RAPOSO *
TATIANE ALMEIDA DE MENEZES †
RICARDO CARVALHO DE ANDRADE LIMA ‡
RICARDO ZIMMERLE DA NÓBREGA §

Abstract

This paper evaluates the peer effects on individual academic performance. The identification strategy considers the architecture of friendship networks within classrooms, in addition to group and individual fixed effects. Estimates of spatial autoregressive models show that an increase of one standard deviation (sd) in peers' math grade improves by 6% sds the student's grade. Furthermore, when we also consider the indirect friendship bonds, the aggregate peer impact raises to 45% sds of the individual math grade.

Keywords: School performance, friendship networks, spatial models.

Resumo

Este artigo avalia o efeito dos pares sobre o desempenho escolar a partir da estrutura da rede de amizades do aluno. A estratégia de identificação explora a arquitetura dessas redes associadas a controles de efeitos fixos individuais e de grupo. Modelos de econometria espacial mostram que um aumento de um desvio-padrão na nota agregada de matemática dos amigos, gera um aumento padronizado do rendimento individual em 6%. Ademais, quando os amigos indiretos são incluídos na análise, esse impacto aumenta para 45%. As estimações utilizam base inédita da FUNDAJ (2013) que traz o levantamento da rede de amizades do aluno na turma.

Palavras-chave: Desempenho escolar, rede de amizades, modelos espaciais.

JEL classification: I20, I21, C21

DOI: <http://dx.doi.org/10.11606/1980-5330/ea134204>

* Researcher at the Fundação Joaquim Nabuco – Ministry of Education, Brazil. Email: i_raposo@hotmail.com

† Professor of economics at the Federal University of Pernambuco, Brazil.

‡ Economic Analyst, Federal Public Ministry (MPF), Brazil.

§ Researcher at the Fundação Joaquim Nabuco – Ministry of Education, Brazil.

1 Introduction

One of the prominent issues within the economics of education is devoted to understanding the role of peers in educational outcomes.¹ The behavioral influence received from friends in the social interaction process might affect educational outcomes not only during the schooling period but also later in life, having an effect on standards of educational attainment to employment decisions. Friends are a source of interaction, motivation and aspiration in the learning process and their influence operates on slightly distinct mechanisms. For example, in a more subtle way, students can benefit from externalities of knowledge created by discussions and questions from other classmates. Peer effects can also be disseminated by imitation or contagion, in which case students have individual motivations for displaying a behavior/performance that is consistent with the group in which he or she is inserted. In an important psychological study, Harris (1995) shows that childhood friends are relatively more important to their development than their parents or home environment. She states that, in the pursuit to be part of the peer group, children can radically change their personality when away from parental supervision. Behaviors that deviate from those expected by the group are generally not accepted and tend to be excluded.

Manski (1993) argues that if individual academic performance increases as the average class performance increases, then peer effects act as a social multiplier and, as a result, have implications for educational policies. Empirical identification of peer effects, however, is an arduous exercise. As first noted by Manski (1993), and after by many other authors, the greatest difficulty is to properly separate endogenous peer effects, from other contextual or correlated effects. In educational contexts, endogenous peer effects represent the influence of peers' academic outcomes on individual achievement. The contextual effect captures the effects of exogenous peers' attributes, such as age, gender, and race, on individual outcome. And the correlated or confounding effects correspond to non-observable characteristics shared by individuals in the same group that are correlated with the peer variable of interest. These common traits occur either for being exposed to the same institutional environment or for homophily, which is the propensity of people with similar attributes to associate with each other.

Depending on the way in which the boundaries of the group are defined, it is impossible to distinguish the endogenous from the exogenous peer effects, even in the absence of correlated effects. For instance, when the behavioral influence of the group is treated as homogenous, in which case all the members inside it are affected by all the others in their groups and by none outside it, the simultaneous nature of educational results creates a perfect collinearity between the expected mean outcome of the group and its mean characteristics – the *reflection problem* as defined by Manski (1993) – making infeasible the distinction between the effect of peers' educational outcome and peers' characteristics that do impact on their outcomes.

¹The terms *peers*, *friends*, *colleagues*, and *partners* will be used interchangeably in this article. *Peers* are all members belonging to a reference group (classroom, school, neighborhood, etc.) but who are not necessarily friends with each other and may not even know each other. The reason why these words are used without distinction here is to be consistent with the term widely used in the literature, *peer effects*, which, in many studies, is used to refer to all of these subjects interchangeably.

Usually, to overcome such constraints, empirical identification strategies rely either on the use of exogenous instruments for peers' educational outcome (e.g. Case & Katz 1991, Evans et al. 1992, Hanushek et al. 2001, Ding & Lehrer 2007, Goux & Maurin 2007, Vigdor & Nechyba 2007, Sund 2009) or on experimental or quasi-experimental data that claim that groups are formed by rules exogenous to individuals, thus avoiding confounding or correlated effects (Eisenkopf et al. 2011, Duflo et al. 2008, Sacerdote 2001, Zimmerman 2003, Oosterbeek & Van Ewijk 2014).

A new segment in the related literature uses as identification strategy the architecture of social networks – structures that keep track of all the links among its members and that build information flow, social norms, and social behaviors (e.g. Patacchini & Venanzoni 2014, Badev 2014, Patacchini et al. 2011, Angelo 2010, Calvó-Armengol et al. 2009, Bramoullé et al. 2009, Ballester et al. 2006). The main idea behind this strategy is that the reference groups are not homogeneous: they might be different in size, but they do not necessarily completely overlap. For example, it is possible that within the same network student i and j are friends, student j and k also, but i and k are not friends. Therefore, the reference group of each member is different in size and composition. Bramoullé et al. (2009) argue that it is exactly this heterogeneous nature of reference groups that permits us to overcome Manski's *reflection problem*, and provide us with formal conditions to identify the endogenous peer effects.

The main objective in this paper is to assess the dissemination of peer effects on school performance within the framework of students' friendship networks. The diffusion of endogenous peer effects is captured by observed friend's school achievement in a *spatial autoregressive model* (SAR). Conditional Maximum Likelihood (CML) estimations together with network and individual fixed effects are adopted as econometric strategies for identifying the influence of peers' academic outcomes.

We closely follow the paper of Calvó-Armengol et al. (2009); however, we advance and provide additional contributions in the following aspects. The empirical estimations are based on a unique educational dataset that comes from a research institute of the Brazilian Ministry of Education (Fundação Joaquim Nabuco/FUNDAJ, 2013), which provides a large information set on students' school environment. This dataset offers two crucial features for identifying peer effects: (i) it raises the direct friendship network within the classroom, and (ii) it provides a longitudinal evaluation of scholastic achievement, in which all the students surveyed are submitted to a math test in the beginning and at the end of scholar year. The initial math grade turns to be especially appropriate to control for preexisting differences between students. Each exam consists of twenty multiple choice questions and was conceived according to the *Item Response Theory* (IRT)². To the best of our knowledge, besides the *National Longitudinal Survey of Adolescent Health (Add Health)*³,

²*Item Response Theory* models the response of each examinee of a given ability to each item in the test. The items are not assumed to be equally difficult, and the difficulty level of each one is treated as information to be incorporated in scaling items.

³*Add Health* is a longitudinal survey of American schools with teens that focuses primarily on the network of friends of these students. The research has a sample of 90,118 students distributed in a representative manner in 132 public and private schools in the United States. Since 1994, the survey has collected five panels (1994, 1995, 1996, 2001 and 2008) to the students of the 7th year by the end of high school. A subsample of these students (approximately 20,000) answers a very

there is no other survey that provides such data; it is thus unique for developing countries.

Besides this introduction, the paper is organized into four additional sections. Section two describes the FUNDAJ ((FUNDAJ - Fundação Joaquim Nabuco and Coordenação de Estudos Econômicos e Populacionais 2013)) dataset and offers a descriptive analysis based on the sample variables used. The third section presents the peers network model and the identification strategies implemented to estimate peer effects on academic performance. The fourth section provides the estimated results, and finally, the last section brings together the main findings of the article.

2 The dataset and descriptive statistics of the variables

This article uses a unique dataset from a survey conducted by the Joaquim Nabuco Foundation in 2013 with a sample of students in the 6th year of public schools in the city of Recife, Pernambuco State, Brazil. The research evaluated student performance on two math tests (which were given at the beginning and the end of the school year). Interviewers also collected a large set of information on internal and external aspects of school life through four types of questionnaires (one for students, one for the primary adult responsible for each child's academic life, one for the math teacher, and another one for the school principal). The main highlight of the survey was the identification of the student friendship networks inside the classrooms. This type of data is crucial for identifying the extent of peer influence on the learning process. In the student's questionnaire, they listed up to five best friends and reported whether they were classmates, whether they studied together, whether they go to each other's houses, and whether they talk about problems. The student's caregivers also reported whether they knew those friends and/or their families and what kind of influence they believed these friends had on their child. Thus, for every friend from the classroom, the data included not only the friend's grade on the two math tests but also all of the additional information collected from the questionnaires.

Other characteristics of the research, such as the student's perceived self-esteem, future aspirations, and anthropometric measurements, provided a rich set of control variables that enabled better specifications of the models.

In total, 4,191 students, 3,670 parents or guardians, 120 principals, and 131 teachers from 120 schools were interviewed. The school units were spatially distributed among 18 micro-regions of Recife⁴. In some schools with higher enrollment in the 6th year, two groups were randomly selected, not just one. For this reason, the total number of classrooms selected for the sample was 146.

After the exclusion of individuals with inadequate or missing information, the final sample consisted of 139 networks/classrooms and 1,670 students.

detailed questionnaire that contains information about their network of friends referring up to ten best friends, five females and five males.

⁴Each political-administrative region of Recife is divided into three micro-regions "in order to define municipal interventions at the local level and in coordination with the population". Each region is composed of one or more of the 94 Districts established by Brazilian Law. The 18 micro-regions correspond to the division of Political-Administrative Regions, which was conceived in 1995 by the Department of Social Policies to organize Participatory Budgeting meetings that had been limited to associations and their representatives (PNUD 2005).

Table 1 shows the definition and descriptive statistics of the variables used in the estimations of this paper. Performance in math was assessed at the beginning and at the end of school year, and on average, students in the sample had about 40% success, with no significant variation during the period of study.

3 Peers network model and identification strategies

The model developed here is based on Ballester et al. (2006) and Calvó-Armengol et al. (2009), and the changes adopted are properly mentioned throughout the paper.

Network: Suppose there is a social network of $1, 2, \dots, n$ agents, represented by matrix \mathbf{G} , which brings together a collection of bilateral influences g_{ij} between them. There is an active connection within the network when $g_{ij} = 1$ and an inactive connection when $g_{ij} = 0$. As per convention, $g_{ii} = 0$.

Preferences: Consider a network game⁵ in which every player $i = 1, 2, \dots, n$ selects a specific educational effort (y_i) and obtains a payoff $u_i(y_1, y_2, \dots, y_n; \mathbf{g})$, which depends on the friends efforts (y_{-i}), defined as:⁶

$$u_i(\mathbf{y}; \mathbf{g}) = \mu y_i^0 y_i - \frac{1}{2} y_i^2 + \rho \sum_{j=1}^n g_{ij} y_i y_j \quad (1)$$

The variable $y_i^0 \geq 0$ represents the initial educational background of individual i and can be interpreted as an innate condition for learning. g_{ij} denotes the friendship relationship within the network. The utility of each individual is strictly concave compared to his/her own efforts, $\frac{\partial^2 u_i}{\partial y_i^2} = -1$, and reflects strategic complementarities in relation to the efforts of the others, $\frac{\partial^2 u_i}{\partial y_i \partial y_j} = \rho g_{ij} \geq 0$. As pointed out by Calvó-Armengol et al. (2009), educational strategic complementarities means that if colleague j increases his/her educational efforts, y_j (e.g., by studying for longer hours), then student i will have his/her marginal utility increased if he or she also broadens effort, y_i . According to Calvó-Armengol et al. (2009), the parameter ρ measures the intensity of the endogenous imitation peer effects. This parameter captures to what extent students are “contaminated” by the behaviors of their companions.

Parameter μ , in turn, means that, regardless of bilateral influences, each

⁵In network games, the interdependence between players’ payoffs comes from the frame of their connections within a network (Ballester et al. 2006). The action of any member of the individual i reference group could affect the outcome of i , either because it is directly connected to this person ($g = 1$) or because even if not directly connected to him/her, this individual i can also be affected indirectly, through the influence of a friend who is connected to him/her.

⁶Model of equation 1 is a special case of the Ballester et al. (2006) model:

$$u_i(\mathbf{y}; \mathbf{g}) = \alpha_i y_i - \frac{1}{2} (\beta - \gamma) y_i^2 - \gamma \sum_{j=1}^n y_i y_j + \lambda \sum_{j=1}^n g_{ij} y_i y_j$$

in which $\alpha_i = \mu y_i^0$, $\beta = 1$, $\gamma = 0$ and $\lambda = \rho$. In the article, the authors decompose the interactions between players in three types of effects: $\Sigma = -\beta \mathbf{I} - \gamma \mathbf{J} + \lambda \mathbf{G}$. $-\beta \mathbf{I}$ represents the effect of the effort itself (concave preference), $-\gamma \mathbf{J}$, global substitution effect and $+\lambda \mathbf{G}$, the effect of the local complementarities. The global substitution effect means that if an agent j increases his/her efforts, then individual i will suffer a reduction in his utility if he/she decides to increase his/her efforts as well. These effects, however, are more natural when there are strategic interactions in the provision of a public non-excludable good, such as the sharing of ideas and information by opinion makers (providers) along with their connections (free-riders) (Jackson, 2008). In the model proposed herein, it is considered the only scenario in which students exhibit complementary strategies in their social interactions, or when $\gamma = 0$.

Table 1: Definition and descriptive statistics of the variables

<i>Dependent Variable (y)</i>	Definition of variables	Average	Standard Deviation
Math grade at the end of the year	Grade of the student i on math test conducted by FUNDAJ ((FUNDAJ - Fundação Joaquim Nabuco and Coordenação de Estudos Econômicos e Populacionais 2013)) at the end of school year. Rating scale ranges from 0 to 100.	40.73	16.00
<i>Individual Characteristic (X)</i>			
Initial math grade (y^0)	Grade of student i on math test conducted by FUNDAJ (2013) at the beginning of school year. Rating scale ranges from 0 to 100.	43.43	16.60
Male	Dummy equals 1 if student is male	0.44	0.50
Whites	Dummy equals 1 if students declare themselves as white and 0 if they describe themselves as black, mixed race, Asian, or indigenous	0.19	0.39
Age	Age of the student in years	11.18	0.86
Newcomer	Dummy equals 1 if student has been attending the school researched for less than one year	0.73	0.44
Dedication to Studies	Student answers the question "How often do you study the school materials:" 1 = every day of the week, 2 = only on school days, 3 = 3 days per week, 4 = less than 3 days per week, 5 = only if there is a test, 6 = never or almost never	2.55	1.51
Praise from Teacher	Student answers the question "the math teacher praises or congratulates you when you get a good grade or do the homework well": 1 = always or almost always, 2 = sometimes, 3 = never or almost never	1.49	0.66
Personality	Student answers the question "I would change something in my personality": 1 = strongly agree, 2 = perhaps, 3 = strongly disagree	2.40	0.88
Left out in classroom	Student answers the question "Do you feel left out in class": 1 = always or almost always, 2 = sometimes, 3 = never or almost never	2.70	0.57

No. of observations: 1,670 students

No. networks/classrooms: 139 classrooms

Source: Original compilation based on FUNDAJ (2013).

Table 1: Definition and descriptive statistics of the variables (continued)

<i>Individual Characteristic (X)</i>	Definition of variables	Average	Standard Deviation
Popular	Student answers the question “I am a popular person, I have many friends”: 1 = strongly agree, 2 = perhaps, 3 = strongly disagree	1.35	0.67
Visits to friends’ houses	Number of friends in the class who students regularly visit	1.05	1.25
Number of friends	Number of nominated friends of student <i>i</i>	2.88	1.33
Religiousness	Student answers the question “Do you ever go to church/mass?”: 1 = always or almost always, 2 = sometimes, 3 = never or almost never	1.76	0.70
Sports clubs, gyms	Dummy equals 1 if student attends a sports club, sports center or fitness facility in their neighborhood	0.23	0.42
Safety in the neighborhood	Dummy equals 1 if student says he/she feels safe in their neighborhood	0.79	0.40
Proportion of peers neighbors	Number of friends of student <i>i</i> living in the same neighborhood over the total number of student <i>i</i> friends	0.30	0.35
Male sex (parent or guardian)	Dummy equals 1 if a parent or person responsible for the student is male	0.13	0.34
White (parent or guardian)	Dummy equals 1 for parents/guardians who declare themselves white and 0 if they declare themselves as black, mixed race, Asian, or indigenous	0.17	0.38
Age (parent or guardian)	Age of student’s primary parent or guardian in years	38.48	8.18
Educational level (parent or guardian)	Parents/guardians answer the question: “What is the highest grade finished successfully?”: 1 = 1 year (literacy), ..., 9 = 9 years; 10 = 1 year of high school, ..., 12 = last year of high school; 13 = 1 year of university ..., 18 = final year of university	8.90	3.45
Marital Status (parent or guardian)	Dummy equals 1 for parents/guardians married with a legally recognized union or common-law marriage	0.55	0.50

No. of observations: 1,670 students
 No. networks/classrooms: 139 classrooms
 Source: Original compilation based on FUNDAJ (2013).

Table 1: Definition and descriptive statistics of the variables (continued)

	Definition of variables	Average	Standard Deviation
<i>Individual Characteristic (X)</i>			
Kinship (parent or guardian) with student	Dummy equals 1 for natural or adoptive parent and 0 for the other cases (grandparents, uncles, brothers, stepfather / stepmother, etc.)	0.82	0.38
Beneficiary of social program (parent or guardian)	Dummy equals 1 if parents/guardian receives any government financial assistance	0.62	0.49
Student grade report (parent or guardian)	Parents/guardians answered the question “Do you check the student’s grade report?”: 1 = always or almost always, 2 = sometimes, 3 = never or almost never	1.19	0.50
<i>Peers’ math grade (Gy)</i>	Aggregate value of math grade (end of the year) over student <i>i</i> friends	93.80	71.30
<i>Peers’ characteristic (GX)</i>	Average values of all students’ control variables among the group of direct friends of student <i>i</i>		
<i>Network/classroom features</i>			
Classroom is disciplined	Teacher responds to the statement, “This classroom is disciplined”: 1 = agree, 2 = partially agree, 3 = disagree	2.09	0.74
Classroom is aggressive	Teacher responds to the statement “This classroom is aggressive/violent”: 1 = agree, 2 = partially agree, 3 = disagree	2.55	0.63
Cumulative tests	Dummy equals 1 if the teacher says that the subjects of the tests are cumulative	0.67	0.47

No. of observations: 1,670 students

No. networks/classrooms: 139 classrooms

Source: Original compilation based on FUNDAJ (2013).

individual has a positive payoff that depends on his/ her initial educational background y_i^0 . Therefore, μy_i^0 measures the return of the educational effort, given the *ex-ante* circumstance of the individual.

Equilibrium: The Nash equilibrium $y_i^*(\mathbf{g})$ is defined as the action of agent i that maximizes payoff $u_i(\mathbf{y};\mathbf{g})$, which is found by maximizing equation (1) relative to y_i . Thus, for each agent $i = 1, 2 \dots, n$, the following best-response function is yielded:

$$y_i^*(\mathbf{g}) = \mu y_i^0 + \rho \sum_{j=1}^n g_{ij} y_j \quad (2)$$

An interesting feature of this equilibrium solution is that it can be expressed as a function of a sociability measure named Katz-Bonacich centrality $\mathbf{b}_{\mathbf{y}^0}$:

$$\mathbf{y}^* = \mu(\mathbf{I} - \rho\mathbf{G})^{-1} \mathbf{y}^0 = \mu \mathbf{b}_{\mathbf{y}^0} \quad (3)$$

This measure of centrality determines the importance of individual i in a network \mathbf{G} and considers his/her direct and indirect ties with each individual of the network, where a lower weight is assigned to more distant connections (Bonacich 1987). As a result, the estimated Katz-Bonacich centrality provides the global influence of peers over student achievement, not only through direct friends, but also through indirect ones. Its mathematical expression is denoted by the following expression (4):

$$\mathbf{b}_{\mathbf{y}^0} = (\mathbf{I} - \rho\mathbf{G})^{-1} \cdot \mathbf{y}^0 = \mathbf{y}^0 + \rho\mathbf{G} \cdot \mathbf{y}^0 + \rho^2\mathbf{G}^2 \cdot \mathbf{y}^0 + \rho^3\mathbf{G}^3 \cdot \mathbf{y}^0 + \dots = \sum_{l=0}^{+\infty} \rho^l \mathbf{G}^l \cdot \mathbf{y}^0 \quad (4)$$

in which \mathbf{G}^l identifies the indirect links that run the length l of each agent and when ρ is sufficiently small, this infinite sum converges to a finite value. In order to provide an empirical counterpart of the theoretical results derived so far, we extend equation (2) and propose the following structural model which will be used for estimations:

$$y_{i,k} = \mu y_i^0 + \rho \sum_{j=1}^{n_k} g_{ij,k} y_{j,k} + \theta_{i,k}(\mathbf{x}) + \phi_{\zeta} \zeta_{i,k} + \theta_k(\mathbf{c}) + \phi_{\eta} \eta_k + \varepsilon_{i,k} \quad (5)$$

where $y_{i,k}$ is the math grade at the end of the school year of student i in network k and $k = 1, \dots, K$ is the set of networks. In each network, there are n_k individuals, and the total number of students is $n = \sum_{k=1}^K n_k$. The components $\theta_{i,k}(\mathbf{x})$ and $\theta_k(\mathbf{c})$ represents control functions in which \mathbf{x} is a vector of individual characteristics, while \mathbf{c} is another vector of classrooms features. The other terms $\zeta_{i,k}$ and η_k are included in order to represent the non-observable effects of the individual and the network, respectively. In matrix notation, equation (5) is denoted by:

$$\mathbf{y} = \mu \mathbf{y}^0 + \rho \mathbf{G} \mathbf{y} + \theta(\mathbf{x}) + \phi_{\zeta} \zeta + \theta(\mathbf{c}) + \phi_{\eta} \eta + \varepsilon \quad (6)$$

where $\theta(\mathbf{x}) = \mathbf{X}\beta + \mathbf{G}\mathbf{X}\gamma$ and $\theta(\mathbf{c}) = \mathbf{C}\alpha$

G is constructed so as to create a block diagonal matrix in which the matrix interaction of every classroom g_k creates a specific block⁷. As a result, students who belong to a particular network or classroom g_k do not connect with students who are members of another network g'_k . It should be remembered that the reference group for each student i is the quantity of nominated friends within the network to which the student belongs and that there may be intersections among these intra-network groups.

The main parameters of interest are (μ, ρ) . The variable linked to μ is the student's initial educational background (y^0), measured here by the math grade in the beginning of the school year, while the variables linked to ρ are the math grades at the end of the school year received by the student's nominated friends in his/her classroom (Gy).

3.1 Identification strategies

In this study, the identification strategies of peer effects consist of two exercises. In the first one, we conduct ordinary least squares estimations on equation (6) in order to capture potential correlations between individual's academic performance and direct friends' achievement. However, even under the control of network fixed effects, the parameters (μ, ρ) might remain biased because of spatial dependence on the error term introduced by the simultaneity nature of the dependent variable y . Therefore, in a second exercise, we treat equation (6) as a spatial autoregressive model and adopt maximum likelihood (ML) estimations in order to control for simultaneity bias in the dependent variable.

Nevertheless, similar to the OLS models, the estimation strategy of spatial econometric models relies on the assumption that the variable of interest is not correlated with the regression error. Therefore, the exogeneity requirement of matrix G is a necessary condition for obtaining consistent estimates. It turns out that this is a strong assumption, because friendships are usually structured by a sense of homophily. Thus, the procedure adopted in literature (Patacchini & Venanzoni 2014, Calvó-Armengol et al. 2009, Bramoullé et al. 2009, Lee 2007) includes group fixed effect η_k to capture unobservable heterogeneity, which may be correlated with the variables of interest, in such a way that matrix G conditional to η may be treated as exogenous, or $E(\varepsilon_{i,k} | \mathbf{x}_k, \mathbf{c}_k, g_{ij,k}, \eta_k) = 0$. Patacchini & Venanzoni (2014) find that the inclusion of network fixed effects is a sufficient strategy to control for individual fixed effects as well, when working with relatively small social networks, as in *Add Health* and also in the sample used herein. Their idea is that, in small networks, unobservable heterogeneity is common among all individuals within a group, and therefore, the network fixed effects would represent a good approximation for capturing the unobserved individual attributes, as well.

⁷As an illustration consider two relational matrices for two hypothetical groups each with three students, g_1 and g_2 . The diagonal joint matrix G , is denoted by:

$$g_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad g_2 = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad G = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

In this paper, besides the inclusion of dummies per classroom/network⁸, we also rely on the initial math score to control for individual fixed effects. The initial math grade works here as a strategy to capture unobservable individual idiosyncrasies. According to Ding & Lehrer (2007), the students' initial grade would be a sufficient statistic for capturing a variety of influences that may confound the analysis and include all observable and unobservable historical background information on students' family, school, and community. The authors assume, hypothetically, that the students' initial grade follows a Markov process, and therefore, previous observable and unobservable factors in $t-1$ conform to the same rate, so that none of these would cease to be represented by $y_{i,t-1}$. The authors' assumption therefore allows the student's initial grade to work as a sort of individual fixed effect, since it brings with it unobservable components (such as effort and skill), which are invariable throughout the school year⁹. Thus, y_i^0 may act as a *proxy* for the student's initial educational background.

Finally, it is worth noting that there may also exist unobservable features at the school level — for example, the quality of teachers or student body, or the factors involved in parents' choices of residence — all of which can affect both academic performance and network formation. However, the strategy of controlling for network fixed effects is also capable of absorbing this fixed effect of school, especially considering that, in the sample used, approximately 80% of the schools surveyed had only one classroom/network under investigation.

4 Results

The empirical estimations for academic peer effects are presented in this section. A variety of models are tested using different methods and increasing order of covariates. The estimations use the indirect friendship network, in which case the G matrix is symmetric¹⁰.

Columns 1 to 3 of Table 2 exhibit the least squares estimates as an initial exercise to investigate the correlation between individual performance and the performance of his/her reference group of friends. The positive and significant coefficient ρ shows that individual's academic performance and the academic performance of the group of friends are directly correlated. The strength of this correlation decreases as additional covariates are included and with network fixed effects, as illustrated in column 3. Correlations μ between the student's initial background and his/her performance at the end of the year are significant in all cases and decrease in magnitude based on the

⁸Another approach commonly used to control for group fixed effects is the deviation to the reference group mean, in which for each variable in the model it is subtracted its group mean.

⁹Consider two models with a regression structure for grades at the beginning and end of the year for student i : $y_i^0 \equiv y_{i1} = \underbrace{\alpha_1 X_{i1}}_{\text{Observable}} + \underbrace{\beta_1 u_i + \varepsilon_{i1}}_{\text{Non-observable}}$ e $y_{i2} = \underbrace{\alpha_2 X_{i2}}_{\text{Observable}} + \underbrace{\beta_2 u_i + \varepsilon_{i2}}_{\text{Non-observable}}$ where u_i

is an unobservable component invariant throughout the scholarly year. According to Boardman & Murnane (1979), if $\alpha_2 = \theta\alpha_1$ and $\beta_2 = \theta\beta_1$, the effects of variables X and μ change at the same constant rate θ between $t-1$ and t . Under such conditions, the inclusion of y^0 in the empirical model (6) allows us to control for this fixed initial condition for each student.

¹⁰Some studies show that the results of peer effects do not change according to the symmetry of matrix G. See, for instance, Patacchini & Venanzoni (2014), Liu et al. (2014), Calvó-Armengol et al. (2009).

amount of control variables.

Columns 4 to 6 of Table 2 provide results of maximum likelihood estimates of the spatial models. The ML coefficients coincide with the OLS coefficients but are slightly smaller in magnitude. Column 4 presents the estimates of the SAR model without the main important identification strategies: the student's initial math grade and the dummies for network fixed effects. The model of column 5 adds the student's initial math and the model of column 6 includes dummies per classroom in order to control for the network fixed effects. The coefficients (μ, ρ) vary depending on the inclusion of additional controls. The influence of peers performance on math over individual achievement measured here by coefficient ρ is reduced in approximately 14% when the model is controlled for student's initial math grade and it is further reduced in 49% when dummies for network fixed effect are included. A possible conclusion is that individual and group unobserved characteristics tend to overestimate the magnitude of peer effects ρ through a positive bias of omitted variable.

Concentrating on our preferred specification model (column 6 of Table 2), we find that a one standard deviation increase in peers' aggregate math grade generates a 6.2% increase of a standard deviation in student grade¹¹.¹² The standardized coefficients (or beta coefficients) permit to compare the importance of different control variables over the outcome variable, independently of their measurement units. Thus, in relative terms, the peers math performance stands out as the third most important predictor of the student math grade at the end of the school year, remaining only less important than own initial math score (40%) and student's age (9%). The complete estimations of model (6) from Table 2 are presented in Table A.1 of the appendix, in which it is possible to verify that the influence of friends' achievement is even higher than the effect of the parental schooling, the student perception about own popularity and the influence of parents who are careful to check for pupil's grade report. Many control variables lose significance when dummies for group fixed effects are included in the model, possibly because they absorb most of individual and group heterogeneity, as already argued. When those dummy variables are not included, as in the models of columns (4) or (5), the results obtained reflect some findings commonly observed in the literature in which parental schooling increases student performance; older students perform lower, and those whose parents check their school records tend to show higher grades¹³.

From the estimates of $\hat{\rho}$, one can calculate the impact of the Katz-Bonacich measure weighted by the student's initial math grade. As previously defined, this measure reflects the effect of direct and indirect friendship connections

¹¹The standardized coefficient, or beta coefficient, is calculated as follows: $\widehat{\rho}_{sd} = \hat{\rho} \left[\frac{sd(Gy)}{sd(y)} \right]$

¹²Prior literature finds considerable variance in the values for peer effects, ranging from near zero (Sanbonmatsu et al. 2006) to 35% standardized impact (Hanushek et al. 2001). In the current study, the peer effects magnitude is very close to other studies using similar approaches. For example, Patacchini et al. (2011) find a peer impact of 10% of a standard deviation of student grade, Calvo-Armengol et al. (2009) find 7%, Lin (2010) find 13% and Sacerdote (2001) find 5%. Comparisons to previously obtained results are difficult to establish because of different methods used, and mainly because of differences in study groups. Just to illustrate, here are a few findings in the literature: Duflo et al. (2008) found a peer effect impact of 14% of a standard deviation of student grade; Sund (2009), 8%; and Vardardottir (2013), 32%.

¹³The complete estimations are omitted here, but they are available on request with the authors. In the appendix, the Table A.1 presents the complete estimates from model (6) of Table 2

Table 2: Estimations for equation 6. Dependent variable: math grade at the end of the year

	OLS			SAR		
	1	2	3	4	5	6
Peer effects ($\hat{\rho}$)	0.158	0.134	0.071	0.138	0.120	0.062
(Test statistic)	(5.775)	(5.401)	(2.496)	(5.488)	(5.109)	(2.417)
Initial math grade ($\hat{\mu}$)		0.432	0.398		0.432	0.399
(Test statistic)	-	(19.446)	(16.913)	-	(19.758)	(17.947)
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Peers' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Classroom features	Yes	Yes	Yes	Yes	Yes	Yes
Network fixed effects	No	No	Yes	No	No	Yes
No. of observations	1,670	1,670	1,670	1,670	1,670	1,670
No. of networks	139	139	139	139	139	139

Source: Original compilation based on FUNDAJ (FUNDAJ - Fundação Joaquim Nabuco and Coordenação de Estudos Econômicos e Populacionais 2013).

on individual performance. The results are presented in Table 3. Calculations of centrality are given according to equation (4), where point estimates $\hat{\rho}$ are obtained from columns (2), (3), (5) and (6) of Table 2. Note it is not possible to obtain the desired weighted centrality measure for the estimates from columns (1) and (3), since the model specification do not include the initial math grade. The average Katz-Bonacich centrality values range from 45.10 to 46.73 depending on the model, and the standard errors are in parentheses. The results of standardized impact (second row of Table 3) show that an increase of one standard error of the sociability measure translates into an increase in performance that will range from 40% to 45% standard errors of the end grade, which corresponds to additional 6.40 points on the math exam, something between one and two questions¹⁴.

Table 3: Impact of Katz-Bonacich centrality on the math grade at the end of the year

	OLS		SAR	
	1	2	3	4
Average Katz-Bonacich: $\widehat{b}_{y0}(g, \hat{\rho})$	46.73	45.10	45.36	44.88
(Standard error)	(17.75)	(16.82)	(17.05)	(16.78)
Impact of one s.d increase of $\widehat{b}_{y0}(g, \hat{\rho})$ on the final grade	0.45	0.40	0.44	0.40

Source: Original compilation based on FUNDAJ ((FUNDAJ - Fundação Joaquim Nabuco and Coordenação de Estudos Econômicos e Populacionais 2013)).

4.1 Robustness tests

In order to check for the robustness of peer effects estimates, we develop additional econometric exercises using two alternative peers' matrices besides the nominated friendship matrix. The first one consists of a "randomized

¹⁴Because the standard deviation of the math grade at the end of the year is 16 (see Table 1), a 40% increase generates, on average, additional 6.40 points: $(0.40 \times 16 = 6.40)$.

friendship matrix” built by drawing at random from the original peer’s matrix, a new friendship network within each classroom, preserving the total number of students in class, but randomizing the friendship links. The other is defined as the “same class matrix”, which assigns value one for all students attending the same class. The robustness tests consist in re-estimate models from columns (4) to (6) of Table 2, but including these alternative peers’ matrices along with the other control variables. If the coefficient attached to the nominated friendship network becomes no longer significant, then peer effects can be interpreted simply as a correlation between similar people in the same classroom rather than causal effect. The results from these additional exercises are reported in Table 4, and it turns out that for all specifications models - columns (1) to (3) - the peer effects coefficient ($\hat{\rho}$) remains significant and similar in magnitude from the estimates presented on Table 2. Notice that we do not find any effect of simulated peer’s network on one’s own outcomes ($\hat{\rho}_r$) for specifications (1) and (2), only for the model of column (3) we capture an inverse relationship between student math score and the math performance of a randomized friendship network, which is probably reflecting an adverse group effect, common amongst students from public schools in Brazil, who share poor school environment. The average influence of all the students in class, depicted from the coefficient ($\hat{\rho}_{all}$), demonstrates, as expected, a positive correlation between own math achievement and the mean performance from all classroom, for all the tested models.

Table 4: Robustness tests for different matrices Dependent variable: math grade at the end of the year

	SAR		
	1	2	3
Peer effects ($\hat{\rho}$) (Test statistic)	0.134 (4.522)	0.116 (4.189)	0.107 (3.743)
Peer effects using the “randomized friendship matrix” ($\hat{\rho}_r$) (Test statistic)	-0.006 (-1.598)	-0.005 (-1.331)	-0.010 (-2.686)
Peer effects using the “same class matrix” ($\hat{\rho}_{all}$) (Test statistic) (Test statistic)	0.100 (3.777)	0.073 (3.078)	na
Initial math grade ($\hat{\mu}$) (Test statistic)	-	0.429 (19.602)	0.398 (17.936)
Individual characteristics	Yes	Yes	Yes
Peers’ characteristics	Yes	Yes	Yes
Classroom features	Yes	Yes	Yes
Network fixed effects	No	No	Yes
No. of observations	1,670	1,670	1,670
No. of networks	139	139	139

Source: Original compilation based on FUNDAJ ((FUNDAJ - Fundação Joaquim Nabuco and Coordenação de Estudos Econômicos e Populacionais 2013)).

$$y = \mu y^0 + \rho G y + \rho_r G_r y + \rho_{all} G_{all} y + \theta(x) + \phi_{\zeta} \zeta + \theta(c) + \phi_{\eta} \eta + \varepsilon.$$

5 Final considerations

This paper evaluates the dissemination of peer effects on school performance from the framework of students’ friendship networks. The identification strategy exploits the architecture of these social networks to separate the endoge-

nous effects of peers from other exogenous or contextual effects. It uses a unique dataset (FUNDAJ, (FUNDAJ - Fundação Joaquim Nabuco and Coordenação de Estudos Econômicos e Populacionais 2013)), which raises the students' friendship network within the classroom, thus bringing crucial information for the identification of peer effects, once it takes into account the intragroup heterogeneity. From theoretical and methodological points of view, this study provides two main contributions. First, it uses a recent dataset very similar to that of the *Add Health Survey* but unique for developing countries. Second, the paper provides a new equilibrium solution for the theoretical model of Calvó-Armengol et al. (2009), in which a student's educational outcome is proportional to a Katz-Bonacich centrality measure weighted by his/her initial learning condition. This *achievement-weighted* centrality permits one to control preexisting differences between students, which is especially appropriate for separating endogenous from correlated effects.

We adopt maximum likelihood estimations for spatial econometric models in order to test for the impact of peers' school performance on individual's math grade. We find that a one standard deviation increase in peer's aggregate math grade generates a 6% increase of a standard deviation in a student's grade. The inclusion of contextual variables, such as individual attributes and average attributes of friends, and especially the use of dummies for controlling for group fixed effects, are all important strategies that can reduce the bias of omitted variables. Robustness tests including alternative peers' matrices demonstrate that peer effects estimates remain significant and invariant in magnitude when the specification models includes a randomized friendship network and another matrix which accounts for all the students in the class.

Furthermore, the study also finds that, besides the influence of direct friends in the classroom, indirect friends (friends distant by more than two links away) also play an important role on student performance. When we consider the direct and indirect friendship bonds, as well as the initial condition for learning, as measured here by the achievement-weighted Katz-Bonacich centrality, we find an aggregate peer impact of 45% standard errors of the final math grade.

Acknowledgements

This work was supported by the Joaquim Nabuco Foundation, a research institute from the Ministry of Education, Brazil.

References

- Angelo, M. (2010), 'A structural model of segregation in social networks'. URL: <https://ideas.repec.org/p/net/wpaper/1016.html>
- Badev, A. (2014), 'Discrete games in endogenous networks: theory and policy. Available in: http://www.antonbadev.com/papers/discr_games_endog_networks.pdf. Access date: 4 July 2014'.
- Ballester, C., Calvó-Armengol, A. & Zenou, Y. (2006), 'Who's who in networks. Wanted: the key player', *Econometrica* 74(5), 1403–1417.

- Boardman, A. E. & Murnane, R. J. (1979), 'Using panel data to improve estimates of the determinants of educational achievement', *Sociology of Education* 52(2), 113–121.
- Bonacich, P. (1987), 'Power and centrality: a family of measures', *American Journal of Sociology* 92(5), 1170–1182.
- Bramoullé, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* 150(1), 41–55.
- Calvo-Armengol, A., Patacchini, E. & Zenou, Y. (2009), 'Peer effects and social networks in education', *The Review of Economic Studies* 76(4), 1239–1267.
- Case, A. C. & Katz, L. F. (1991), 'The company you keep: the effects of family and neighborhood on disadvantaged youths', *National Bureau of Economic Research Working Paper Series* n. 3705.
URL: <http://www.nber.org/papers/w3705>
- Ding, W. & Lehrer, S. F. (2007), 'Do peers affect student achievement in China's secondary schools?', *The Review of Economics and Statistics* 89(2), 300–312.
- Duflo, E., Dupas, P. & Kremer, M. (2008), 'Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya', *National Bureau of Economic Research Working Paper Series* n. 14475.
URL: <http://www.nber.org/papers/w14475>
- Eisenkopf, G., Hessami, Z., Fischbacher, U. & Ursprung, H. (2011), Academic performance and single-sex schooling: evidence from a natural experiment in Switzerland, Report, Department of Economics, University of Konstanz.
URL: <https://ideas.repec.org/p/knz/dpteco/1134.html>
- 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.
- FUNDAJ - Fundação Joaquim Nabuco and Coordenação de Estudos Econômicos e Populacionais (2013), 'Acompanhamento longitudinal do desempenho escolar de alunos da rede pública de ensino fundamental do Recife'.
- Goux, D. & Maurin, E. (2007), 'Close neighbours matter: neighbourhood effects on early performance at school', *Economic Journal* 117(523), 1193–1215.
- Hanushek, E. A., Kain, J. F., Markman, J. M. & Rivkin, S. G. (2001), 'Does peer ability affect student achievement?', *National Bureau of Economic Research Working Paper Series* n. 8502.
URL: <http://www.nber.org/papers/w8502>
- Harris, J. R. (1995), 'Where is the child's environment? A group socialization theory of development', *Psychological Review* 102(3), 458–489.
- Lee, L.-f. (2007), 'Identification and estimation of econometric models with group interactions, contextual factors and fixed effects', *Journal of Econometrics* 140(2), 333–374.

- Lin, X. (2010), 'Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables', *Journal of Labor Economics* **28**(4), 825–860.
- Liu, X., Patacchini, E. & Zenou, Y. (2014), 'Endogenous peer effects: local aggregate or local average?', *Journal of Economic Behavior & Organization* **103**, 39–59.
- Manski, C. F. (1993), 'Identification of endogenous social effects: the reflection problem', *The Review of Economic Studies* **60**(3), 531–542.
- Oosterbeek, H. & Van Ewijk, R. (2014), 'Gender peer effects in university: evidence from a randomized experiment', *Economics of Education Review* **38**(C), 51–63.
- Patacchini, E., Rainone, E. & Zenou, Y. (2011), Dynamic aspects of teenage friendships and educational attainment, Report, Stockholm University, Department of Economics.
URL: https://ideas.repec.org/p/hhs/sunrpe/2011_0004.html
- Patacchini, E. & Venanzoni, G. (2014), 'Peer effects in the demand for housing quality', *Journal of Urban Economics* **83**(C), 6–17.
- PNUD (2005), *Metodologia de divisão do território do Recife adotada no Atlas Municipal do desenvolvimento humano*, Recife.
- Sacerdote, B. (2001), 'Peer effects with random assignment: results for Dartmouth roommates', *National Bureau of Economic Research Working Paper Series n. 7469*.
URL: <http://www.nber.org/papers/w7469>
- Sanbonmatsu, L., Kling, J. R., Duncan, G. J. & Brooks-Gunn, J. (2006), 'Neighborhoods and academic achievement: results from the moving to opportunity experiment', *National Bureau of Economic Research Working Paper Series n. 11909*.
URL: <http://www.nber.org/papers/w11909>
- Sund, K. (2009), 'Estimating peer effects in Swedish high school using school, teacher, and student fixed effects', *Economics of Education Review* **28**(3), 329–336.
- Vardardottir, A. (2013), 'Peer effects and academic achievement: a regression discontinuity approach', *Economics of Education Review* **36**, 108–121.
- Vigdor, J. & Nechyba, T. (2007), Peer Effects in North Carolina Public Schools, in 'Schools and the equal opportunity problem, ed. Ludger Woessmann and Paul E. Peterson', MIT Press.
- Zimmerman, D. J. (2003), 'Peer effects in academic outcomes: evidence from a natural experiment', *The Review of Economics and Statistics* **85**(1), 9–23.

Appendix A

Table A.1: Complete estimates: model from column (6), Table 2

	Estimate	Standardized	Std. Error	z value	Pr(> z)
Peers' math grade	0.014	0.062	0.006	2.417	0.016
Initial math grade	0.384	0.028	0.021	17.947	0.000
Male	0.894	0.027	1.120	0.798	0.425
Whites	-1.108	-0.096	0.831	-1.333	0.183
Age	-1.777	-0.017	0.420	-4.229	0.000
Newcomer	0.611	0.002	1.013	0.603	0.546
Dedication to Studies	0.025	0.034	0.221	0.115	0.908
Praise from Teacher	-0.814	-0.027	0.504	-1.615	0.106
Personality	0.487	0.023	0.366	1.330	0.183
Left out in classroom	0.655	0.044	0.565	1.161	0.246
Popular	1.056	0.032	0.480	2.201	0.028
Visits to friends' houses	-0.411	-0.006	0.273	-1.507	0.132
Sports clubs, gyms	-0.217	0.000	0.770	-0.282	0.778
Religiousness	0.010	0.020	0.458	0.021	0.983
Safety in the neighborhood	0.808	0.033	0.802	1.007	0.314
Number of friends	-0.398	-0.037	0.278	-1.433	0.152
Educational level (parent or guardian)	0.173	0.003	0.100	1.722	0.085
Male sex (parent or guardian)	0.147	0.014	0.971	0.151	0.880
White (parent or guardian)	0.604	0.030	0.871	0.693	0.489
Age (parent or guardian)	0.058	0.003	0.043	1.361	0.174
Marital Status (parent or guardian)	-0.111	-0.018	0.655	-0.169	0.866
Kinship (parent or guardian) with student	0.745	0.008	0.910	0.819	0.413
Beneficiary of social program (parent or guardian)	-0.279	-0.045	0.705	-0.396	0.692
Student grade report (parent or guardian)	-1.419	-0.005	0.641	-2.212	0.027

No. of observations: 1,670 students

No. networks/classrooms: 139 classrooms

Source: Original compilation based on FUNDAJ ((FUNDAJ - Fundação Joaquim Nabuco and Coordenação de Estudos Econômicos e Populacionais 2013)).

Table A.1: Complete estimates: model from column (6), Table 2 (continued)

	Estimate	Standardized	Std. Error	z value	Pr(> z)
Male - peers average	0.187	0.021	1.267	0.148	0.883
Whites - peers average	-1.223	-0.013	1.262	-0.969	0.332
Religiousness - peers average	-0.435	-0.011	0.698	-0.623	0.533
Newcomer - peers average	-0.440	-0.001	1.428	-0.308	0.758
Sports clubs, gyms - peers average	-0.046	-0.016	1.204	-0.039	0.969
Dedication to Studies - peers average	0.226	0.017	0.325	0.696	0.486
Personality - peers average	0.455	0.012	0.564	0.806	0.420
Left out in classroom - peers average	-0.488	-0.015	0.894	-0.546	0.585
Popular - peers average	-0.526	-0.031	0.753	-0.698	0.485
Safety in the neighborhood - peers average	-1.691	-0.013	1.168	-1.447	0.148
Praise from Teacher - peers average	-0.458	-0.026	0.787	-0.583	0.560
Educational level (parent or guardian) - peers average	0.158	0.003	0.145	1.086	0.277
Male sex (parent or guardian) - peers average	-0.207	-0.025	1.406	-0.148	0.883
White (parent or guardian) - peers average	1.441	0.034	1.259	1.145	0.252
Age (parent or guardian) - peers average	0.092	0.007	0.062	1.480	0.139
Marital Status (parent or guardian) - peers average	-0.308	-0.001	0.971	-0.318	0.751
Kinship (parent or guardian) with student - peers average	-0.079	-0.028	1.317	-0.060	0.952
Beneficiary of social program (parent or guardian) - peers average	-1.229	-0.014	1.047	-1.174	0.240
Student grade report (parent or guardian) - peers average	-0.622	-0.012	0.945	-0.658	0.510
Initial math grade - peers average	0.015	0.024	0.032	0.490	0.624
Proportion of peers neighbors	-1.107	-0.062	1.076	-1.030	0.303
(Intercept)	33.394	0.399	8.990	3.715	0.000

No. of observations: 1,670 students

No. networks/classrooms: 139 classrooms

Source: Original compilation based on FUNDAJ ((FUNDAJ - Fundação Joaquim Nabuco and Coordenação de Estudos Econômicos e Populacionais 2013)).